[SPECIAL EDITION] CANCER

Publish your research in
the *Science* family of journals

The *Science* family of journals (*Science*, *Science Advances*, *Science Immunology*, *Science Robotics*, *Science Signaling*, and *Science Translational Medicine*) are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!
Learn more at **ScienceMag.org/journals**

SPECIAL EDITION: Cancer

Select research published in *Science, Science Advances, Science Immunology, Science Signaling,* and *Science Translational Medicine*

Cancer remains one of the leading causes of death worldwide, but the rapid pace of innovation and discovery in this field is increasingly transforming cancer into a survivable and sometimes even curable condition. In this booklet are samples of empirical research from five featured journals from the *Science* family, showcasing various discoveries in cancer biology and treatment.

Image: Davide Bonazzi/@salzmanart

Editor-in-Chief
Jeremy Berg, Ph.D.
Science family of journals
AAAS, Washington, DC

Science | AAAS

Science is a leading outlet for scientific news, commentary, and cutting-edge research. The journal publishes original peer-reviewed research across the sciences, with articles that consistently rank among the most cited in the world.

Editors
Lisa D. Chong, Ph.D.
Valda J. Vinson, Ph.D.
Jake S. Yeston, Ph.D.

Cancer Editors
Priscilla Kelly, Ph.D.
Paula Kiberstis, Ph.D.

Science Advances | AAAS   ⛓ OPEN ACCESS

Science Advances is the digital, open access expansion of *Science* magazine. The journal publishes significant, innovative original research that advances the frontiers of science across a broad range of disciplines, from bioscience and medicine to neuroscience, physics, ecology, and material sciences.

Managing Editor
Philippa Benson, Ph.D.

Editor
Ali Shilatifard, Ph.D.

Editorial Team
Lynden Archer, Ph.D.
Aaron Clauset, Ph.D.
Dale Dorsett, Ph.D.
Sarah M. Fortune, M.D.
Kip V. Hodges, Ph.D.
Jeremy Jackson, Ph.D.

Zakya H. Kafafi, Ph.D.
Kevin S. LaBar, Ph.D.
Shahid Naeem, Ph.D.
Leslie Vosshall, Ph.D.
Philip Yeagle, Ph.D.
Warren Warren, Ph.D.

Science Immunology | AAAS

Science Immunology publishes original, peer-reviewed, science-based research articles that report critical advances in all areas of immunological research, including important new tools and techniques.

Chief Scientific Advisors
Abul K. Abbas, M.D.
University of California, San Francisco

Federica Sallusto, Ph.D.
Università della Svizzera Italiana

Editor
Ifor Williams, M.D., Ph.D.

Editorial Team
Anand Balasubramani, Ph.D.

Science Signaling | AAAS

Science Signaling publishes peer-reviewed, original research investigating cell signaling that underlies physiology and disease, on the molecular, cellular, intercellular and organismal levels.

Chief Scientific Advisor and Academic Editor
Michael B. Yaffe, M.D., Ph.D.
Massachusetts Institute of Technology

Editor
John F. Foley, Ph.D.

Editorial Team
Leslie K. Ferrarelli, Ph.D.
Annalisa M. VanHook, Ph.D.
Erin R. Williams, Ph.D.
Wei Wong, Ph.D.

Science Translational Medicine | AAAS

Science Translational Medicine is an interdisciplinary journal that publishes translational research with impact for human health that fills the knowledge gaps between preclinical studies and clinical applications.

Chief Scientific Advisors
Elazer R. Edelman, M.D., Ph.D.
Massachusetts Institute of Technology

Garret FitzGerald, M.D.
University of Pennsylvania

Editor
Orla M. Smith, Ph.D.

Editorial Team
Catherine A. Charneski, Ph.D.
Caitlin A. Czajka, Ph.D.
Mattia Maroso, Ph.D.
Yevgeniya Nusinovich, M.D., Ph.D.
Lindsey Pujanandez, Ph.D.

Learn more and submit your research today: aaas.org/science-journals

IN THIS BOOKLET

Learn more and submit your research today: aaas.org/science-journals

SOMETIMES THE GRASS REALLY IS GREENER SOMEPLACE NEW.

🔍 **Find your next job at ScienceCareers.org**

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place. Start here: sciencecareers.org

Science Careers

FROM THE JOURNAL SCIENCE AAAS

ORGANOIDS

Patient-derived organoids model treatment response of metastatic gastrointestinal cancers

Georgios Vlachogiannis,[1] Somaieh Hedayat,[1] Alexandra Vatsiou,[2] Yann Jamin,[3] Javier Fernández-Mateos,[1,2] Khurum Khan,[1,4] Andrea Lampis,[1] Katherine Eason,[1] Ian Huntingford,[1] Rosemary Burke,[5] Mihaela Rata,[3] Dow-Mu Koh,[3,6] Nina Tunariu,[3,6] David Collins,[3] Sanna Hulkki-Wilson,[1] Chanthirika Ragulan,[1] Inmaculada Spiteri,[2] Sing Yu Moorcroft,[4] Ian Chau,[4] Sheela Rao,[4] David Watkins,[4] Nicos Fotiadis,[6] Maria Bali,[3,6] Mahnaz Darvish-Damavandi,[1] Hazel Lote,[1,4] Zakaria Eltahir,[1] Elizabeth C. Smyth,[4] Ruwaida Begum,[4] Paul A. Clarke,[5] Jens C. Hahne,[1] Mitchell Dowsett,[7] Johann de Bono,[8] Paul Workman,[5] Anguraj Sadanandam,[1] Matteo Fassan,[9] Owen J. Sansom,[10] Suzanne Eccles,[5] Naureen Starling,[4] Chiara Braconi,[4,5] Andrea Sottoriva,[2] Simon P. Robinson,[3] David Cunningham,[4] Nicola Valeri[1,4]*

Patient-derived organoids (PDOs) have recently emerged as robust preclinical models; however, their potential to predict clinical outcomes in patients has remained unclear. We report on a living biobank of PDOs from metastatic, heavily pretreated colorectal and gastroesophageal cancer patients recruited in phase 1/2 clinical trials. Phenotypic and genotypic profiling of PDOs showed a high degree of similarity to the original patient tumors. Molecular profiling of tumor organoids was matched to drug-screening results, suggesting that PDOs could complement existing approaches in defining cancer vulnerabilities and improving treatment responses. We compared responses to anticancer agents ex vivo in organoids and PDO-based orthotopic mouse tumor xenograft models with the responses of the patients in clinical trials. Our data suggest that PDOs can recapitulate patient responses in the clinic and could be implemented in personalized medicine programs.

High-throughput sequencing has been extensively used in precision medicine to identify somatic mutations that can be exploited for cancer treatment and drug development (1). However, the limited role of genomic profiling in predicting response to targeted therapies and the limitations of preclinical models used for drug validation are important obstacles hampering the success of personalized medicine (2). Co-clinical trials are parallel studies in which drug responses in patients are matched to laboratory preclinical models to personalize treatment and understand mechanisms of chemosensitivity through functional genomics and reverse translation (3). Most co-clinical trials rely on the use of genetically

engineered mouse models or patient-derived xenografts, posing logistic, ethical, and economic issues (4).

LGR5[+] stem cells can be isolated from a number of organs and propagated as epithelial organoids in vitro to study physiology and neoplastic transformation (5). Most studies on human colorectal cancer (CRC) organoids have been conducted on cultures derived from primary tumors (6). In contrast, examples of PDOs from metastatic cancer sites remain sparse (7–9). Furthermore, very limited evidence is available on the ability of PDOs to predict response to treatment in the clinic (10). Here we present a living biobank of PDOs from heavily pretreated metastatic gastrointestinal cancer patients and show examples of how the drug responses of these cancer organoids can be compared with those of the actual patient.

A total of 110 fresh biopsies from 71 patients enrolled in four prospective phase 1/2 clinical trials were processed between October 2014 and February 2017. In line with previous data (7), PDOs were grown from 70% of biopsies with a cellularity of 2+ and above, and their establishment rate strongly correlated with tumor cellularity in the parental biopsy (χ^2 test, $P < 0.0001$). No inverse correlation was observed between PDO establishment rate and presence of necrosis (cutoff ≥ 20%). Tumor percentage is a key limiting factor for genomic and transcriptomic analyses. When the 60% threshold used in large

sequencing studies of primary CRC (11) or gastro-esophageal cancer (GOC) (12) was applied in our cohort, we found no correlation between PDO take-up rate and tumor percentage, suggesting that PDOs can also be established in cases of a low tumor/stroma ratio, thus allowing the ex vivo expansion of the cancer population in samples that would have otherwise failed quality-control tests for next-generation sequencing (NGS).

PDOs presented in this study were derived from ultrasound (n = 20), computed tomography (CT)–guided (n = 7), or endoscopic (n = 2) biopsies of metastatic CRC (mCRC; n = 16), metastatic GOC (mGOC; n = 4), and metastatic cholangiocarcinoma (n = 1) patients (fig. S1). Liver, pelvic, peritoneal, and nodal metastases of chemorefractory patients were used to establish PDOs. In several cases, PDOs were established from sequential biopsies at baseline (BL), at the time of best response [partial response (PR) or stable disease (SD)], and at the time of disease progression (PD), as well as from multiregion biopsies (table S1).

Histological evaluation revealed notable morphological similarities between PDOs and the patient biopsies from which they were originally derived (Fig. 1, A and B, and figs. S2A and S2B). Immunohistochemistry markers routinely used in the diagnosis of CRC (CDX-2 and CK7) showed that the parental tumor's expression pattern was maintained in PDOs, even when derived from sequential biopsies during treatment (fig. S2, C to E). Similarly, amplification of oncogenic drivers such as ERBB2 (Fig. 1C and fig. S2F) and rearrangements in FGFR2 (fig. S2G) were retained in PDOs from mGOC and metastatic cholangiocarcinoma, respectively.

NGS was used to profile 151 cancer-related genes in both PDOs (n = 23) and their parental biopsies; archival material from primary cancer or pretreatment diagnostic biopsies was also sequenced for eight patients, and whole-genome sequencing (WGS) was performed for one PDO (tables S2 and S3). The molecular landscape of our PDOs (Fig. 1D) largely overlapped with that reported for mCRC and mGOC in the MSK-IMPACT study (1), with the exception of SRC and EGFR amplifications and ATM and BRCA2 mutations that were more frequent in our mCRC PDO cohort (table S4). Overall, a 96% overlap in mutational spectrum was observed between PDOs and their parental biopsies (Fig. 1D), whereas intratumor heterogeneity was observed between archival material (primary cancer) and metastatic deposits (biopsies or PDOs) (fig. S3A and table S2). PDOs were able to capture spatiotemporal intratumor heterogeneity when established from multiple biopsies at the time of disease progression and when compared with PDOs established at the beginning of treatment (Fig. 1D, fig. S3A, and table S2). Similar results were observed for copy number alterations (CNAs) in PDOs and biopsies collected at different time points during treatment (figs. S3B and S4). WGS confirmed CNAs extrapolated from targeted NGS of PDOs or PDO-derived orthotopic tumors (PDO-xenografts) (figs. S3B and S4); CNAs detected in key oncogenic

Fig. 1. Histopathological, molecular, and functional characterization of patient-derived organoids (PDOs). (**A**) Phase-contrast image of a mCRC PDO culture (top) and hematoxylin and eosin staining comparing organoids to their matching patient biopsy (bottom). (**B**) Intestinal and diffuse growth patterns are retained in mGOC PDOs. (**C**) *ERBB2* amplification and overexpression in mGOC PDOs and parental tissue biopsy (CISH, chromogenic in situ hybridization; IHC, immunohisto- chemistry). (**D**) Heatmap showing the most frequently mutated and/or copy number–altered genes in PDOs (left) and Venn diagram demon- strating 96% mutational overlap between PDOs and parental tissue biopsies (right). (**E**) Target engagement in genotype–drug phenotype combinations: pathway analysis downstream of ERBB2 in *ERBB2*- amplified and nonamplified PDOs treated with lapatinib (24 hours) (right), BRAF inhibition (24 hours) (center), and AKT inhibition (4 hours)

(left). wt, wild type. (**F**) Concentration-dependent effect of the dual PI3K/mTOR inhibitor GDC-0980 in three PDOs from patient R-009, all carrying an acquired PIK3CA mutation (H1047R). PDOs established from a liver metastasis biopsied at disease progression (R-009 PD-A) that also harbored *PIK3CA* amplification showed concentration-dependent response to GDC-0980. *PIK3CA*-mutant but nonamplified PDOs established before regorafenib treatment (R-009 BL) or from a different liver metastasis biopsied at disease progression (R-009 PD-B) did not respond to GDC-0980. Viability data shown are means ± SEM of indicated independent experiments. (**G**) Correlation (Fisher's exact test) between presence of *RB1* amplification in PDOs (D) and response to the CDK4/CDK6 inhibitor palbociclib in the reported drug screen (fig. S9A). BL, baseline; SD, stable disease; PD, posttreatment/progressive disease.

drivers were further validated by digital-droplet polymerase chain reaction (fig. S5). High concordance was observed in mutational, CNA, and transcriptomic profiling over successive passages when PDOs were tested before and after several months of continuous culture [passage range, 5 to 13; mutations, coefficient of determination (R^2) = 0.96, $P < 0.0001$; CNA, R^2 = 0.97, $P < 0.0001$; gene expression (RNA sequencing), R^2 = 0.7, $P < 0.001$] (fig. S6).

Next we tested the feasibility of using PDOs derived from metastatic cancers as drug-screening tools and validated the robustness of our approach by identifying several genotype–drug phenotype correlations across the PDO panel. We ran three-dimensional (3D) screening assays over a period of 2 weeks (figs. S7 and S8), using a library of 55 drugs now in phase 1 to 3 clinical trials or in clinical practice (table S5). The heatmap shown in fig. S9A summarizes the screening data; hit validation at lower drug concentrations is reported in fig. S9B. For all 19 screens, a very high correlation was observed among each screen's three replicate assays and controls (fig. S10).

F-013 was the only *ERBB2*-amplified PDO in our cohort (Fig. 1C), and it exhibited the strongest response to lapatinib [dual ERBB2/epidermal growth factor receptor (EGFR) inhibitor]; lapatinib potently inhibited the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K)/AKT signaling downstream of EGFR/ERBB2, inducing apoptosis in the F-013 PDO (Fig. 1E and fig. S9A). In a PDO (F-014) in which *EGFR* was amplified but *ERBB2* was not, lapatinib had no effect on viability and only modestly reduced MAPK and PI3K/AKT signaling (Fig. 1E and fig. S9A).

Similarly, across all PDOs, F-016 was the only tumor carrying an *AKT1* amplification and E17K mutation (E, glutamic acid; K, lysine) (Fig. 1D) and was the only one to respond strongly to both AKT inhibitors present in the drug library (MK-2206
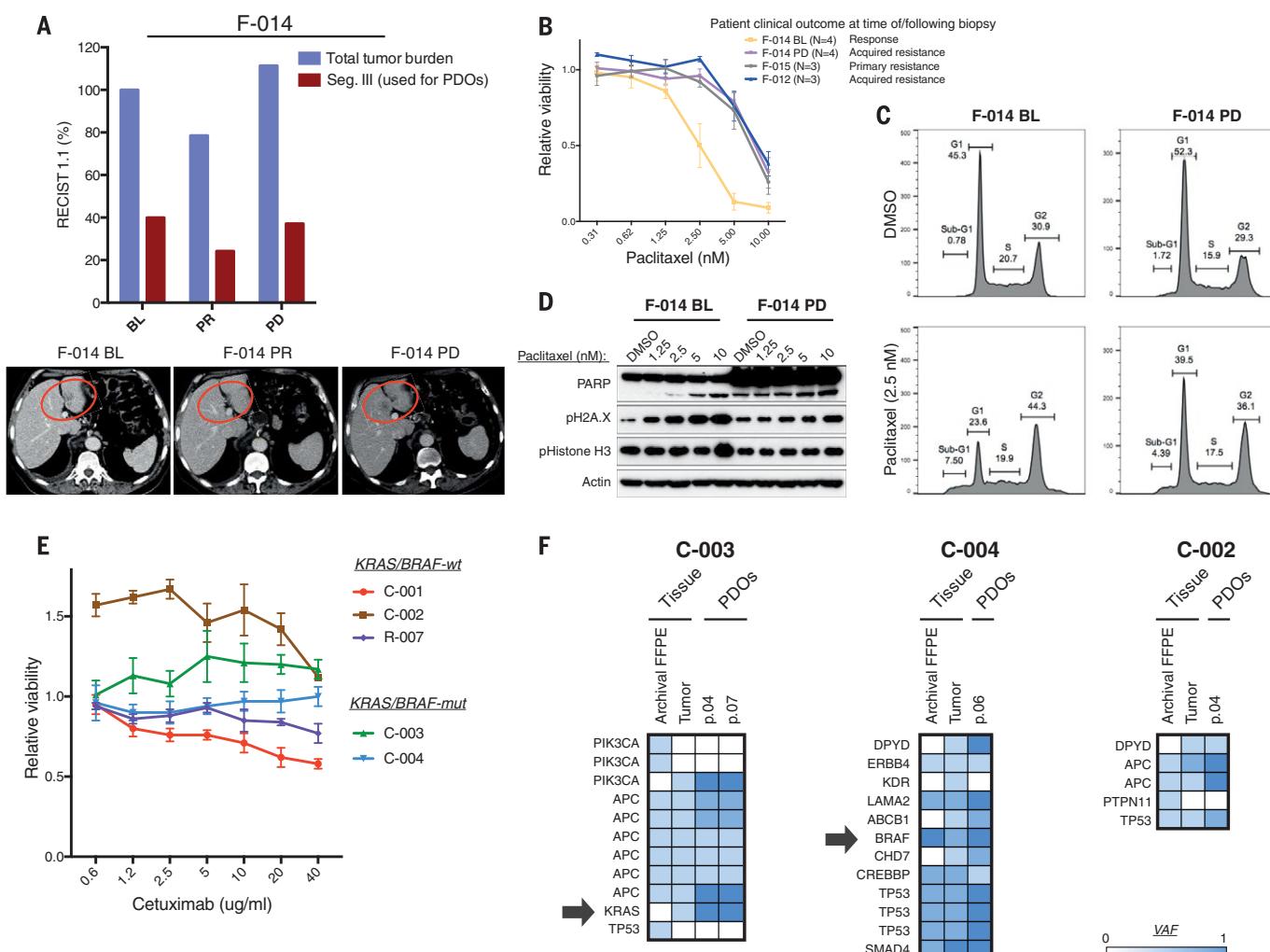
Fig. 2. PDO-based ex vivo co-clinical trials in mGOC and mCRC.
(**A**) PDOs were generated from sequential biopsies of a liver metastasis (red circles in the bottom panel) of mGOC patient F-014 that showed initial response to paclitaxel (F-014 BL) and subsequently progressed (F-014 PD). Violet bars indicate overall tumor volume [according to RECIST (Response Evaluation Criteria in Solid Tumors) 1.1], and red bars indicate volume of the target metastasis used to generate PDOs. (**B**) Cell viability upon paclitaxel treatment was compared in BL and PD PDOs from patient F-014 and PDOs from patients that exhibited primary (F-015) or acquired (F-012) resistance to paclitaxel in the clinic. Viability data shown are means ± SEM of indicated independent experiments. (**C**) Cell cycle analysis upon paclitaxel treatment in the F-014 BL PDO compared with the F-014 PD PDO. DMSO, dimethyl sulfoxide. (**D**) Concentration-dependent DNA damage was observed in the F-014 BL PDO in response to paclitaxel but not in PDOs from the same patient established at PD. (**E**) PDOs were established from BL (C-003 and C-004) and PD (C-001 and C-002) biopsies from patients treated with the anti-EGFR monoclonal antibody cetuximab. PDOs were treated with cetuximab in vitro; data shown are means ± SD from independent experiments performed in triplicate. (**F**) Molecular analysis of BL and PD PDOs, matching biopsy (tumor), and primary bowel cancer (archival); arrows indicate the presence of clonal or subclonal mutations in *BRAF* or *KRAS*, respectively, in two patients. VAF, variant allele frequency; FFPE, formalin-fixed paraffin-embedded.
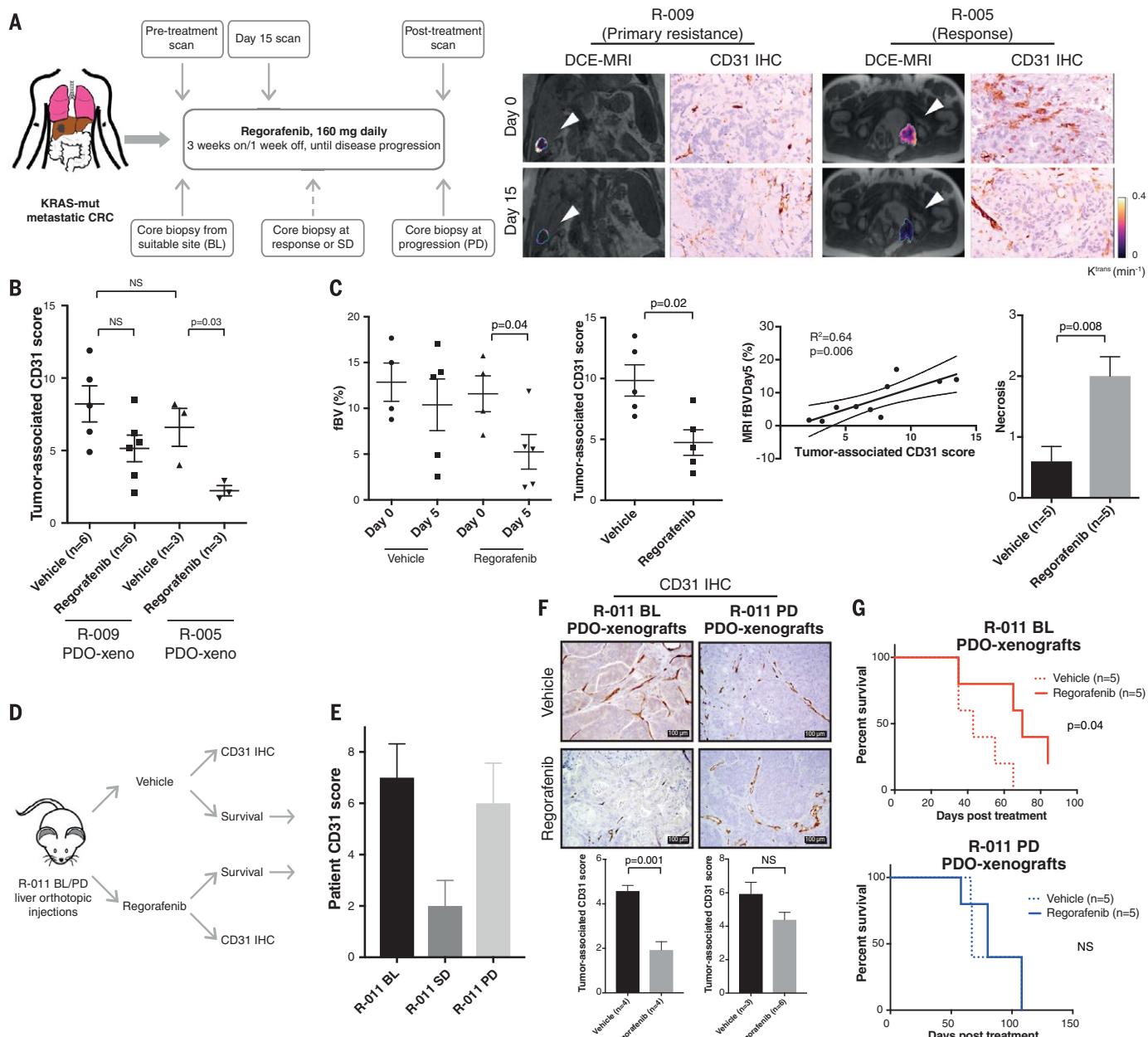
Fig. 3. PDO-based co-clinical trials mimic primary and acquired resistance to regorafenib in mice. (**A**) mCRC patients on regorafenib treatment underwent biopsies at BL, SD, or PD. An early reduction (15 days) in functional imaging (DCE-MRI) parameters correlated with changes in microvasculature assessed by CD31 staining and clinical benefit from regorafenib (right). Arrowheads indicate CRC metastases; K^{trans}, volume transfer constant. (**B**) Changes in microvasculature in response to regorafenib were assessed in PDO-xenografts in mice by quantification of tumor-associated CD31-positive vessels. Data show PDO-xenografts from a primary resistant patient (R-009) and a long-term responder (R-005) to regorafenib. Means ± SD from the indicated number of mice (*n*) in a representative experiment are shown; significance was determined using Student's unpaired *t* test. (**C**) Reduction in fractional blood volume (fBV) in regorafenib-treated mice carrying long-term regorafenib responder (R-005) PDO-xenografts. A total of 10 animals were analyzed (five in each arm); shown are the means ± SD of an individual experiment. Day 0 fBV values could not be obtained for two animals owing to respiratory movement. Significance was determined using Student's paired *t* test for fBV and unpaired *t* test for CD31 and necrosis.

(**D**) Schematic representation of the animal experiment using PDOs from patient R-011, established pre- and posttreatment with regorafenib. Mice carrying liver orthotopic R-011 pretreatment (BL) and posttreatment (PD) PDO-xenografts were randomized to control and treatment arms and treated with vehicle or regorafenib for 10 days. After treatment, each arm was further randomized to a cohort culled for histopathological analysis and a survival cohort, which was monitored over time. (**E**) CD31 immunostaining in the parental patient BL, SD, and PD biopsies, demonstrating an initial reduction in tumor microvasculature in response to regorafenib. Shown are means ± SD calculated by scoring 10 high-power-field tumor areas. (**F**) Representative images (top) and analysis (bottom) of CD31 immunostaining in the BL and PD R-011 PDO-xenografts. Shown are means ± SD calculated by scoring at least 10 high-power-field tumor areas per animal in an individual experiment; *n*, number of animals analyzed in each group. Significance was determined using Student's unpaired *t* test. (**G**) Kaplan-Mayer curves for regorafenib- or vehicle-treated mice bearing BL and PD PDO-xenografts from patient R-011 from an individual experiment (*n*, number of mice analyzed). Significance was determined using the Mantel-Cox log-rank test.
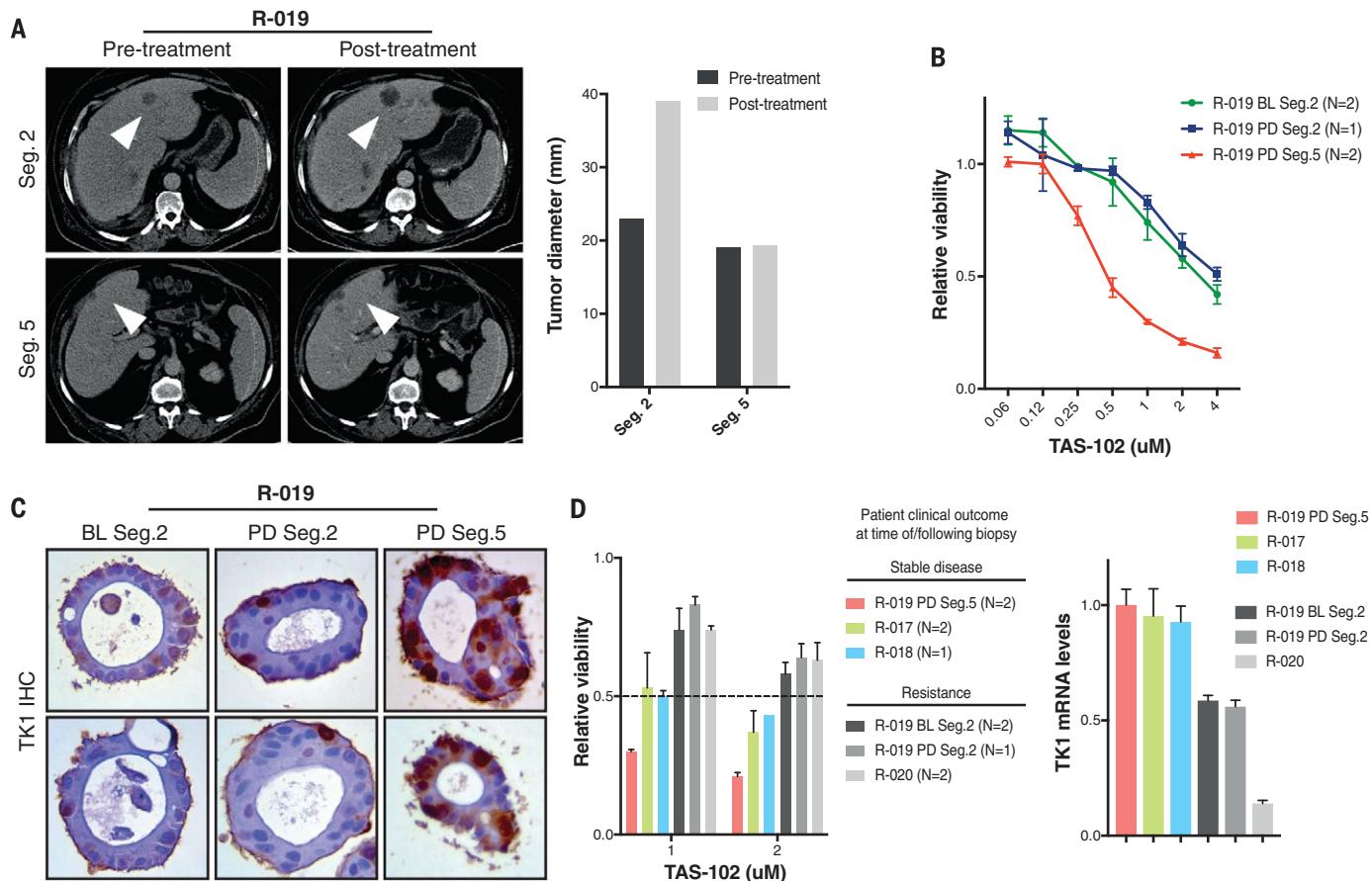
Fig. 4. PDOs recapitulate intra- and interpatient heterogeneity in response to TAS-102. (**A**) PDOs were established from a patient (R-019) with mixed response to TAS-102. Whereas the segment 2 metastasis rapidly progressed, the segment 5 metastasis remained stable upon TAS-102 treatment (white arrowheads in the CT scan indicate metastases; bars represent pre- and posttreatment measurements of the indicated metastases). (**B**) Ex vivo concentration-response curves in BL and PD multiregion PDOs from patient R-019 (with mixed response to TAS-102). (**C**) TK1 IHC expression in TAS-102–refractory (segment 2) and –sensitive (segment 5) PDOs. (**D**) Cell viability (left) and TK1 mRNA expression (right) in PDOs from TAS-102–responsive and –refractory patients. In (B) and (D), *N* indicates the number of independent experiments, and viability values are expressed as means ± SEM.

and GSK690693) (Fig. 1E and fig. S9A). One mCRC PDO (C-004) harbored a BRAF V600E mutation (V, valine) (Fig. 1D) and was the only PDO that showed significantly decreased viability after treatment with the BRAF inhibitor vemurafenib (fig. S9A). Consistent with this, vemurafenib selectively inhibited MEK/ERK (MAPK kinase/extracellular signal–regulated kinase) signaling in the C-004 PDO (Fig. 1E) but failed to induce apoptosis, in keeping with the lack of efficacy of single-agent BRAF inhibitors in mCRC (*13*).

Overall, *PIK3CA* mutations were not predictive of response to GDC-0980 (a dual PI3K/mTOR inhibitor) in the PDOs panel (Fig. 1D and fig. S9A). In line with this observation, in a patient from whom pre- and posttreatment PDOs were established from multiple metastases (R-009 BL, PD-A, and PD-B), a PIK3CA H1047R mutation (H, histidine; R, arginine) common to all the PDOs was not associated with any response to GDC-0980. However, PDOs carrying a synchronous *PIK3CA* amplification (R-009 PD-A) showed a concentration-dependent reduction in cell viability in response to GDC-0980 (Fig. 1F and fig. S3A). Last, in keeping with

published data (*14*), a significant correlation was observed between *RB1* amplification and sensitivity of PDOs to palbociclib [cyclin-dependent kinase 4 (CDK4)/CDK6 inhibitor] (Fig. 1G).

After extensive molecular and functional characterization of our PDOs, we examined their clinical predictive value in 21 comparisons of clinical responses observed in patients with ex vivo response data gathered in organoids (table S6). Taxanes are a standard second-line treatment option for metastatic gastric cancer; however, efficacy is modest, and no predictive biomarkers are available to inform clinical decisions (*15*). We compared response to paclitaxel in sequential PDOs established before and after treatment in a paclitaxel-sensitive patient (F-014) with that in PDOs established from liver metastases of two paclitaxel-resistant patients (Fig. 2, A and B). PDOs derived from the responsive metastasis showed a GI_{50} (concentration that inhibits growth of cancer cells by 50%) for clinically relevant paclitaxel concentrations (*16*) that was about one-fourth that for PDOs from the same patient derived at progression; these resistant PDOs dem-

onstrated an identical paclitaxel concentration-response profile to the two PDOs established from paclitaxel-refractory patients (Fig. 2B). Cell cycle analysis showed marked apoptosis and G_2 arrest upon taxane treatment in the pretreatment F-014 PDOs, whereas no significant difference was observed in PDOs established at progression (Fig. 2C and fig. S11A). Similarly, paclitaxel induced concentration-dependent DNA damage, mitotic arrest, and apoptosis in the pretreatment F-014 PDOs but had a much weaker impact on the progression (and thus resistant) PDOs (Fig. 2D). Consistent with data observed for second-line treatment, a ~10-fold difference in GI_{50} was observed in response to the combination of 5-fluorouracil and cisplatin in PDOs collected from chemosensitive and chemorefractory mGOC patients receiving first-line treatment (fig. S11B), highlighting the clinical potential of PDOs for treatment selection in cancers of unmet need.

Anti-EGFR monoclonal antibodies, regorafenib, and TAS-102 are U.S. Food and Drug Administration–approved options for treatment of chemorefractory mCRC; however, with the exception of RAS pathway

mutations for anti-EGFR therapy, there are no validated clinical biomarkers for patient selection in this setting. We initially tested the predictive value of PDOs in mCRC by comparing response to anti-EGFR treatment (cetuximab) in five PDOs and their respective patients (Fig. 2E). Two PDOs established from BL biopsies before anti-EGFR treatment in the PROSPECT-C trial showed no response to cetuximab, in keeping with the primary resistance observed in these two patients in the clinic. Unsurprisingly (17), both PDOs and their respective patient biopsies harbored either *KRAS* G12D (subclonal) (G, glycine; D, aspartic acid) or *BRAF* V600E (clonal) mutations (Fig. 2F). The third cetuximab-resistant PDO (C-002) was established from the progression biopsy of a patient who initially responded to cetuximab, and it harbored an *EGFR* amplification (Fig. 1D and fig. S5), no RAS pathway mutational aberrations (Fig. 2F), and high amphiregulin mRNA levels. Despite these molecular markers being suggestive of responsiveness to cetuximab, the C-002 PDO showed no response (and, in fact, paradoxically showed enhanced proliferation) upon cetuximab treatment, in line with the respective patient's clinical outcome, thus highlighting the potential of PDOs to predict clinical outcomes better than molecular pathology alone. Another *KRAS* wild-type PDO derived from a slow-growing progressing metastasis in a patient with otherwise stable disease (C-001) (fig. S11C) showed a marginal response to cetuximab. Last, the *KRAS* wild-type PDO established from a BL biopsy of a patient enrolled in the PROSPECT-R trial (R-007) (fig. S11C) showed response to cetuximab at concentrations higher than 5 μg/ml; this, however, could not be compared with clinical response, because the patient did not receive anti-EGFR monoclonal antibodies.

Next we tested the ability of PDOs to recapitulate response to regorafenib, a multiple tyrosine kinase inhibitor blocking oncogenic and angiogenic signaling pathways. No response to regorafenib was observed in our 3D ex vivo screening assays (fig. S9A), an observation in keeping with our recently reported clinical results from the PROSPECT-R trial (18), suggesting that response to regorafenib is mainly driven by its antiangiogenic effect (Fig. 3A).

To match responses to regorafenib in the clinic and in aligned PDOs, we established an orthotopic human tumor xenograft model by implanting luciferase-expressing (Luc$^+$) PDOs in the liver of NSG mice (PDO-xenografts) (fig. S12A). We initially compared response to regorafenib in PDO-xenografts from a patient with primary resistance (R-009; n = 11) and from a patient who achieved a durable (10-month) response (R-005; n = 6) to regorafenib (fig. S12, B and C). In keeping with the clinical responses (Fig. 3A), PDO-xenografts from the regorafenib-sensitive patient displayed a significant (P = 0.03) reduction in their microvasculature in response to regorafenib, as revealed by CD31 immunostaining; in contrast, no significant changes were observed in PDO-xenografts from the regorafenib-resistant patient (Fig. 3B). To mimic our clinical obser-

vations, we performed functional susceptibility contrast magnetic resonance imaging (MRI) of PDO-xenografts of the responding patient (R-005; n = 10) before and after treatment (fig. S12D). In line with dynamic contrast-enhanced MRI (DCE-MRI) results in patients (Fig. 3A), susceptibility contrast MRI revealed a significant reduction in tumor fractional blood volume (fBV) in regorafenib-treated mice (Fig. 3C). These changes were associated with a reduction in CD31 staining and increased necrosis (Fig. 3C). Notably, across all animals, a robust correlation was observed between the fBV values obtained from susceptibility contrast MRI and the microvasculature assessment (CD31) of the same samples (R^2 = 0.64, P = 0.006) (Fig. 3C). Consistent with our clinical data, changes in microvasculature indicative of response appeared to be independent of changes in tumor volume (fig. S12E) (18). Three different histopathological growth patterns (HGPs)—desmoplastic HGP, pushing HGP, and replacement HGP—have been associated with different degrees of response to anti-angiogenic drugs, with the replacement HGP being frequently associated with vessel co-option and primary resistance (19). In our experiments, a predominance of replacement HGP, and thus vessel co-option, was observed in PDO-xenografts from the resistant patient, whereas tumors established from the PDOs of the sensitive patient showed a prevalence of desmoplastic and pushing HGPs (fig. S12F), suggesting that vessel co-option might be the mechanism underpinning primary resistance to regorafenib. When the responder to regorafenib (R-005) progressed and received subsequent treatment, he was enrolled in a phase 1 trial of the ATR inhibitor VX-970. No response was observed in this patient with VX-970 monotherapy, and this was in keeping with the lack of response to ATM/ATR inhibitors observed in his PDOs in the drug screening reported in fig. S9A.

To test the PDOs' ability to capture tumor evolution and acquired resistance to treatment, we generated xenografts using PDOs from the same liver metastasis before (BL) and after (PD) treatment in mCRC patient R-011 that exhibited initial response to regorafenib and subsequently progressed (fig. S13A). Mice were randomized to treatment and control arms, and, after treatment, each arm was further randomized for survival or functional analysis (Fig. 3D). In line with clinical findings (Fig. 3E) (18), CD31 immunostaining revealed a ~60% reduction in microvasculature in response to regorafenib in BL PDO-xenografts (P = 0.001), whereas no significant change was observed in PD PDO-xenografts (Fig. 3F). More importantly, regorafenib treatment offered a selective survival benefit in mice carrying BL PDO-xenografts (Fig. 3G and fig. S13B), confirming the predictive value of PDOs and their ability to reflect cancer evolution upon treatment.

TAS-102, a combination of the nucleoside analog trifluridine and the thymidine phosphorylase inhibitor tipiracil, is approved for the treatment of chemorefractory mCRC, but no validated biomarkers are available (20). We compared clinical and preclinical response to TAS-102 in six organ-

oids from four different patients treated with TAS-102. Initially, we tested response to TAS-102 in PDOs from a patient (R-019) who had a mixed response, with stability of disease in one of the liver metastases (segment 5) and rapid progression in another (segment 2) (Fig. 4A). Ex vivo concentration-response data showed about an eightfold difference in GI$_{50}$ between PDOs derived from the TAS-102–sensitive metastasis and those derived from pre- and posttreatment biopsy of the rapidly progressing metastasis (Fig. 4B), highlighting the ability of PDOs to recapitulate intrapatient heterogeneity. TK1 has been proposed as a potential biomarker of response to TAS-102 (21); TK1 protein expression was indeed higher in PDOs from the responding metastasis than in those from the nonresponding site (Fig. 4C). When we extended the TAS-102 sensitivity analysis to PDOs from three other patients, we confirmed that PDOs from patients who achieved disease control were sensitive to low micromolar concentrations of TAS-102, whereas no significant effect on cell viability was observed in PDOs from resistant (primary or acquired) patients (Fig. 4D, left); consistent with previous data, TK1 mRNA expression was higher in PDOs from patients that achieved stable disease in response to TAS-102 (Fig. 4D, right).

Overall, for the PDOs that we analyzed, we found 100% sensitivity, 93% specificity, 88% positive predictive value, and 100% negative predictive value in forecasting response to targeted agents or chemotherapy in patients (Fisher's exact test, P < 0.0001) (table S7). Our data suggest that PDOs can be exploited for functional genomics to simulate cancer behavior ex vivo and integrate molecular pathology into the decision-making process of early-phase clinical trials.

REFERENCES AND NOTES

1. A. Zehir *et al.*, *Nat. Med.* **23**, 703–713 (2017).
2. E. E. Voest, R. Bernards, *Cancer Discov.* **6**, 130–132 (2016).
3. A. T. Byrne *et al.*, *Nat. Rev. Cancer* **17**, 254–268 (2017).
4. J. G. Clohessy, P. P. Pandolfi, *Nat. Rev. Clin. Oncol.* **12**, 491–498 (2015).
5. A. L. Bredenoord, H. Clevers, J. A. Knoblich, *Science* **355**, eaaf9414 (2017).
6. M. van de Wetering *et al.*, *Cell* **161**, 933–945 (2015).
7. F. Weeber *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13308–13311 (2015).
8. M. Fujii *et al.*, *Cell Stem Cell* **18**, 827–838 (2016).
9. C. Pauli *et al.*, *Cancer Discov.* **7**, 462–477 (2017).
10. J. F. Dekkers *et al.*, *Sci. Transl. Med.* **8**, 344ra84 (2016).
11. Cancer Genome Atlas Network, *Nature* **487**, 330–337 (2012).
12. Cancer Genome Atlas Research Network, *Nature* **513**, 202–209 (2014).
13. D. M. Hyman *et al.*, *N. Engl. J. Med.* **373**, 726–736 (2015).
14. C. J. Sherr, D. Beach, G. I. Shapiro, *Cancer Discov.* **6**, 353–367 (2016).
15. E. C. Smyth *et al.*, *Ann. Oncol.* **27**, v38–v49 (2016).
16. L. M. Zasadil *et al.*, *Sci. Transl. Med.* **6**, 229ra43 (2014).
17. C. S. Verissimo *et al.*, *eLife* **5**, e18489 (2016).
18. K. Khan *et al.*, *Gut* 10.1136/gutjnl-2017-314178 (2017).
19. S. Frentzas *et al.*, *Nat. Med.* **22**, 1294–1302 (2016).
20. R. J. Mayer *et al.*, *N. Engl. J. Med.* **372**, 1909–1919 (2015).
21. Y. Kuboki *et al.*, *Lancet Oncol.* **18**, 1172–1181 (2017).

FORCE**FOR**SCIENCE.ORG

STAND TOGETHER
Be a Force for Science

GET THE FACTS

Understand the science behind the issues that matter.

FOLLOW AAAS ADVOCACY

Champion public discussion and evidence-based policy.

TAKE ACTION

Learn ways you can become an advocate and stand up for science.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

AAAS

CANCER

The chromatin accessibility landscape of primary human cancers

M. Ryan Corces[1]*, Jeffrey M. Granja[1,2,3]*, Shadi Shams[1], Bryan H. Louie[1], Jose A. Seoane[2,4,5], Wanding Zhou[6], Tiago C. Silva[7,8], Clarice Groeneveld[9], Christopher K. Wong[10], Seung Woo Cho[1], Ansuman T. Satpathy[1], Maxwell R. Mumbach[1,2], Katherine A. Hoadley[11], A. Gordon Robertson[12], Nathan C. Sheffield[13], Ina Felau[14], Mauro A. A. Castro[9], Benjamin P. Berman[7], Louis M. Staudt[14], Jean C. Zenklusen[14], Peter W. Laird[6], Christina Curtis[2,4,5], The Cancer Genome Atlas Analysis Network†, William J. Greenleaf[1,2,3,15,16]‡, Howard Y. Chang[1,2,17,18]‡

We present the genome-wide chromatin accessibility profiles of 410 tumor samples spanning 23 cancer types from The Cancer Genome Atlas (TCGA). We identify 562,709 transposase-accessible DNA elements that substantially extend the compendium of known cis-regulatory elements. Integration of ATAC-seq (the assay for transposase-accessible chromatin using sequencing) with TCGA multi-omic data identifies a large number of putative distal enhancers that distinguish molecular subtypes of cancers, uncovers specific driving transcription factors via protein-DNA footprints, and nominates long-range gene-regulatory interactions in cancer. These data reveal genetic risk loci of cancer predisposition as active DNA regulatory elements in cancer, identify gene-regulatory interactions underlying cancer immune evasion, and pinpoint noncoding mutations that drive enhancer activation and may affect patient survival. These results suggest a systematic approach to understanding the noncoding genome in cancer to advance diagnosis and therapy.

Cancer is a highly heterogeneous group of diseases, with each tumor type exhibiting distinct clinical features, patient outcomes, and therapeutic responses. The Cancer Genome Atlas (TCGA) was established to characterize this heterogeneity and understand the molecular underpinnings of cancer (*1*). Through large-scale genomic and molecular analyses, TCGA has revealed an exquisite diversity of genomic aberrations, altered transcriptional networks, and tumor subtypes that have engendered a more comprehensive understanding of disease etiologies and laid the foundations for new therapeutics and impactful clinical trials.

Work from TCGA and many others has demonstrated the importance of the epigenome to cancer initiation and progression (*2*). Profiling of cancer-specific coding mutations through whole-exome sequencing has identified prominent driver mutations in genes encoding chromatin remodeling enzymes and modifiers of DNA methylation. These mutations drive alterations in the epigenome which, in turn, can establish the dysregulated cellular phenotypes that have become known as the hallmarks of cancer (*3*). Although many principles of chromatin regulation have been elucidated in cultured cancer cells, epigenomic studies of primary tumors are especially valuable, capturing the genuine ecosystem of heterotypic tumor and stromal cell interactions and the impacts of factors in the tumor microenvironment such as hypoxia, acidosis, and matrix stiffness (*4*). TCGA has carried out targeted DNA methylation profiling of more than 10,000 samples and, more recently, whole-genome bisulfite sequencing (WGBS) of 39 TCGA tumor samples (*5*). This data-rich resource has identified cancer-specific differentially methylated regions, providing an unprecedented view of epigenetic heterogeneity in cancer. Integration of DNA methylation and additional TCGA data types has enabled the prediction of functional regulatory elements (*6*–*8*) and the identification of previously unknown cancer subtypes (*9*–*13*). Additional work has identified cancer-relevant variable enhancer loci by using histone modifications (*14*) and enhancer RNA sequencing (*15*). These studies represent, to date, the largest genome-wide epigenomic profiling efforts in primary human cancer samples.

Recently, the advent of the assay for transposase-accessible chromatin using sequencing (ATAC-seq) (*16*) has enabled the genome-wide profiling of chromatin accessibility in small quantities of frozen tissue (*17*). Because accessible chromatin is a hallmark of active DNA regulatory elements, ATAC-seq makes it possible to assess the gene regulatory landscape in primary human cancers. Combined with the richness of diverse, orthogonal data types in TCGA, the chromatin accessibility landscape in cancer provides a key link between inherited and somatic mutations, DNA methylation, long-range gene regulation, and, ultimately, gene expression changes that affect cancer prognosis and therapy.

Results

ATAC-seq in frozen human cancer samples is highly robust

We profiled the chromatin accessibility landscape for 23 types of primary human cancers, represented by 410 tumor samples derived from 404 donors from TCGA (protocol S1). These 23 cancer types are representative of the diversity of human cancers (Fig. 1A and data S1). From the 410 tumor samples, we generated technical replicates from 386 samples, yielding 796 genome-wide chromatin accessibility profiles (data S1). Given the size of this cohort, we first ensured that all generated ATAC-seq data could be uniquely mapped to the expected donor through comparison with single-nucleotide polymorphism (SNP) genotyping calls (fig. S1A). In all samples, the genotype from the ATAC-seq data generated in this study correlated most highly with previously published genotyping array data for the expected donor compared with that of all other 11,126 TCGA donors. All ATAC-seq data included in this study passed a minimum threshold of enrichment of signal over background (fig. S1, B to D, and data S1) with most samples showing a characteristic fragment size distribution with clear nucleosomal periodicity (fig. S1E). With this high-quality set of 410 tumor samples, we identified 562,709 reproducible (observed in more than one replicate) pan-cancer

[1]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Program in Biophysics, Stanford University School of Medicine, Stanford, CA 94305, USA. [4]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. [5]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA. [7]Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA. [8]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP CEP 14.040-905, Brazil. [9]Bioinformatics and Systems Biology Laboratory, Polytechnic Center, Federal University of Paraná, Curitiba, PR CEP 80.060-000, Brazil. [10]Department of Biomolecular Engineering, Center for Biomolecular Sciences and Engineering, University of California–Santa Cruz, Santa Cruz, CA 95064, USA. [11]Department of Genetics, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [12]Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada. [13]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA. [14]National Cancer Institute, NIH, Bethesda, MD 20892, USA. [15]Department of Applied Physics, Stanford University, Stanford, CA 94025, USA. [16]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [17]Program in Epithelial Biology, Stanford University, Stanford, CA 94305, USA. [18]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA.
*These authors contributed equally to this work. †The Cancer Genome Atlas Analysis Network collaborators and affiliations are listed in the supplementary materials.
‡Corresponding author. Email: howchang@stanford.edu (H.Y.C.); wjg@stanford.edu (W.J.G.)
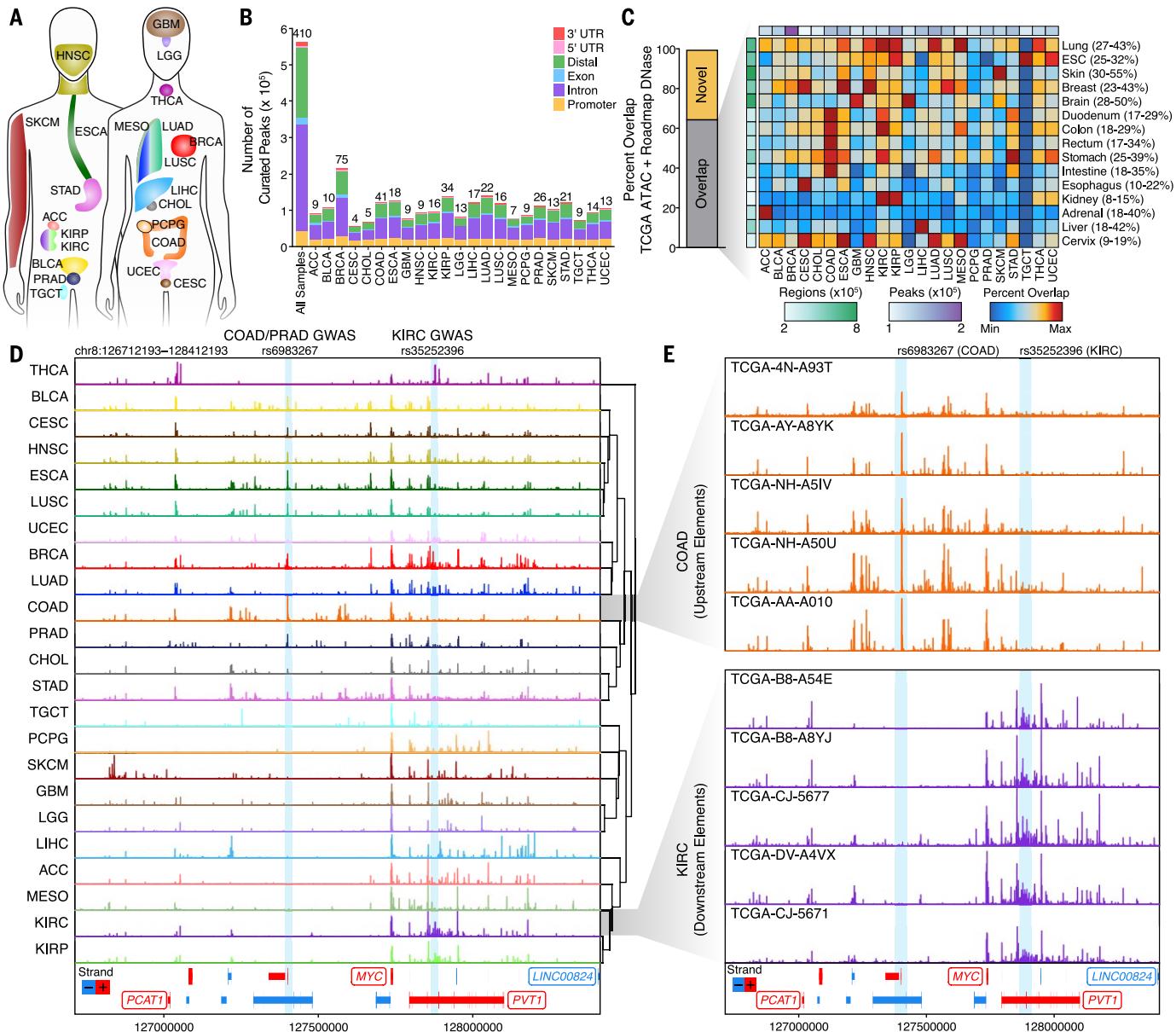
Fig. 1. Pan-cancer ATAC-seq of TCGA samples identifies diverse regulatory landscapes. (**A**) Diagram of the 23 cancer types profiled in this study. Colors are kept consistent throughout figures. (**B**) Pan-cancer peak calls from ATAC-seq data. Peak calls from each cancer type are shown individually in addition to the 562,709 peaks that represent the pan-cancer merged peak set. Color indicates the type of genomic region overlapped by the peak. The numbers shown above each bar represent the number of samples profiled for each cancer type. UTR, untranslated region. (**C**) Overlap of cancer type-specific ATAC-seq peaks with Roadmap DNase-seq peaks from various tissues and cell types. Left: The percent of ATAC-seq peaks that are overlapped by one or more Roadmap peaks. Right: A heatmap of the percent overlap observed for each ATAC-seq peak set within the Roadmap DNase-seq peak set. Colors are scaled according to the minimum and maximum overlaps, which are indicated numerically to the right of the DNase-seq peak set names. The total number of ATAC-seq peaks (white to purple) or Roadmap DNase-seq regions (white to green) are shown colorimetrically. (**D**) Normalized ATAC-seq sequencing tracks of all 23 cancer types at the *MYC* locus. Each track represents the average accessibility per 100-bp bin across all

replicates. Known GWAS SNPs rs6983267 (COAD, PRAD) and rs35252396 (KIRC) are highlighted with light blue shading. Region shown represents chromosome 8 (chr8):126712193 to 128412193. (**E**) Normalized ATAC-seq sequencing tracks of five different COAD samples (top, orange) and KIRC samples (bottom, purple) shown across the same *MYC* locus as in Fig. 1D. Known GWAS SNPs rs6983267 (COAD, PRAD) and rs35252396 (KIRC) are highlighted with light blue shading. Region shown represents chr8:126712193 to 128412193. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

peaks of chromatin accessibility (Fig. 1B and data S2). These peaks were identified using a normalized peak score metric to enable direct comparison of peaks across samples of unequal sequencing depth, with each cancer type having an average of 105,585 peaks (range 56,125 to 215,978; Fig. 1B and fig. S1F; see methods). Reproducibility within the pan-cancer peak set was high for technical replicates (different nuclei from the same tumor sample; fig. S1, G and H), intratumor replicates (different samples from the same tumor; fig. S1I), and intertumor replicates (tumor samples from different donors; fig. S1, J and K).

Cancer chromatin accessibility extends the dictionary of DNA regulatory elements

The pan-cancer and cancer type–specific peak sets generated in this study enabled quantification of the number of DNA regulatory elements identified. To do this, we compared the regions defined by our pan-cancer and cancer type–specific peak sets to the regions defined by the Roadmap Epigenomics Project deoxyribonuclease I hypersensitive sites sequencing (DNase-seq) studies (*18*), finding a median of 34.4% overlap between the cancer type–specific peak sets and the various Roadmap tissue-type peak sets, with the strongest overlap occurring in the expected combinations (Fig. 1C and data S3). In total, about 65% of the pan-cancer peaks identified in this study had overlap with previously observed regulatory elements, highlighting both the consistency of our results with published datasets and the large number of additional putative regulatory elements observed in this study (Fig. 1C). Given the extensive coverage of Roadmap DNase-seq studies in healthy tissues, our results suggested that the disease context of cancer unveils the activity of additional DNA regulatory elements. Moreover, overlap of the ATAC-seq–defined DNA regulatory elements with chromatin immunoprecipitation sequencing (ChIP-seq)–defined ChromHMM regulatory states shows a strong enrichment of accessible chromatin sites in promoter and enhancer regions, as expected (fig. S1L). Although we profiled many samples in some cancer types [i.e., breast invasive carcinoma (BRCA), 75 tumor samples], we profiled fewer samples in multiple other cancer types (i.e., cervical squamous cell carcinoma, four tumor samples) (Fig. 1B). By estimating the number of unique peaks added with each additional sample, we found that cancer types have an estimated average of 169,822 total peaks (range 97,995 to 309,313) at saturation (fig. S1, M and N, and data S3), suggesting that profiling of additional samples of each cancer type would further expand the repertoire of regulatory elements.

Noncoding DNA elements reveal distinct cancer gene regulation and genetic risks

The *MYC* proto-oncogene locus provides a prime illustration of the diversity of the chromatin accessibility landscape across cancer types. *MYC* is embedded in a region with multiple DNA regulatory elements and noncoding transcripts that regulate *MYC* in a tissue-specific fashion (*19*). We observed sufficient diversity in the chromatin accessibility landscape of the *MYC* locus to enable clustering of cancer types into two primary categories: (i) cancer types with extensive chromatin accessibility at 5′ and 3′ DNA elements, such as colon adenocarcinoma (COAD), and (ii) cancer types with chromatin accessibility primarily at 3′ regulatory elements, such as kidney renal clear cell carcinoma (KIRC) (Fig. 1D). This trend is consistent across different samples of the same cancer type, as shown for COAD and KIRC (Fig. 1E) and is similar to the regulation observed in the *HOXD* locus (*20*).

Genome-wide association studies (GWAS) have identified numerous inherited risk loci for cancer susceptibility. However, many of these SNPs reside in the noncoding genome within known DNA regulatory elements. In the *MYC* locus, we identify known sites of chromatin accessibility, including peaks surrounding functionally validated GWAS cancer susceptibility SNPs (rs6983267 and rs35252396; Fig. 1, D and E). SNP rs6983267 is associated with increased susceptibility to colon adenocarcinoma and prostate adenocarcinoma (PRAD) (*21–23*), consistent with the presence of focal chromatin accessibility in these cancer types. However, SNP rs6983267 has not been previously associated with breast cancer or any squamous tumor types, which also have strong chromatin accessibility at this regulatory element in our ATAC-seq data (Fig. 1D). Similarly, SNP rs35252396 has been associated with KIRC and, in our data, shows strong accessibility in samples from kidney cancer types as well as breast and thyroid carcinoma, suggesting a potential role for these SNPs in previously unappreciated cancer contexts.

To visualize global patterns from our diverse ATAC-seq datasets, we performed Pearson correlation hierarchal clustering on distal and promoter elements (Fig. 2A). We found that distal elements exhibited a greater specificity and wider dynamic range of activity in association with cancer types, whereas promoter element accessibility was less cancer type–specific and showed similar patterns of correlation to global gene expression, as measured by RNA-seq (Fig. 2A). This functional specificity of distal regulatory elements was also previously observed in healthy tissues and in development (*24, 25*). Using *t*-distributed stochastic neighbor embedding (*26*) (t-SNE; Fig. 2B) and density clustering (*27*) (fig. S2A), we identified 18 distinct clusters, which we labeled based on the observed cancer-type enrichment (fig. S2B and data S3). We found strong concordance between this ATAC-seq–based clustering and the published multiomic iCluster scheme using TCGA mRNA-seq, microRNA (miRNA)–seq, DNA methylation, reverse-phase protein array (RPPA), and DNA copy number data (*28*) (Fig. 2, C and D). Comparing this clustering scheme to other TCGA-based clustering schemes, we observed the strongest concordance of our ATAC-seq clustering scheme with mRNA and cancer type (Fig. 2E).

This is consistent with the connection of chromatin accessibility to transcriptional output and the observation that ATAC-seq is strongly cell type–specific. Multiple observations can be made from these clusters: (i) Some cancer types split into two distinct clusters such as breast cancer (i.e., basal and nonbasal) and esophageal cancer (i.e., squamous and adenocarcinoma), (ii) cancer samples derived from the same tissue type often group together [i.e., kidney renal papillary cell carcinoma (KIRP) and KIRC], and (iii) some cancers group together across tissues as observed for squamous cell types (Fig. 3A and fig. S2B).

Cluster-specific regulatory landscapes identify patterns of transcription factor usage and DNA hypomethylation

Grouping of samples into defined clusters enables the determination of patterns in chromatin accessibility that are unique to each cluster. Using a framework that we term "distal binarization," we identified the distal regulatory elements that are accessible only in a single cluster or small group of clusters (Fig. 3B, fig. S2C, and data S4). Of the 516,927 pan-cancer distal elements, 203,260 were found to be highly accessible in a single cluster or group of clusters (up to four clusters). These cluster-specific peak sets are enriched for motifs of transcription factors (TFs) with correlated gene expression that are known to be important for cancer and tissue identity (Fig. 3C, fig. S2D, and data S4). These include the androgen receptor (AR) in prostate cancer, forkhead box A1 (FOXA1) in nonbasal breast cancer, and melanogenesis-associated transcription factor (MITF) in melanoma. Moreover, these cluster-specific peak sets are enriched for known GWAS SNPs that are associated with cancers of the corresponding type (fig. S2E and data S5), highlighting that cancer-related GWAS SNPs tend to be located within or near cancer type–specific regulatory elements. The concordance of GWAS risk loci and cancer chromatin state has often been evaluated using cancer cell lines in the past, and our work provides a foundational map to evaluate noncoding GWAS SNPs in primary human cancers.

Consistent with published reports (*12, 18, 29, 30*), the degree of DNA methylation was anticorrelated with chromatin accessibility at regulatory elements, and regions lacking chromatin accessibility were more frequently methylated (fig. S2F). In particular, cluster-specific peak sets are hypomethylated in the relevant cancer types, though frequently methylated in other cancer types that lack accessibility in those peaks (fig. S2G). Consistent with these observations, which are based on DNA methylation array data, we see a strong depletion of DNA methylation at the center of both distal peaks and promoter peaks in a single patient profiled by WGBS (fig. S2H) (*5*). In our analysis of methylation levels within cluster-specific peak sets, we also identified a subgroup of brain cancers that exhibits DNA hypermethylation of peaks specific to nonbrain cancers (fig. S2G), likely caused by mutations in genes

that affect DNA methylation, such as isocitrate dehydrogenase 1 (*IDH1*) (fig. S3A). Similarly, we found that the subset of testicular germ cell tumors that are seminomas show a pattern of genome-wide DNA hypomethylation, consistent with a published report (*31*) (fig. S3B). Thus, a small number of TFs dominate the cis-regulatory landscape in each cancer type. These TFs are often the known key drivers of the respective

cancer or tissue type, and TF occupancy is associated with, and possibly causes, DNA hypomethylation of the corresponding DNA elements in cancer.
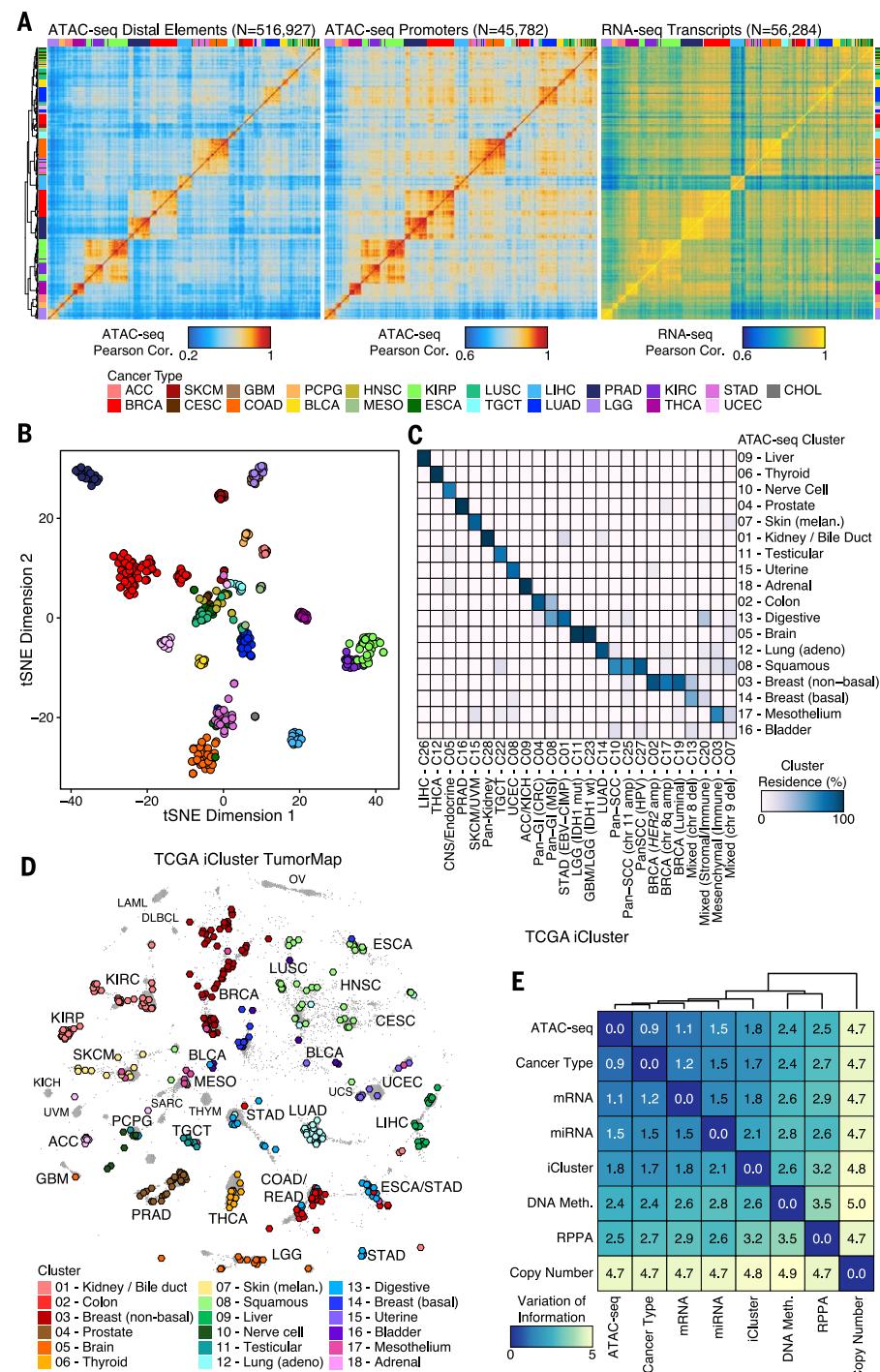
De novo identification of cancer subtypes from ATAC-seq data

Given the richness of the chromatin accessibility landscape, we explored the capacity of

ATAC-seq data to define molecular subtypes of cancer de novo. This analysis was limited to cancer types with sufficient available donors: BRCA (N = 74), PRAD (N = 26), and KIRP (N = 34). In KIRP, a gap statistic identified three distinct subgroups that are clearly separable by the first two principal components (Fig. 3D). The smallest of these subgroups contains four donors with very clear differences in ATAC-seq

Fig. 2. Chromatin accessibility profiles reveal distinct molecular subtypes of cancers. (**A**) Pearson correlation heatmaps of ATAC-seq distal elements (left), ATAC-seq promoters (middle), and RNA-seq of all genes (right). Clustering orientation is dictated by the ATAC-seq distal element accessibility, and all other heatmaps use this same clustering orientation. Color scale values vary between heatmaps. Promoter peaks are defined as occurring between −1000 and +100 bp of a transcriptional start site. Distal peaks are all nonpromoter peaks. The total number of features (N) used for correlation is indicated above each Pearson correlation heatmap. (**B**) Unsupervised t-SNE on the top 50 principal components for the 250,000 most variable peaks across all cancer types. Each dot represents the merge of all technical replicates from a given sample. Color represents the cancer type shown above the plot. (**C**) Cluster residence heatmap showing the percent of each TCGA iCluster that overlaps with each ATAC-seq–based cluster. (**D**) ATAC-seq t-SNE clusters shown on the PanCanAtlas iCluster TumorMap. Each hexagon represents a cancer patient sample, and the positions of the hexagons are computed from the similarity of samples in the iCluster latent space. The color and larger size of the hexagon indicates the ATAC-seq cluster assignment. Samples that were not included in the ATAC-seq analysis are represented by smaller gray hexagons. The text labels indicate the cancer disease type. (**E**) Variation of information analysis of clustering schemes derived by using various data types from TCGA.
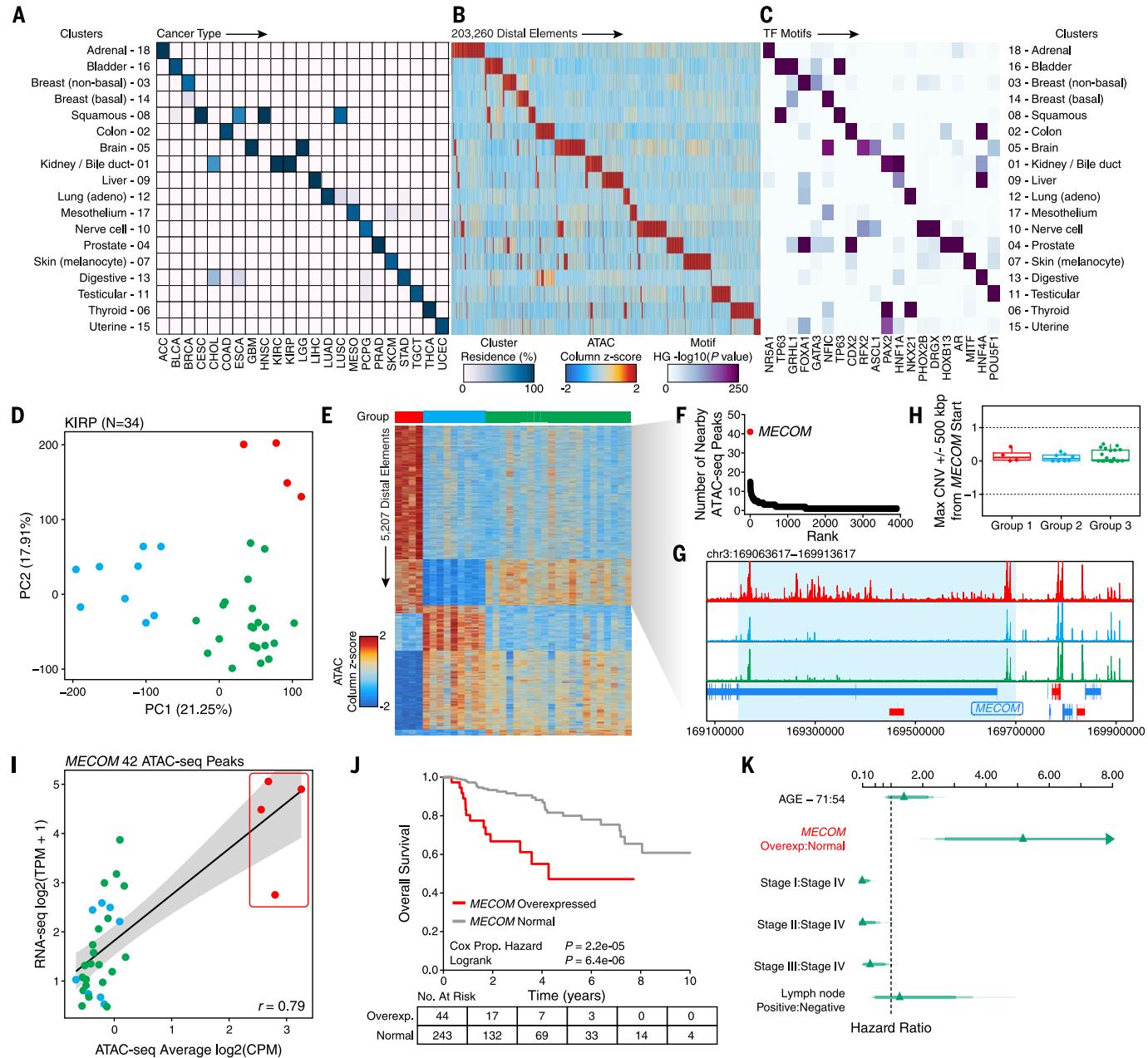
Fig. 3. ATAC-seq clusters cancer samples to show cancer- and tissue-specific drivers. (**A**) Cluster residence heatmap showing the percent of samples from a given cancer type that reside within each of the 18 annotated ATAC-seq clusters. (**B**) Heatmap showing the ATAC-seq accessibility at distal elements (*N* = 203,260) identified to be cluster-specific by distal binarization. (**C**) Enrichment of TF motifs in peak sets identified in Fig. 3B. Enrichment is determined by a hypergeometric (HG) test −log$_{10}$(*P* value) of the motif's representation within the cluster-specific peaks compared to the pan-cancer peak set. Transcription factors shown represent a manually trimmed set of factors whose expression is highly correlated (*r* > 0.4) with the accessibility of the corresponding motif. Color represents the −log$_{10}$(*P* value) of the hypergeometric test. (**D**) Principal component analysis of the top 25,000 distal ATAC-seq peaks within the KIRP cohort (*N* = 34 samples). Each dot represents an individual sample. The color of the dots represents *K*-means clustering (*K* = 3 by gap statistic). (**E**) Distal binarization analysis based on the three *K*-means–defined groups identified and shown (by color) in

Fig. 3D. (**F**) Dot plot showing the number of nearby ATAC-seq peaks per gene from the group 1 distal binarization. Each dot represents a different gene. The *MECOM* gene (also called *EVI1*) is highlighted in red. (**G**) Normalized average sequencing tracks of *K*-means–defined groups 1, 2, and 3 at the *MECOM* locus. Peaks specific to group 1 are highlighted by light blue shading. (**H**) DNA copy number data at the *MECOM* locus in the three *K*-means–defined groups. Each dot represents an individual sample. CNV, copy number variation. (**I**) Average chromatin accessibility at peaks near the *MECOM* gene (*N* = 42 peaks) and RNA-seq gene expression of *MECOM* in KIRP samples (*N* = 34 samples). Each dot represents an individual donor. Dots are colored according to the clustering group colors shown in Fig. 3D. CPM, counts per million. (**J**) Kaplan-Meier analysis of overall survival of all KIRP donors in TCGA (*N* = 287) stratified by *MECOM* overexpressed (*N* = 44) and normal *MECOM* expression (*N* = 243). (**K**) Hazard plot of risk of dying from KIRP based on multiple covariates, including *MECOM* expression (hazard ratio = 5.2, 95% confidence interval = 2.4 to 11.0). Lines represent 95% confidence intervals.

accessibility identified by distal binarization (red coloring in Fig. 3E). Within the set of regulatory elements that are specific to this subgroup, we found 42 ATAC-seq peaks near the MDS1 and EVI1 complex locus (*MECOM*) gene (Fig. 3, F and G). Notably, the high chromatin accessibility of these *MECOM* peaks is not related to copy number amplification, as determined by DNA copy number array data (Fig. 3H). The expression of the *MECOM* gene is highly correlated with the mean ATAC-seq accessibility at these 42 ATAC-seq peaks [correlation coefficient (r) = 0.79, Fig. 3I]. Additionally, overexpression of *MECOM* is significantly associated with poorer overall survival across all available KIRP data from TCGA ($P = 2.2 \times 10^{-5}$, Cox proportional hazard test, Fig. 3J) with a hazard ratio of 5.2 (95% confidence interval = 2.4 to 11.0). This association is more substantial than lymph node status or patient age and is independent of cancer stage (Fig. 3K), indicating a potential prognostic role for these findings. Importantly, *MECOM* overexpression is not readily explained by any previously identified subgroups of KIRP, including subgroups with a CpG island methylator phenotype or mutations in the gene encoding fumarate hydratase, which have also been shown to confer poor overall survival (*13*). These results suggest that *MECOM* activation in KIRP identifies a previously unappreciated subgroup of patients with adverse outcomes, a finding that was uncovered by notable changes in the chromatin accessibility landscape of these samples.

Similarly, we found multiple distinct subgroups of PRAD and BRCA based on *K*-means clustering of the top 25,000 variable distal ATAC-seq peaks (fig. S3, C and D). In PRAD, these include subgroups driven by activity of AR, tumor protein P63 (*TP63*), and forkhead box–family TFs (fig. S3C). From an unsupervised analysis of breast cancer, we identified motifs of known TF drivers of luminal subtype identity, such as GATA binding protein 3 (GATA3) and FOXA1, as being enriched in the peak clusters specific to a subset of luminal samples (clusters 3 and 4, fig. S3D). We also identified a potential role for grainyhead-like (GRHL) TF motifs in basal breast cancer (*32*) (cluster 1, fig. S3D) and an overlapping role for nuclear factor I (NFI) in both basal and luminal A breast cancer (cluster 2, fig. S3D). Additionally, ATAC-seq data can be used to identify regions of copy number amplification de novo (*33*), enabling the classification of *HER2*-amplified cases of breast cancer (fig. S3, E to G).

Footprinting analysis defines TF activities in cancer

The high sequencing depth of the ATAC-seq data generated in this study (median of 56.7 million unique reads per technical replicate) enabled the profiling of TF occupancy at base-pair resolution through TF footprinting. TF binding to DNA protects the protein-DNA binding site from transposition while the displacement or depletion of one or more nucleosomes creates high DNA accessibility in the immediate flanking se-

quence. Collectively, these phenomena are referred to as the TF footprint. To characterize TF footprints, we adapted a recent approach (*34*) that quantifies the "flanking accessibility," a measure of the accessibility of the DNA adjacent to a TF motif, and "footprint depth," a measure of the relative protection of the motif site from transposition (Fig. 4A and data S6). To calculate these variables, we aggregated all insertions relative to the TF motif center, genome-wide (fig. S4A). To attempt to account for known Tn5 transposase insertion bias, we computed the hexamer frequency centered at Tn5 insertions and normalized for the expected bias at each position relative to the motif center (*34*) (see methods for potential limitations). Depending on the binding properties of a TF and its ability to affect local chromatin accessibility, changes in these properties would be detectable through this approach genome-wide (fig. S4, B and C). ChromVAR (*35*), a similar genome-wide approach which assesses the ability of a TF to affect flanking accessibility, identified a highly overlapping list of TFs (fig. S4D).

To uncover transcriptionally driven TF binding patterns, we correlated the RNA-seq gene expression of a given TF to its corresponding footprint depth and estimated flanking accessibility (data S6). A factor whose expression is sufficient to generate robust DNA binding would have a footprint depth and flanking accessibility that are significantly correlated to its gene expression [false discovery rate (FDR) < 0.1, purple dots in Fig. 4B], such as TP63 (Fig. 4, C and D) or NK2 homeobox 1 (NKX2-1) (Fig. 4, E and F). Increases in flanking accessibility and decreases in footprint depth are likewise accompanied by decreases in methylation (bottom of Fig. 4, D and F), consistent with the hypothesis that methylated DNA is less likely to be bound by TFs (*36*). Although footprint depth and flanking accessibility are often correlated, their divergence can suggest the modes of TF-DNA interaction. For example, factors whose expression is sufficient to cause opening of chromatin around the motif site but not to protect the motif site from transposition would be expected to only exhibit a significant correlation between gene expression and flanking accessibility (blue dots in Fig. 4B). This pattern of correlation could be caused by effects such as rapid TF off rates or low occupancy (fig. S4, E and F). Conversely, a small number of TFs have expression that is only significantly correlated with footprint depth (red dots in Fig. 4B). Though likewise rare, we also identified potential negative regulators whose expression is inversely correlated to gain of flanking accessibility and loss of footprint depth, such as the cut-like homeobox 1 (CUX1) TF (*37*) (Fig. 4B and fig. S4, G and H). This is the expected behavior of repressive TFs that bind DNA and lead to compaction of the neighboring sequence. These results predicted dozens of positive and negative regulators whose expression is strongly correlated with chromatin accessibility patterns near to their corresponding motif (fig. S4I and data S6). Overall, our
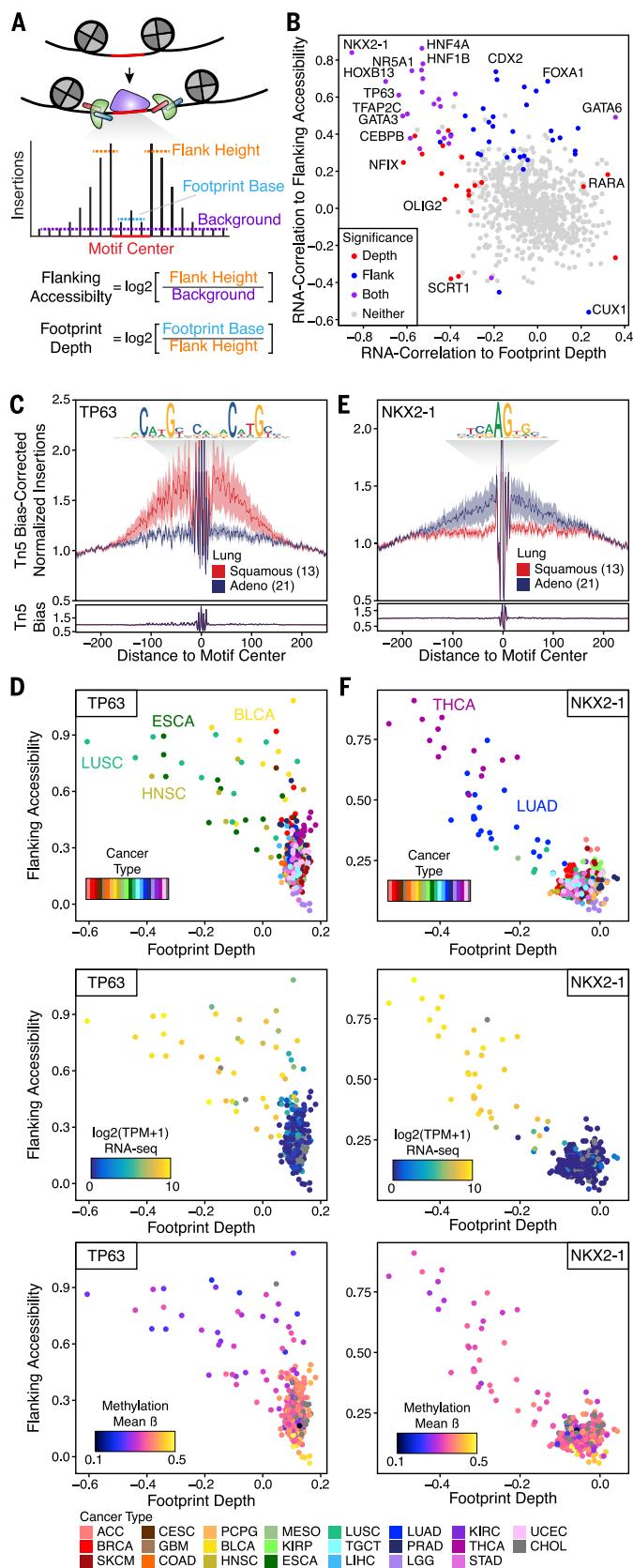
footprinting analysis identified putative TFs with activities correlated with gene expression.

Linking of DNA regulatory elements to genes predicts interactions relevant to cancer biology

The breadth and depth of this sequencing study enabled a robust association of ATAC-seq peaks with the genes that they are predicted to regulate. To do this, we implemented a strategy based on the correlation of ATAC-seq accessibility and gene expression across all samples (Fig. 5A, N = 373 with matched RNA-seq and ATAC-seq). Because promoter capture Hi-C data suggested that >75% of three-dimensional (3D) promoter-based interactions occur within a 500–kilobase pair (kbp) distance (*38*), we restricted the length scale of this analysis to 500 kbp to avoid spurious predictions. Using a conservative FDR cutoff of 0.01, we identified 81,323 unique links between distal ATAC-seq peaks and genes (Fig. 5B and data S7). Some of these links are driven by correlation across many cancer types (Fig. 5, C to E), whereas 70% are strongly driven by one cluster (Fig. 5F and data S7). To derive a final list of peak-to-gene links (Fig. 5B), putative links were filtered against (i) links whose correlation is strongly driven by DNA copy number amplification ("CNA"; fig. S5, A and B), (ii) regions with broad and high local correlation ("diffuse"; fig. S5, B and C), and (iii) links involving an ATAC-seq peak that overlaps the promoter of any gene (Fig. 5G). As expected, the histogram of distances between a peak and its target gene decays sharply with distance (*39*) (Fig. 5H). The expression of most genes is correlated with the activity of fewer than five different peaks (Fig. 5I), whereas most peaks are predicted to interact with a single gene (Fig. 5J). Additionally, this analysis found that only 24% of predicted links occur between an ATAC-seq peak and the nearest gene, indicating that the majority of predicted interactions skip over one or more genes and would not be possible to predict from primary sequence alone (Fig. 5K). In total, we predicted at least one peak-to-gene link for 8552 protein-coding genes, accounting for nearly half of all protein-coding genes in the human genome, including 48% of the curated Catalogue of Somatic Mutations in Cancer (COSMIC) cancer-relevant genes (data S7).

In addition to predicting peak-to-gene links across cancer types, we also predicted peak-to-gene links within breast cancer (N = 74 donors), identifying 9711 unique peak-to-gene links (fig. S5D and data S7). Of these links, 36% were also identified in our analysis of all cancer types (fig. S5E). Particularly important in these BRCA-specific links was the contribution of recurrent DNA CNA as a strong driver for spurious peak-to-gene correlation (Fig. 5G). These false-positive associations were removed through the use of published TCGA DNA copy number array data and a local correlation correction model, as mentioned above (see methods). The final predicted BRCA-specific links follow a similar distance

Fig. 4. Footprinting analysis identifies distinct TF activities in cancer.

Fig. 4. Footprinting analysis identifies distinct TF activities in cancer. (**A**) Schematic illustrating the dynamics of TF binding (purple) and Tn5 insertion (green). (**B**) Classification of TFs by the correlation of their RNA expression to the footprint depth and flanking accessibility of their motifs. Color represents whether the depth (red), flank (blue), or both (purple) are significantly correlated to TF expression below an FDR cutoff of 0.1. Each dot represents an individual deduplicated TF motif (see methods). (**C**) TF footprinting of the TP63 motif (CIS-BP M2321_1.02) in lung cancer samples from the squamous (cluster 8) or adenocarcinoma (cluster 12) subtype. The Tn5 insertion bias track of TP63 motifs is shown below. (**D**) Dot plots showing the footprint depth and flanking accessibility of TP63 motifs across all lung cancer samples studied. Each dot represents a unique sample. Color represents cancer type (top), RNA-seq gene expression (middle), or methylation beta value (bottom). Samples without matching RNA or methylation data are shown in gray. (**E**) TF footprinting of the NKX2-1 motif (CIS-BP M6374_1.02) in lung cancer samples from the squamous (cluster 8) and adenocarcinoma (cluster 12) subtype. The Tn5 insertion bias track of NKX2-1 motifs is shown below. (**F**) Dot plots showing the footprint depth and

flanking accessibility of NKX2-1 motifs across all lung cancer samples studied. Each dot represents a unique sample. Color represents cancer type (top), RNA-seq gene expression (middle), or methylation beta value (bottom). Samples without matching RNA or methylation data are shown in gray.

distribution and peak-to-gene linking specificity as observed in the pan-cancer predicted links (fig. S5, F to I).

Many of these predicted peak-to-gene links occur in clusters where multiple nearby peaks are predicted to be linked to the same gene, indicating that these clusters of peak-to-gene links may function as part of a single regulatory unit or enhancer. Extending the width of the linked ATAC-seq peaks to 1500 bp allows for joining of these peaks into defined merged putative enhancer units (fig. S5J). This resulted in a total of 58,092 pan-cancer and 7622 BRCA-specific enhancer-to-gene links (data S7).

Validation and utility of predicted links between distal elements and genes

To verify a regulatory interaction for the predicted peak-to-gene links, we used a CRISPR interference (CRISPRi) (40) strategy using a catalytically dead Cas9 (dCas9) fused to a Kruppel-associated box (KRAB) domain, which mediates focal heterochromatin formation and functional silencing of noncoding DNA regulatory elements (Fig. 6A). In this way, targeting the distal peak region of a predicted peak-to-gene link would be expected to cause a decrease in the expression of the linked gene, located tens to hundreds of kilobases away. CRISPRi of a predicted distal regulatory element linked to *BCL2* (164 kbp, Fig. 5C) led to a significant reduction in *BCL2* gene expression in the luminal-like breast cancer MCF7 cell line but not in the basal-like MDA-MB-231 cell line (Fig. 6B), consistent with the role of *BCL2* as a luminal-specific survival factor (41). Similarly, CRISPRi of a distal regulatory element linked to the *SRC* oncogene (−49 kbp, Fig. 5D) led to a significant reduction in gene expression in both MCF7 cells and MDA-MB-231 cells (Fig. 6B). On a genome-wide scale, the predicted BRCA-specific peak-to-gene links show a strong enrichment in 3D chromosome conformation data from MDA-MB-231 cells (42), providing further support for our link prediction strategy (Fig. 6C). Moreover, we found that, of the peak-to-gene links predicted from BRCA ATAC-seq data that are also associated with a DNA methylation array CpG probe, 35% overlap with links predicted jointly from DNA methylation array and RNA-seq data in an ELMER analysis (8, 43) of the complete TCGA BRCA dataset (N = 858 tumors) (P << 0.001; Fig. 6D, fig. S6A, and data S8). These overlaps contain many luminal-specific and basal-specific links (fig. S6A), with a clear delineation between luminal (fig. S6B) and basal (fig. S6C) breast cancer samples. Integrating WGBS and ATAC-seq demonstrated the dynamics of methylation and chromatin accessibility and the overlap of predicted interactions at the non-basal *FOXA1* and basal forkhead box C1 (*FOXC1*) loci (fig. S6, D and E).

Similarly, previous work has leveraged TCGA RNA-seq data to infer transcriptional networks that consist of regulons, each of which is based on a TF regulator and its associated positive and negative target genes (fig. S7A) (44). For each

regulon, every donor in the cohort can be assigned a positive, undefined, or negative regulon activity as measured by a differential enrichment score (dES) (45). Certain patterns of chromatin accessibility are expected on the basis of the target gene set and dES status of the donor (fig. S7B). For example, in donors with positive dES, chromatin at sites linked to positive target genes should be more accessible, whereas chromatin at sites linked to negative targets should be less accessible (fig. S7B). Examination of the estrogen receptor 1 (ESR1) regulon in the 74 BRCA donors profiled in this study identified 482 ATAC-seq distal peak-to-gene links corresponding to 124 ESR1 target genes (fig. S7C and data S8). Accessibility at these peaks is strongly concordant with expectations, further supporting the predicted links ($P < 1 \times 10^{-20}$, fig. S7D). Examination of this regulon across all TCGA

BRCA donors (N = 1082) showed a significant difference in overall survival between ESR1 dES-positive and -negative samples (fig. S7, E and F).

Together, pan-cancer and BRCA-specific peak-to-gene links further informed cancer-related GWAS polymorphisms, allowing the linkage of SNPs to putative gene targets with about 65% of all GWAS polymorphisms targeting a gene other than the closest gene on the linear genome (data S5). SNPs falling within peak-to-gene links were predicted to act on important cancer-related genes, including master regulators of cancer and tissue identity such as NKX2-1 (fig. S7G) and TP63 (fig. S7H). Focusing specifically on the BRCA peak-to-gene links for which published 3D chromosome conformation data are available, we found clear examples of GWAS SNPs interacting with distant, non-neighboring genes, such as OSR1 (Fig. 6E and fig. S7I). More-

over, overlapping of the pan-cancer and breast cancer–specific peak-to-gene links with expression quantitative trait loci (eQTLs, where genetic variation at noncoding elements is associated with gene expression differences) from the Genotype-Tissue Expression (GTEx) project showed significant overlap in almost all comparisons (N = 44 of 48 comparisons) (fig. S7J and data S5). These results underscored our ability to use these predicted peak-to-gene links to generate key insights into published data and inform poorly understood aspects of cancer biology.

Identification of DNA regulatory elements related to immunological response to cancer

Of particular interest to current cancer therapy, immune infiltrates represent a substantial
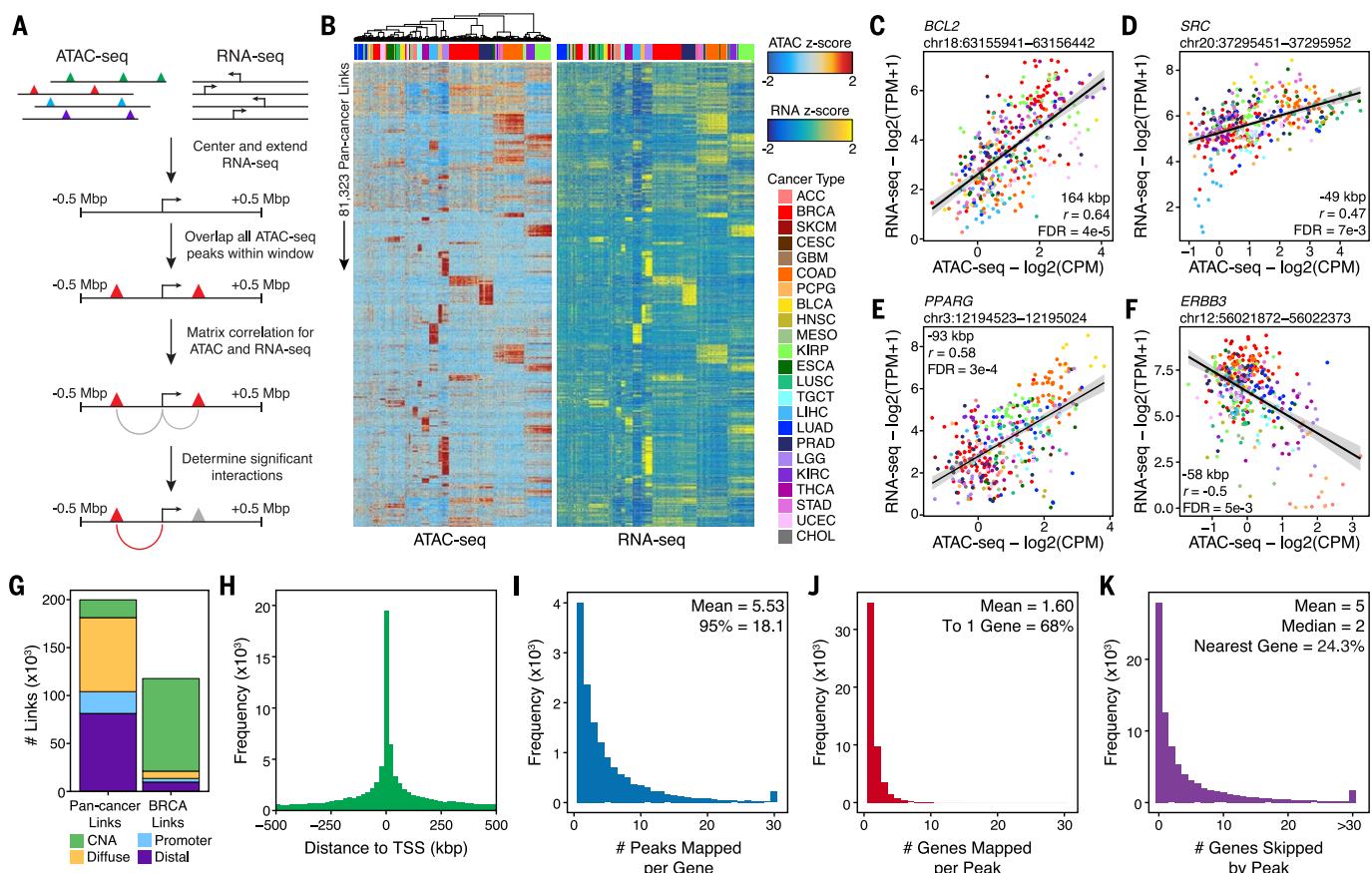
Fig. 5. In silico linking of ATAC-seq peaks to genes. (**A**) Schematic of the in silico approach used to link ATAC-seq peaks in distal noncoding DNA elements to genes via correlation of chromatin accessibility and RNA expression. (**B**) Heatmap representation of the 81,323 pan-cancer peak-to-gene links predicted. Each row represents an individual link between one ATAC-seq peak and one gene. Color represents the relative ATAC-seq accessibility (left) or RNA-seq gene expression (right) for each link as a z-score. (**C**) Dot plot of the ATAC-seq accessibility and RNA-seq gene expression of a peak-to-gene link located 164 kbp away from the transcription start site of the BCL2 gene (peak 498895) that is predicted to regulate its expression. Color represents the cancer type. Each dot represents an individual sample. (**D**) Same as in Fig. 5C but for a peak that is located 49 kbp away from the SRC gene (peak 525295). (**E**) Same

as in Fig. 5C but for a peak that is located 93 kbp away from the PPARG gene (peak 98874). (**F**) Same as in Fig. 5C but for a peak that is located 58 kbp away from the ERBB3 gene (peak 381116). (**G**) Bar plot showing the number of predicted links that were filtered for various reasons. First, regions whose correlation is driven by DNA copy number amplification were excluded ("CNA"). Next, regions of high local correlation were filtered out ("Diffuse"). Lastly, peak-to-gene links where the peak overlapped a promoter region were excluded ("Promoter"). The remaining links ("Distal") are used in downstream analyses. (**H**) Distribution of the distance of each peak to the transcription start site (TSS) of the linked gene. (**I**) Distribution of the number of peaks linked per gene. (**J**) Distribution of the number of genes linked per peak. (**K**) Distribution of the number of genes "skipped" by a peak to reach its predicted linked gene.

contribution to the overall tumor composition in solid tumors (46–48). We reasoned that infiltrating immune cells could contribute to our ATAC-seq data, both through actions on tumor cells and through increased chromatin accessibility at known immune-specific regulatory elements. Leveraging published ATAC-seq datasets from the human hematopoietic system (25) and data generated here from human dendritic cell subsets (Fig. 6F), we characterized each of our linked peaks by comparing its accessibility in immune cell types to its accessibility in bulk cancer samples (Fig. 6G). We reasoned that peaks that are more accessible in immune cells compared with our cancer cohort might be generated from immune cells associated with the tumor tissue (Fig. 6G). Additionally, we correlated each linked peak to the cytolytic activity score (49) of the tumor. The cytolytic activity score is based on the log-average gene expression of granzyme A and perforin 1, two CD8 T cell–specific markers. Linked peaks that exhibit high correlation to cytolytic activity might also be considered to be related to immune infiltration. Combining these two metrics, we identified peak-to-gene links expected to be highly relevant to immune infiltration, including links to genes relevant to antigen presentation and T cell response (Fig. 6H and data S9). The accessibility of these peak-to-gene links that were predicted to be immune-related is highly correlated with computationally predicted metrics of immune infiltration (46, 47) and inversely correlated with tumor purity (48) (Fig. 6I). One notable linked gene is programmed death ligand 1 (*PDL1*, also known as *CD274*), a key mediator of immune evasion by cancer and an important target for cancer immunotherapy. *PDL1* is linked to four putative distal regulatory elements that exhibit distinct chromatin accessibility across cancer types and are located as far as 43 kbp away from the *PDL1* transcription start site (Fig. 6, J and K). CRISPRi of each of these four putative *PDL1* regulatory elements significantly decreased, but did not abrogate, the expression of *PDL1* mRNA in at least one of the two breast cancer cell lines tested (MCF7 and MDA-MB-231 cells, Fig. 6L). These results support a model where the expression of *PDL1* is affected by the combined activity of multiple distal regulatory elements.

## Identification of cancer-relevant noncoding mutations
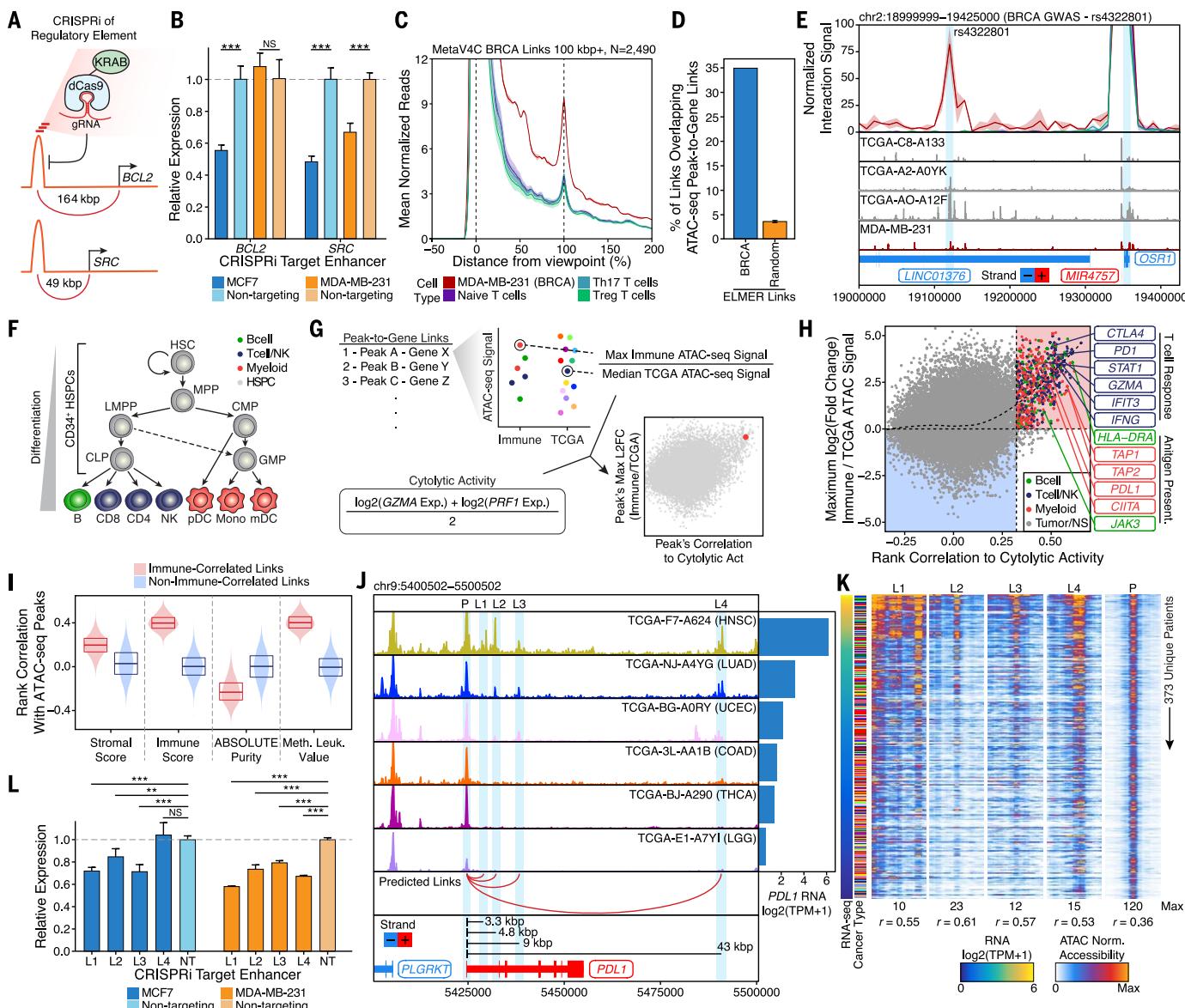
In addition to identifying gene regulatory interactions in cancer, ATAC-seq combined with whole-genome sequencing (WGS) can be used to identify regulatory mutations driving cancer initiation and progression. For example, if a noncoding somatic mutation causes the generation of a TF binding site, this mutation could lead to an increase in chromatin accessibility in cis and a concomitant increase in the observed frequency of the mutant allele in ATAC-seq as compared with that in WGS (Fig. 7A). Similarly, a mutation that inactivates a TF binding site can lead to a decrease in chromatin accessibility and a concomitant decrease in the observed frequency of the mutant allele. If such mutations in regulatory elements were to be functional in cancer, we might also expect that they increase or decrease chromatin accessibility beyond the expected distribution observed in nonmutated samples.

From the 404 donors profiled in this study, high-depth WGS data was available for 35 donors across 10 cancer types. These 35 donors had 374,705 called somatic mutations, with 32,696 falling within annotated ATAC-seq peaks and 2259 having at least 30 reads in both ATAC-seq and WGS data (data S10). Among these mutations were three separate occurrences of telomerase reverse transcriptase (*TERT*) gene promoter mutations (Fig. 7B), previously shown to generate de novo E26 transformation-specific (ETS) motif sites. ATAC-seq is especially well suited to identifying these *TERT* promoter mutations because the variant allele frequency is skewed owing to the increase in accessibility on the mutant allele (fig. S8A). Compared with the publicly available exome sequencing data from TCGA, where the *TERT* capture probes do not extend into the promoter region, ATAC-seq provided significantly

Fig. 6. Validation of long-range gene regulation of cancer in peak-to-gene links. (**A**) Schematic of CRISPRi experiments performed. Each experiment uses three guide RNAs (gRNAs) to target an individual peak. The effect of this perturbation on the expression of the linked gene is determined with quantitative polymerase chain reaction (qPCR). (**B**) Gene expression changes by qPCR after CRISPRi of peaks predicted to be linked to the *BCL2* (peak 498895) and *SRC* (peak 525295) genes in MCF7 and MDA-MB-231 cells. Error bars represent the standard deviation of four technical replicates. ***P < 0.001 and NS is not significant by two-tailed Student's *t* test. (**C**) Meta-virtual circular chromosome conformation capture (4C) plot of predicted BRCA-specific peak-to-gene links with distances greater than 100 kbp. HiChIP interaction frequency is shown for the MDA-MB-231 basal breast cancer cell line as well as multiple populations of primary T cells. Th17, T helper 17 cell; Treg, regulatory T cell. (**D**) Bar plot showing the overlap of predicted ATAC-seq–based peak-to-gene links and DNA methylation–based ELMER predicted probe-to-gene links in BRCA, as a percentage of all ATAC-seq–based peak-to-gene links with a peak overlapping a methylation probe. The percentage of peak-to-gene links overlapping an ELMER probe-to-gene link (34.9%) is compared to the overlap with 1000 sets of randomized ELMER probe-to-gene links (3.6 ± 0.6%, P << 0.001). (**E**) Virtual 4C plot of the peak-to-gene link between rs4322801 and the *OSR1* gene. Normalized HiChIP interaction signal is shown for the MDA-MB-231 basal breast cancer cell line as well as multiple populations of primary T cells using the colors shown in Fig. 6C. ATAC-seq sequencing tracks are shown below for four BRCA samples and MDA-MB-231 cells with increasing levels of *OSR1* gene expression. The rs4322801 SNP (left) and *OSR1* gene (right) are highlighted by light blue shading. Region shown represents chr2:18999999 to 19425000. (**F**) Diagram of the hematopoietic differentiation hierarchy with differentiated cells colored as either B cells (green), T cell or natural killer (Tcell/NK) cells (blue), or myeloid cells (red). HSC, hematopoietic stem cell; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; MPP, multipotent progenitor; CMP, common myeloid progenitor, GMP, granulocyte macrophage progenitor; HSPC, hematopoietic stem and progenitor cells; pDC, plasmacytoid dendritic cell; mDC, myeloid dendritic cell. (**G**) Schematic of the analysis shown in Fig. 6H. Peak-to-gene links are classified as related to immune infiltration if their accessibility is higher in immune cells than TCGA cancer samples and they are highly correlated to cytolytic activity. (**H**) Dot plot showing ATAC-seq peak-to-gene links with relevance to immune infiltration. Each dot represents an individual peak with a predicted gene link. Peaks that are related to immune cells have higher ATAC-seq accessibility in immune cell types compared to TCGA cancer samples. Peaks related to immune infiltration have a higher correlation to cytolytic activity. Color represents the cell type of the observation. The vertical dotted line represents the mean + 2.5 standard deviations above the mean for all ATAC-seq peak correlations to the cytolytic activity. The red shading indicates peak-to-gene links that are predicted to be related to immune infiltration. The blue shading indicates peak-to-gene links that are not predicted to be related to immune infiltration. NS, not significant. (**I**) Violin plots of the distribution of Spearman correlations across all peak-to-gene links predicted to be related to immune infiltration (red) or not (blue) with various metrics of tumor purity. (**J**) Normalized ATAC-seq sequencing tracks of the *PDL1* gene locus in six samples with variable levels of expression of the *PDL1* gene (right). Predicted links (red) are shown below for four peak-to-gene links (L1 to L4, peaks 293734, 293735, 293736, and 293740, respectively) to the promoter of *PDL1*. One of these peak-to-gene links (L2) overlaps an alternative start site for *PDL1* and was therefore labeled as a "promoter" peak during filtration. This peak-to-gene link was added to this analysis after manual observation. Region shown represents chr9:5400502 to 5500502. (**K**) Heatmap representation of the ATAC-seq chromatin accessibility of the 5000-bp region centered at each of the four peak-to-gene links shown in Fig. 6J. Each row represents a unique donor (N = 373) ranked by *PDL1* expression. The correlation of the chromatin accessibility of each peak with the expression of *PDL1* is shown below the plot. Color represents normalized accessibility. (**L**) Gene expression changes of the *PDL1* gene by qPCR after CRISPRi of peaks predicted to be linked to the *PDL1* gene in MCF7 and MDA-MB-231 cells. Error bars represent the standard deviation of four technical replicates. ***P < 0.0001 and **P < 0.05 by two-tailed Student's *t* test.

higher sequencing coverage of the *TERT* promoter locus per read sequenced, enabling a more robust classification of *TERT* promoter mutations ($P < 1 \times 10^{-7}$, fig. S8B). Of the three *TERT* promoter mutations identified in the subset of donors with matched WGS, one mutation, in particular, leads to a significant increase in accessibility compared to the other nonmutated members of that cancer type (FDR < 0.0001, blue dot in Fig. 7, B and C). As expected, this increase in *TERT* promoter accessibility is associated with a concomitant increase in *TERT* gene expression (blue dot in Fig. 7C). *TERT* promoter mutations, however, are not the only way to increase *TERT* gene expression, because high *TERT* expression can also be observed in samples without identifiable *TERT* promoter mutations (Fig. 7C). Consistent with a previous report (*50*), differential motif analysis at the site of this *TERT* promoter mutation identified E74-like ETS tran-

scription factor 1 (ELF1) or ELF2 as the TF that likely binds to the de novo ETS motif (fig. S8C). In addition, we identified several mutations overlapping CCCTC-binding factor (CTCF) motif occurrences that are associated with decreased accessibility at that site (fig. S8, D and E). However, these mutations were relatively rare and often had only small effects on the accessibility of the CTCF motif site despite a known enrichment of somatic mutations in CTCF motif sites in cancer (*51*, *52*).

In addition to known *TERT* promoter mutations, integrative analysis of WGS and ATAC-seq data uncovered a mutation upstream of the FYVE RhoGEF and PH domain-containing 4 (*FGD4*) gene, a regulator of the actin cytoskeleton and cell shape. This mutation occurs in a bladder cancer sample where the variant allele frequency observed in ATAC-seq is markedly higher than the variant allele frequency observed in

WGS (Fig. 7B). This mutation is associated with a significant increase in accessibility compared to other bladder cancer samples in this cohort (Fig. 7, B and D) and is accompanied by a similar increase in *FGD4* mRNA (Fig. 7D). Moreover, this mutation upstream of the *FGD4* gene (referred to as eFGD4 for enhancer FGD4) leads to a level of accessibility that is higher than any of the other samples profiled by ATAC-seq in this study (fig. S8F) and a level of *FGD4* gene expression that is in the top 3% of all bladder cancer samples in TCGA (fig. S8G). As estimated by WGS data, this eFGD4 mutation is present in a subclone comprising about 13% of the tumor (Fig. 7E); however, the mutant allele is present in 96% of all ATAC-seq reads spanning this locus (Fig. 7E), demonstrating a strong preference for accessibility on the mutant allele. This eFGD4 mutation is analogous to, but potentially more potent than, the *TERT* promoter mutation
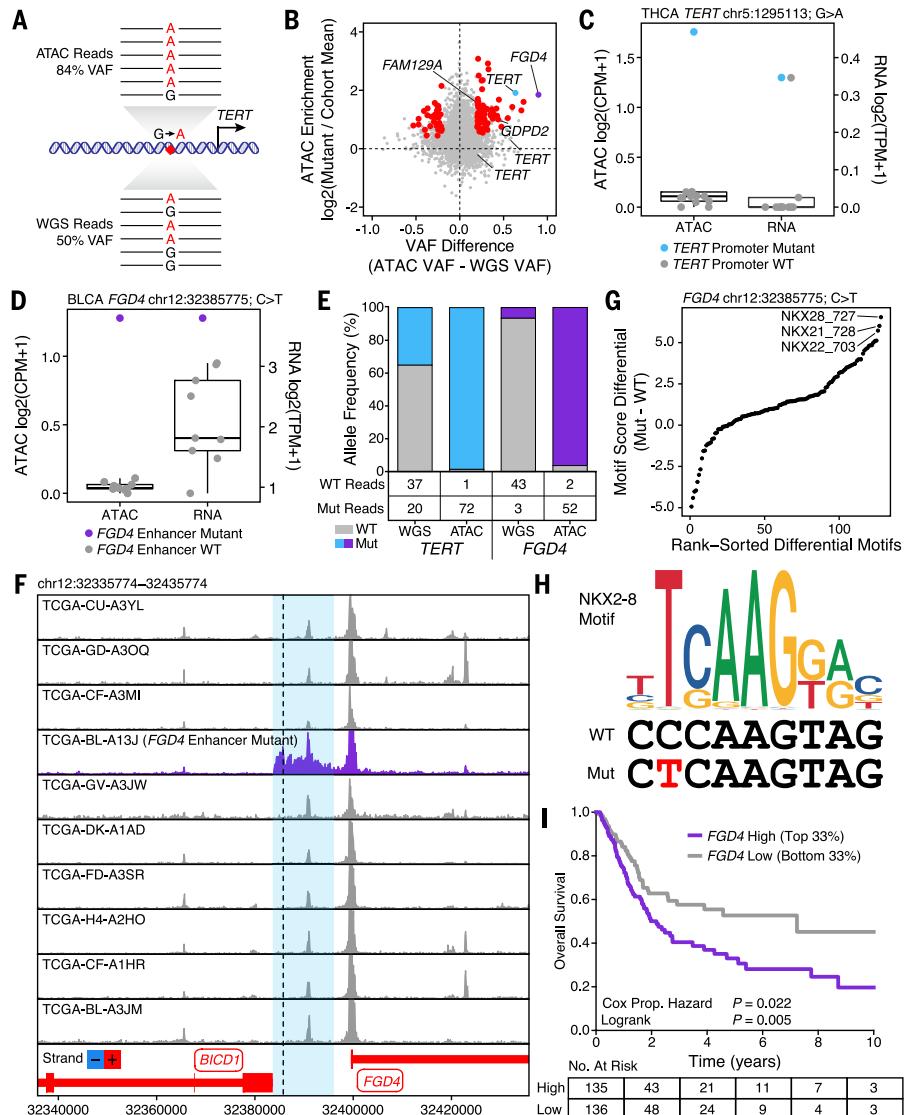
Fig. 7. Integration of WGS and ATAC-seq identifies cancer-relevant regulatory mutations.
(**A**) Schematic of how functional variants are identified in regulatory elements. The example shown depicts the *TERT* promoter. (**B**) Dot plot of the difference in variant allele frequency (VAF) of ATAC-seq and WGS and the changes in chromatin accessibility caused by the given variant with respect to other samples of the same cancer type. Variants with a higher variant allele frequency in ATAC-seq than WGS would be expected to cause an increase in accessibility. Each dot represents an individual somatic mutation. (**C**) Normalized ATAC-seq and RNA-seq of thyroid cancer donors profiled in this study. Each dot represents an individual donor. Blue dot represents the donor with a *TERT* promoter mutation shown in Fig. 7B. Other thyroid cancer donors known to harbor a *TERT* promoter mutation were excluded from this plot. The hinges of the box represent the 25th to 75th percentile. WT, wild type. (**D**) Normalized ATAC-seq and RNA-seq of bladder cancer donors profiled in this study. Each dot represents an individual donor. Purple dot represents the donor with a mutation upstream of the *FGD4* gene shown in Fig. 7B. The hinges of the box represent the 25th to 75th percentile. (**E**) Comparison of wild-type and mutant reads in WGS and ATAC-seq data at the *TERT* promoter and *FGD4* upstream region for the donors highlighted in (D) and (E). (**F**) Normalized ATAC-seq sequencing tracks of the *FGD4* locus in the 10 bladder cancer samples profiled in this study, including the one sample with a mutation predicted to generate a de novo NKX motif (TCGA-BL-A13J). Locus shown represents chr12:32335774 to 32435774. The mutation position is indicated by a black dotted line. The predicted enhancer region surrounding this mutation is highlighted by light blue shading. (**G**) Difference in motif score in the wild-type and mutant *FGD4* upstream region. Motif score represents the degree of similarity between the sequence of interest and the relevant motif. Each dot represents an individual motif. (**H**) Overlay of the NXK2-8 motif (CIS-BP M6377_1.02) and the wild-type and mutant sequences of the *FGD4* upstream region. (**I**) Kaplan-Meier survival analysis of TCGA bladder cancer patients with high (top 33%) and low (bottom 33%) expression for the *FGD4* gene.

described above (Fig. 7E). In the case of the eFGD4 mutation, this dramatic allele bias occurs because chromatin at this locus is not normally accessible in any of the bladder cancer samples profiled in this study (gray dots and tracks in Fig. 7, D and F) but becomes highly accessible in the context of the eFGD4 mutation (purple dot and track in Fig. 7, D and F). Differential motif analysis identified NKX factor motifs as the most strongly enriched in the sequence corresponding to the eFGD4 mutation (Fig. 7G), where a C-to-T transition at position two generated a perfect NKX2-8 motif de novo from a latent site (Fig. 7H). RNA-seq data from the mutated sample identified multiple expressed NKX TFs [transcripts per million (TPM) > 0.5], nominating NKX3-1, NKX2-3, and NKX2-5 as potential mediators of this DNA binding event (fig. S8H). From this, we hypothesized that the eFGD4 mutation creates a de novo binding site for an NKX TF which, upon binding to the DNA, leads to a broad increase in accessibility across the entire 12-kbp region upstream of the *FGD4* gene. This hypothesis was further supported by the observation that the ATAC-seq accessibility of the entire *FGD4* upstream locus occurs on a single phased allele (fig. S8I). Moreover, separation of subnucleosomal and nucleosome-spanning reads in the ATAC-seq data are consistent with protein binding at the site of the eFGD4 mutation (light blue shading in fig. S8I). Lastly, because higher *FGD4* expression is significantly associated with worse overall survival in bladder cancer (Fig. 7I and fig. S8J), this mutation could have functional consequence in this particular cancer. Whether the eFGD4 mutation or other enhancer mutations emerge as recurrent drivers of human cancer should be addressed in future studies. Our data identified multiple additional noncoding mutations associated with a concomitant gain of chromatin accessibility and increase in RNA expression (fig. S8, K to Q), and we anticipate that future work will uncover mechanisms underlying this type of regulatory mutation across all cancer types.

Discussion

Here we provide an initial characterization of the chromatin regulatory landscape in primary human cancers. This dataset identified hundreds of thousands of accessible DNA elements, expanding the dictionary of regulatory elements discovered through previous large-scale efforts such as The Roadmap Epigenomics Project. The identification of these additional elements was made possible through (i) our analysis of primary cancer specimens, (ii) greater saturation of some cancer and tissue types in our dataset, or (iii) potential differences between ATAC-seq and DNase-seq platforms. Nevertheless, the high overlap between the two datasets demonstrates the robustness of both platforms and the consistency of the observed results.

The exquisite cell type–specificity of distal regulatory elements from our ATAC-seq data enabled the classification of cancer types and the discovery of previously unappreciated cancer

subtypes. De novo clustering of TCGA samples based on chromatin accessibility strongly overlaps previous integrative clustering methods, identifying 18 distinct cancer clusters. Comparing this clustering scheme to other clustering schemes defined by cancer type, mRNA, miRNA, DNA methylation, RPPA, and DNA copy number alterations, we observed the strongest concordance of our clustering scheme with mRNA and cancer type, consistent with a close functional linkage between chromatin accessibility and transcriptional output. The strength of the observed associations is influenced by the features represented for each platform. For example, the DNA methylation clusters are based on cancer-specific promoter hypermethylation (28). Clustering based on DNA methylation at distal regulatory elements would likely show a stronger correlation with the ATAC-seq groupings, but distal regulatory element representation on the DNA methylation array used for these samples was too sparse to allow such an analysis. We also identified epigenetically distinct subtypes of kidney renal papillary cancer that have clear differences in overall survival. This cancer type–specific activity in DNA regulatory elements may arise via mutations within the regulatory element, pathologic transcription factor activity, or reflect the regulatory state of the tumor's cell of origin (e.g., stem cells). As the chromatin accessibility landscapes of additional primary cancer samples are profiled, we anticipate the identification of further epigenetic subdivisions with prognostic implications, potentially nominating avenues for therapeutic intervention.

The data generated in this study fully represents the cellular complexity of primary human tumors, comprising signals from tumor cells, infiltrating immune cells, stromal cells, and other normal cell types. In many ways, this complexity is advantageous because it allows complex systems-level analyses to be performed in the future, including cellular deconvolution approaches to understand the contributions of various cell types or cell states to the overall landscape of chromatin accessibility. However, the admixed nature of this signal also highlights the need for future work to profile the chromatin accessibility of matched healthy tissues to further refine the specific changes that drive cancer. Nevertheless, the chromatin accessibility profiles generated in this study represent the largest effort to date to characterize the regulatory landscape in primary human cancer cells.

Using this data-rich resource, we identified classes of TFs whose expression leads to different patterns in TF occupancy and motif protection. By integrating RNA-seq and ATAC-seq, we found factors whose expression is sufficient for both motif protection and nucleosome repositioning and demonstrated this binding to be inversely correlated with the level of DNA methylation at those binding sites. Despite this strong correlation, many sites of differential chromatin accessibility do not show differential methylation, demonstrating the complemen-

tarity of these two data types, perhaps owing to the presence of intermediate chromatin states such as poised promoters or enhancers (53, 54).

Moreover, integration of RNA-seq and ATAC-seq across the 373 donors with paired datasets enabled a quantitative model to link the accessibility of a regulatory element to the expression of predicted target genes. This workflow identified putative links for more than half of the protein-coding genes in the genome, informing the target genes of poorly understood GWAS SNPs and increasing our understanding of cancer gene regulatory networks. These predictions were further supported using 3D chromosome conformation data, and a subset were validated through CRISPRi experiments in breast cancer cell lines. However, profiling of chromosome conformation in primary cancer samples has not been performed on a large scale. Future work to produce maps of chromosome conformation in these or other primary cancer samples will improve our understanding of gene regulatory networks in cancer and further clarify the roles for certain GWAS-identified SNPs in cancer initiation and progression.

Lastly, through integration of WGS and ATAC-seq, we revealed a class of somatic mutations that occur in regulatory regions and lead to strong gains in chromatin accessibility. We demonstrated that these mutations likely lead to changes in nearby gene expression and affect genes whose expression is linked to poorer overall survival. Some of these mutations, such as those occurring in the *TERT* promoter, have been found to be recurrent whereas others, such as the mutation upstream of the *FGD4* gene, may be rare but functionally important. Because the enhancer functions are often distributed and latent enhancer sequences are pervasive in the genome, noncoding mutations in cancer may be especially challenging and require high-throughput functional assessment. Future larger-scale efforts to combine genome and epigenome sequencing will pave the way to tackling the noncoding genome in cancer.

Materials and methods summary

ATAC-seq data was generated from 410 tissue samples from the TCGA collection of primary human tumors. These samples spanned 23 different tumor types. These ATAC-seq data were used to cluster samples, identifying epigenetically defined patient subgroups. Moreover, TF regulators of cancer were defined, and footprinting of these regulators was correlated to gene expression to identify putative classes of TFs. A correlation-based model was developed to link ATAC-seq peaks to putative target genes. These putative links were validated using CRISPRi-based perturbation of the peak region followed by quantification of changes in gene expression. Publicly available HiChIP data and GTEx eQTL data were further used to support genome-wide peak-to-gene linkage predictions. Lastly, WGS and ATAC-seq were combined to identify non-coding mutations that affect chromatin accessibility in an allele-specific manner.

REFERENCES AND NOTES

1. C. Hutter, J. C. Zenklusen, The Cancer Genome Atlas: Creating lasting value beyond its data. *Cell* **173**, 283–285 (2018). doi: 10.1016/j.cell.2018.03.042; pmid: 29625045
2. W. A. Flavahan, E. Gaskell, B. E. Bernstein, Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017). doi: 10.1126/science.aal2380; pmid: 28729483
3. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011). doi: 10.1016/j.cell.2011.02.013; pmid: 21376230
4. M. Egeblad, E. S. Nakasone, Z. Werb, Tumors as organs: Complex tissues that interface with the entire organism. *Dev. Cell* **18**, 884–901 (2010). doi: 10.1016/j.devcel.2010.05.012; pmid: 20627072
5. W. Zhou *et al.*, DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018). doi: 10.1038/s41588-018-0073-4; pmid: 29610480
6. M. Almamun *et al.*, Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia. *Epigenetics* **10**, 882–890 (2015). doi: 10.1080/15592294.2015.1078050; pmid: 26308964
7. Y. He *et al.*, Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1633–E1640 (2017). doi: 10.1073/pnas.1618353114; pmid: 28193886
8. L. Yao, H. Shen, P. W. Laird, P. J. Farnham, B. P. Berman, Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.* **16**, 105 (2015). doi: 10.1186/s13059-015-0668-3; pmid: 25994056
9. M. Ceccarelli *et al.*, Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016). doi: 10.1016/j.cell.2015.12.028; pmid: 26824661
10. H. Noushmehr *et al.*, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010). doi: 10.1016/j.ccr.2010.03.017; pmid: 20399149
11. T. Hinoue *et al.*, Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* **22**, 271–282 (2012). doi: 10.1101/gr.117523.110; pmid: 21659424
12. P. A. Northcott *et al.*, The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017). doi: 10.1038/nature22973; pmid: 28726821
13. The Cancer Genome Atlas Research Network, Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016). doi: 10.1056/NEJMoa1505917; pmid: 26536169
14. B. Akhtar-Zaidi *et al.*, Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012). doi: 10.1126/science.1217277; pmid: 22499810
15. H. Chen *et al.*, A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell* **173**, 386–399.e12 (2018). doi: 10.1016/j.cell.2018.03.027; pmid: 29625054
16. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). doi: 10.1038/nmeth.2688; pmid: 24097267
17. M. R. Corces *et al.*, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017). doi: 10.1038/nmeth.4396; pmid: 28846090
18. A. Kundaje *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
19. J. Schuijers *et al.*, Transcriptional dysregulation of *MYC* reveals common enhancer-docking mechanism. *Cell Reports* **23**, 349–360 (2018). doi: 10.1016/j.celrep.2018.03.056; pmid: 29641996
20. G. Andrey *et al.*, A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013). doi: 10.1126/science.1234167; pmid: 23744951
21. M. Yeager *et al.*, Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007). doi: 10.1038/ng2022; pmid: 17401363
22. I. P. M. Tomlinson *et al.*, A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008). doi: 10.1038/ng.111; pmid: 18372905

23. I. K. Sur *et al*., Mice lacking a *Myc* enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012). doi: 10.1126/science.1228606; pmid: 23118011

24. R. E. Thurman *et al*., The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi: 10.1038/nature11232; pmid: 22955617

25. M. R. Corces *et al*., Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016). doi: 10.1038/ng.3646; pmid: 27526324

26. L. J. P. van der Maaten, G. E. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

27. A. Rodriguez, A. Laio, Clustering by fast search and find of desity peaks. *Science* **344**, 1492–1496 (2014). doi: 10.1126/science.1242072; pmid: 24970081

28. K. A. Hoadley *et al*., Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304.e6 (2018). doi: 10.1016/j.cell.2018.03.022; pmid: 29625048

29. M. Kulis *et al*., Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* **47**, 746–756 (2015). doi: 10.1038/ng.3291; pmid: 26053498

30. H. S. Kim *et al*., Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature* **556**, 510–514 (2018). doi: 10.1038/s41586-018-0048-8; pmid: 29670286

31. H. Shen *et al*., Integrated molecular characterization of testicular germ cell tumors. *Cell Reports* **23**, 3392–3406 (2018). doi: 10.1016/j.celrep.2018.05.039; pmid: 29898407

32. S. Werner *et al*., Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *J. Biol. Chem.* **288**, 22993–23008 (2013). doi: 10.1074/jbc.M113.456293; pmid: 23814079

33. S. K. Denny *et al*., Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016). doi: 10.1016/j.cell.2016.05.052; pmid: 27374332

34. S. Baek, I. Goldstein, G. L. Hager, Bivariate genomic footprinting detects changes in transcription factor activity. *Cell Reports* **19**, 1710–1722 (2017). doi: 10.1016/j.celrep.2017.05.003; pmid: 28538187

35. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: 10.1038/nmeth.4401; pmid: 28825706

36. Y. Yin *et al*., Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017). doi: 10.1126/science.aaj2239; pmid: 28473536

37. T. Ellis *et al*., The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001). doi: 10.1101/gad.200101; pmid: 11544187

38. B. M. Javierre *et al*., Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016). doi: 10.1016/j.cell.2016.09.037; pmid: 27863249

39. C. P. Fulco *et al*., Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016). doi: 10.1126/science.aag2445; pmid: 27708057

40. L. S. Qi *et al*., Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013). doi: 10.1016/j.cell.2013.02.022; pmid: 23452860

41. Y. H. Eom, H. S. Kim, A. Lee, B. J. Song, B. J. Chae, BCL2 as a subtype-specific prognostic marker for breast cancer. *J. Breast Cancer* **19**, 252–260 (2016). doi: 10.4048/jbc.2016.19.3.252; pmid: 27721874

42. S. W. Cho *et al*., Promoter of lncRNA gene *PVT1* is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018). doi: 10.1016/j.cell.2018.03.068; pmid: 29731168

43. T. C. Silva, S. G. Coetzee, L. Yao, D. J. Hazelett, H. Noushmehr, B. P. Berman, Enhancer linking by methylation/expression relationships with the R package ELMER version 2. bioRxiv 148726 [Preprint]. 11 June 2017. doi: 10.1101/148726

44. A. G. Robertson *et al*., Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* **171**, 540–556.e25 (2017). doi: 10.1016/j.cell.2017.09.007; pmid: 28988769

45. M. A. A. Castro *et al*., Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* **48**, 12–21 (2016). doi: 10.1038/ng.3458; pmid: 26618344

46. V. Thorsson *et al*., The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018). doi: 10.1016/j.immuni.2018.03.023; pmid: 29628290

47. K. Yoshihara *et al*., Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013). doi: 10.1038/ncomms3612; pmid: 24113773

48. S. L. Carter *et al*., Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012). doi: 10.1038/nbt.2203; pmid: 22544022

49. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015). doi: 10.1016/j.cell.2014.12.033; pmid: 25594174

50. M. M. Makowski *et al*., An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. *Proteomics* **16**, 417–426 (2016). doi: 10.1002/pmic.201500327; pmid: 26553150

51. R. Katainen *et al*., CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47**, 818–821 (2015). doi: 10.1038/ng.3335; pmid: 26053496

52. D. Hnisz *et al*., Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016). doi: 10.1126/science.aad9024; pmid: 26940867

53. R. D. Hawkins *et al*., Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010). doi: 10.1016/j.stem.2010.03.018; pmid: 20452322

54. T. K. Kelly *et al*., Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012). doi: 10.1101/gr.143008.112; pmid: 22960375

CANCER

Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival

Charles E. Gast[1]*, Alain D. Silk[1]*, Luai Zarour[2]*, Lara Riegler[3]†, Joshua G. Burkhart[4],
Kyle T. Gustafson[5,6], Michael S. Parappilly[1], Minna Roh-Johnson[7]‡, James R. Goodman[8],
Brennan Olson[1], Mark Schmidt[1], John R. Swain[1], Paige S. Davies[1], Vidya Shasthri[1], Shinji Iizuka[1],
Patrick Flynn[1], Spencer Watson[1]§, James Korkola[3,9], Sara A. Courtneidge[1,9], Jared M. Fischer[5,9,10],
Jerry Jaboin[9,11], Kevin G. Billingsley[2,9], Charles D. Lopez[3,9], Julja Burchard[12]¶, Joe Gray[6,9],
Lisa M. Coussens[1,9], Brett C. Sheppard[2,9], Melissa H. Wong[1,9]‖

High lethality rates associated with metastatic cancer highlight an urgent medical need for improved understanding of biologic mechanisms driving metastatic spread and identification of biomarkers predicting late-stage progression. Numerous neoplastic cell intrinsic and extrinsic mechanisms fuel tumor progression; however, mechanisms driving heterogeneity of neoplastic cells in solid tumors remain obscure. Increased mutational rates of neoplastic cells in stressed environments are implicated but cannot explain all aspects of tumor heterogeneity. We present evidence that fusion of neoplastic cells with leukocytes (for example, macrophages) contributes to tumor heterogeneity, resulting in cells exhibiting increased metastatic behavior. Fusion hybrids (cells harboring hematopoietic and epithelial properties) are readily detectible in cell culture and tumor-bearing mice. Further, hybrids enumerated in peripheral blood of human cancer patients correlate with disease stage and predict overall survival. This unique population of neoplastic cells provides a novel biomarker for tumor staging, as well as a potential therapeutic target for intervention.

INTRODUCTION

Historic dogma describing tumor evolution is based on outgrowth and expansion of clonal tumor populations; however, it is now appreciated that both genetic and nongenetic mechanisms drive tumor evolution fostering phenotypic variability of neoplastic cells and their clones. These changes underlie aggressive tumor growth, metastatic spread, acquisition of tumor heterogeneity, and therapeutic response or resistance (1, 2). While our understanding of the molecular and cellular mechanisms contributing to intratumoral heterogeneity has significantly expanded, there are no effective therapies quelling het-

erogeneity to improve patient stratification or response to anticancer therapies, highlighting the need for advances in this area.

Heterotypic cell fusion is a fundamental developmental mechanism serving to enhance cellular diversity; the most notable and best-studied example is fusion of sperm and egg. In adult murine intestines, we previously reported that fusion between hematopoietic and epithelial cells is readily detected in response to injury (3, 4); similar findings have been reported with various cells, including hepatocytes, cardiomyocytes, and skeletal muscles (5–8). In cancer, however, despite a century-old hypothesis that cell fusion contributes to tumor initiation (9–11) and acquisition of metastatic behaviors (10, 12, 13), few experimental studies have mechanistically addressed the functional underpinnings or consequences of cell fusion in the etiology of malignant progression. Reports using various approaches identify human tumor cells with immune and malignant characteristics (13–19); the etiologic mechanism for these cells is attributed to cell fusion, developmental mimicry, transdifferentiation, or other unidentified mechanisms. These previous studies do not address the biologic significance of the hybrid tumor cells or present evidence from experimental models to support the mechanism. Since underlying mechanisms for these cells cannot easily be determined in human subjects, murine models and in vitro studies provide a more appropriate and tractable platform for investigation.

In our previous studies, we reported that in vivo fusion between intestinal epithelial cells and macrophages (MΦs) yields hybrid offspring retaining epithelial characteristics defined by their gene expression profile (12). On the basis of this report, and recognizing that MΦs are inherently migratory, we sought to determine physiologic relevance of cell fusion to tumor heterogeneity by enhanced somatic diversity of neoplastic hybrids, through increased migratory or invasive properties, and by instilling a selective metastatic advantage. Herein,

[1]Department of Cell, Developmental and Cancer Biology, Oregon Health & Science University, Portland, OR 97239, USA. [2]Department of Surgery, Oregon Health & Science University, Portland, OR 97239, USA. [3]Department of Medicine, Oregon Health & Science University, Portland, OR 97239, USA. [4]Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR 97239, USA. [5]Center for Early Detection Advanced Research, Oregon Health & Science University, Portland, OR 97239, USA. [6]Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR 97239, USA. [7]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA. [8]Department of Physiology and Pharmacology, Oregon Health & Science University, Portland, OR 97239, USA. [9]Knight Cancer Institute, Oregon Health & Science University, Portland, OR 97239, USA. [10]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR 97239, USA. [11]Department of Radiation Medicine, Oregon Health & Science University, Portland, OR 97239, USA. [12]Department of Computational Biology, Oregon Health & Science University, Portland, OR 97239, USA.
*These authors contributed equally to this work.
†Present address: Division of Hematology/Oncology, Pediatrics, PO Box 800386, University of Virginia, Charlottesville, VA 22908, USA.
‡Present address: Department of Biochemistry, University of Utah, Salt Lake City, UT 94112, USA.
§Present address: Ludwig Center for Cancer Research of the University of Lausanne, Chemin des Boveresses 155, Epalinges VD, 1006 Lausanne, Switzerland.
¶Present address: Sera Prognostics, 2749 East Parleys Way, Suite 200, Salt Lake City, UT 84109, USA.
‖Corresponding author. Email: wongme@ohsu.edu

we present a systematic analysis of MΦ–neoplastic cell fusion (referred to as MΦ–cancer cell fusion or fusion hybrid) using ex vivo and in vivo murine cancer models to provide evidence that hybrids acquire functional MΦ-associated phenotypes that enhance tumor progression. Analyses of human tumor biopsies and peripheral blood reveal a novel circulating hybrid cell (CHC) population (defined as cells harboring hematopoietic and epithelial/tumor properties), whose numbers correlate with disease stage and predict overall outcome, thereby representing a biomarker for patient stratification.

RESULTS

In vitro–derived MΦ–neoplastic cell fusion hybrids display biparental lineage

On the basis of our previous findings that MΦs are the prominent fusogenic bone marrow–derived cell partner for epithelial cells (12), we used in vitro validation and analyses of MΦ–cancer cell fusion hybrids to examine contributions of the neoplastic cell and MΦ to the identity of the hybrid cells. To generate in vitro–derived hybrids, we engineered two murine cancer cell lines, colon adenocarcinoma (MC38) and melanoma (B16F10), to stably express Cre recombinase and histone 2B fused to red fluorescent protein (H2B-RFP). In cocultures, engineered MC38 and B16F10 cancer cells spontaneously fused with bone marrow–derived MΦs isolated from transgenic mice expressing either actin–green fluorescent protein (GFP) (20) or a yellow fluorescent protein (YFP) Cre reporter (21). This resulted in MΦ–cancer cell fusion hybrids identified by coexpression of nuclear RFP and cytoplasmic GFP or YFP (Fig. 1, A and B; figs. S1A and S2; and movie S1). YFP expression enabled subsequent fluorescence-activated cell sorting (FACS) isolation of hybrid cells and downstream validation of their identity using immunoblot analyses and YFP expression (Fig. 1C). Notably, in control experiments where conditioned media from GFP-expressing MΦs incubated on MC38s or conditioned media from Cre-expressing MC38s incubated on YFP Cre reporter MΦs, fusion hybrids were not detected.

To demonstrate the biparental lineage of hybrid cells, we used three discrete approaches. First, MΦs labeled with 5-ethynyl-2′-deoxyuridine (EdU) before coculture with H2B-RFP–expressing neoplastic cells produced MΦ–cancer cell fusion hybrids that initially harbored two nuclei, one labeled with EdU (MΦ origin) and the other expressing H2B-RFP (neoplastic cell origin; Fig. 1D). Upon the first mitotic division, binucleated hybrids underwent nuclear fusion, yielding a single nucleus containing both EdU- and H2B-RFP–labeled DNA (Fig. 1D).

A second approach, karyotype analyses of sex chromosomes, demonstrated that male MΦs (XY) fused to neoplastic cells (XO) generated hybrids containing three sex chromosomes (XXY; Fig. 1E), consistent with a fusion event. Chromosome enumeration revealed that hybrids existed as a unique cell population defined by their sex chromosome and total chromosome content when compared to parental MΦs or cancer cells (Fig. 1F, red spheres are hybrids, black spheres are MΦs, and white spheres are MC38s). Loss of chromosomes observed in hybrid clones occurred with temporal in vitro passage (fig. S3A); karyotype analyses of single hybrid cells revealed variable chromosome numbers (Fig. 1F), indicating that cell fusion contributes to tumor cell heterogeneity.

Finally, transcriptome analyses revealed that MΦ–cancer cell hybrids predominantly exhibited neoplastic cell transcriptional identity, while notably, retained MΦ gene expression signatures (Fig. 1G,

red bar, and table S1) that clustered into gene ontology (GO) biologic functions attributed to MΦ behavior (table S2). Of the five independently analyzed hybrid clones, each displayed a high degree of heterogeneity with respect to their MΦ gene expression. Together, these findings support the tenet that cell fusion between MΦs and neoplastic cells produces heterogeneous hybrid cells sharing characteristics of both parental predecessors but possessing their own characteristics.

Fusion hybrids acquire differential response to the microenvironment

Despite acquiring MΦ gene expression profiles, MΦ–cancer cell fusion hybrids initially retained in vitro proliferative capacity similar to unfused neoplastic cells, as opposed to MΦs (fig. S3B). However, with prolonged culture, that is, past confluence, unfused neoplastic cells pile on themselves, forming cellular aggregates, whereas MΦ–cancer cell fusion hybrids remained sheet-like with mesenchymal histologic features, indicating an acquired contact inhibition (fig. S3B and movie S2). These data indicate that, although hybrids have similar division rates, they gain differential growth properties as compared to unfused cancer cells. To determine whether these in vitro differences were recapitulated when cells were grown in vivo, in vitro–derived hybrids from MC38 or B16F10 cells were respectively injected subcutaneously into the flank or intradermally into syngeneic immunocompetent mice. Hybrids retained tumorigenic potential, with MC38 hybrids displaying shorter doubling times as compared to unfused parental cancer cells [data from two hybrid clones and two unfused cancer cell clones (Fig. 2A); B16F10 not shown], indicating that hybrids gain growth advantage in an in vivo microenvironment.

To determine whether hybrid cells acquired enhanced ability to seed and/or proliferate in ectopic microenvironments, we conducted experimental metastases assays. MC38-derived hybrids injected into spleens readily trafficked to the liver and resulted in increased metastatic foci per area compared to unfused parental cancer cells (Fig. 2B), indicating that cancer fusion hybrids gained enhanced properties required for trafficking to metastatic sites, seeding, and/or growing in a new microenvironment. Likewise, B16F10-derived fusions injected intravenously resulted in greater metastatic lung tumor area, relative to control, unfused B16F10 (Fig. 3G), indicating that they trafficked, adhered, or proliferated more efficiently within the lung. These acquired phenotypic behaviors aligned with data from gene expression analyses that identified increased fusion-associated expression of GO pathway genes implicated in metastatic spread (table S2). In particular, those pathways contributing to tumor invasion—attachment, matrix dissolution, and migration—as well as pathways involving response to specific microenvironmental cues (22–24) were up-regulated in hybrids relative to unfused tumor cells.

These results led us to ask whether tumor hybrids gain selective advantages in different microenvironments that may reflect primary tumor or metastatic sites. To directly test distinct microenvironmental interactions, we evaluated adhesion phenotypes and cytokine-dependent growth responsiveness of MC38-derived fusion hybrids versus unfused tumor cells using a microenvironment microarray (MEMA) platform (25). This high-throughput assay specifically measures cellular behavior in distinct engineered microenvironments containing variable extracellular matrix (ECM) molecules, growth factors, and chemokines, spotted combinatorially in rows and columns, thus permitting comparison of adhesion phenotypes among unfused cancer cells, MΦs, and hybrids. Analysis of microenvironment-specific adhesion revealed that MC38 cells harbor distinct growth
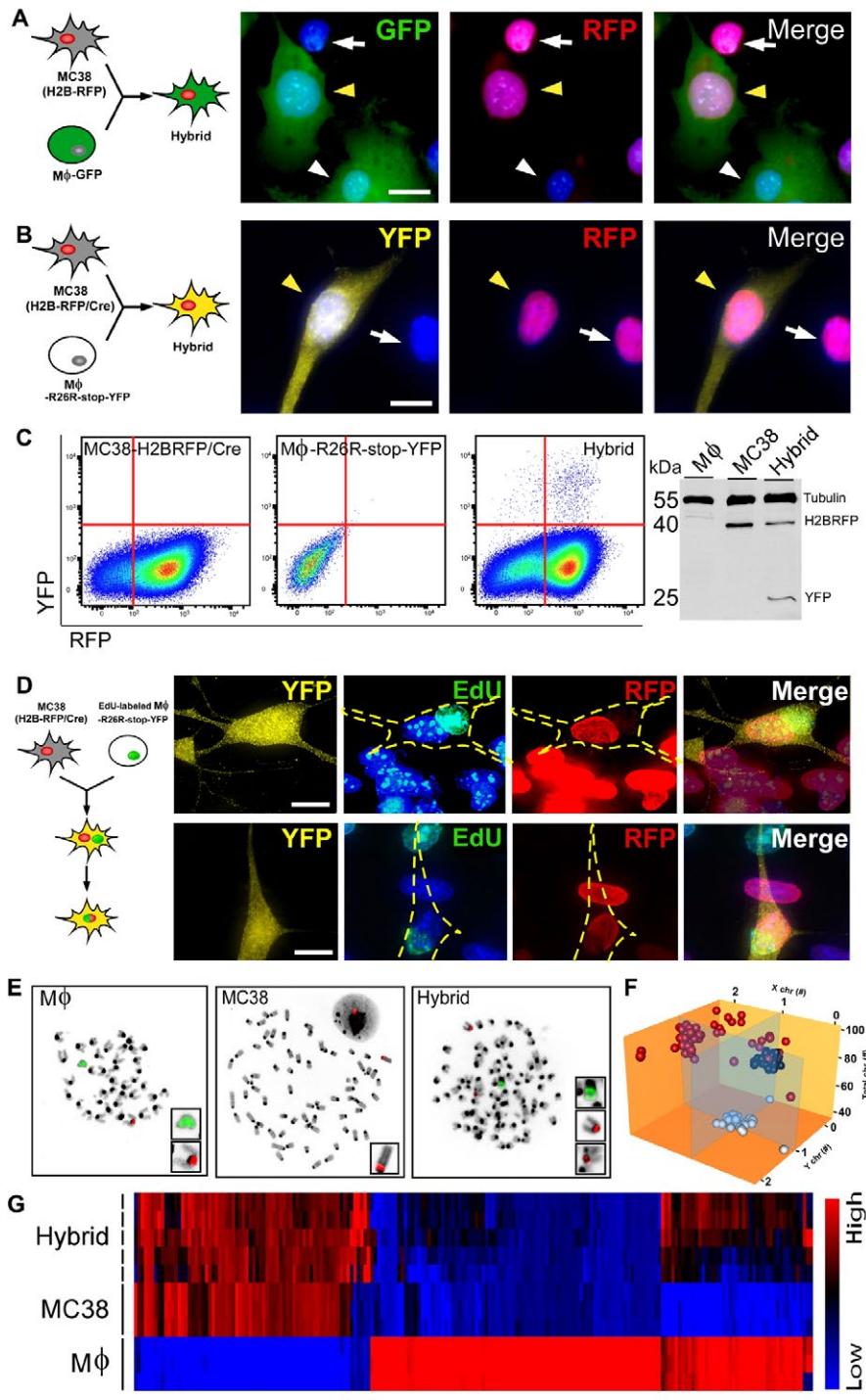
Fig. 1. In vitro–derived MΦ–cancer cell fusion hybrids. (**A**) MC38 (H2B-RFP) cancer cells and GFP-expressing MΦs cocultured in a ratio of 1:2 result in hybrid cells with RFP nuclei and GFP-expressing cytoplasm (yellow arrowhead) among unfused cancer cells (white arrow) and MΦs (white arrowhead). (**B**) MC38 (H2B-RFP/Cre) cancer cells cocultured with MΦs expressing the Cre reporter, R26R-stop-YFP results in YFP-expressing hybrid cells (yellow arrowhead). (**C**) YFP-expressing hybrids can be FACS-isolated to purify YFP-expressing hybrid cells confirmed by immunoblot. A representative FACS plot is shown. (**D**) Cocultured MΦs labeled with EdU (green) and MC38 (H2B-RFP/Cre) cancer cells produce YFP-expressing hybrids that initially harbor two nuclei—one from each parent; upon mitotic division, these undergo nuclear fusion, resulting in a single nucleus with EdU-labeled and RFP-expressing DNA. Hybrid cell outlined in yellow. Scale bar, 10 μm. (**E**) Karyotype and X (red) and Y chromosome (green) fluorescence in situ hybridization (FISH) analyses of parental MΦs, unfused MC38 cancer cells, and fusion hybrids. (**F**) Fusion hybrids (red sphere, *n* = 45) cluster as a unique population based on their chromosome number and sex chromosomes, relative to MΦs (white sphere, *n* = 27) and MC38s (black sphere, *n* = 28). (**G**) Microarray analyses of *n* = 5 independent hybrid isolates and *n* = 3 each for MC38 and MΦ populations. The yellow bar denotes hybrid gene expression unique from MC38s and MΦs. The red bar marks hybrid gene expression that is similar to that in MΦs.
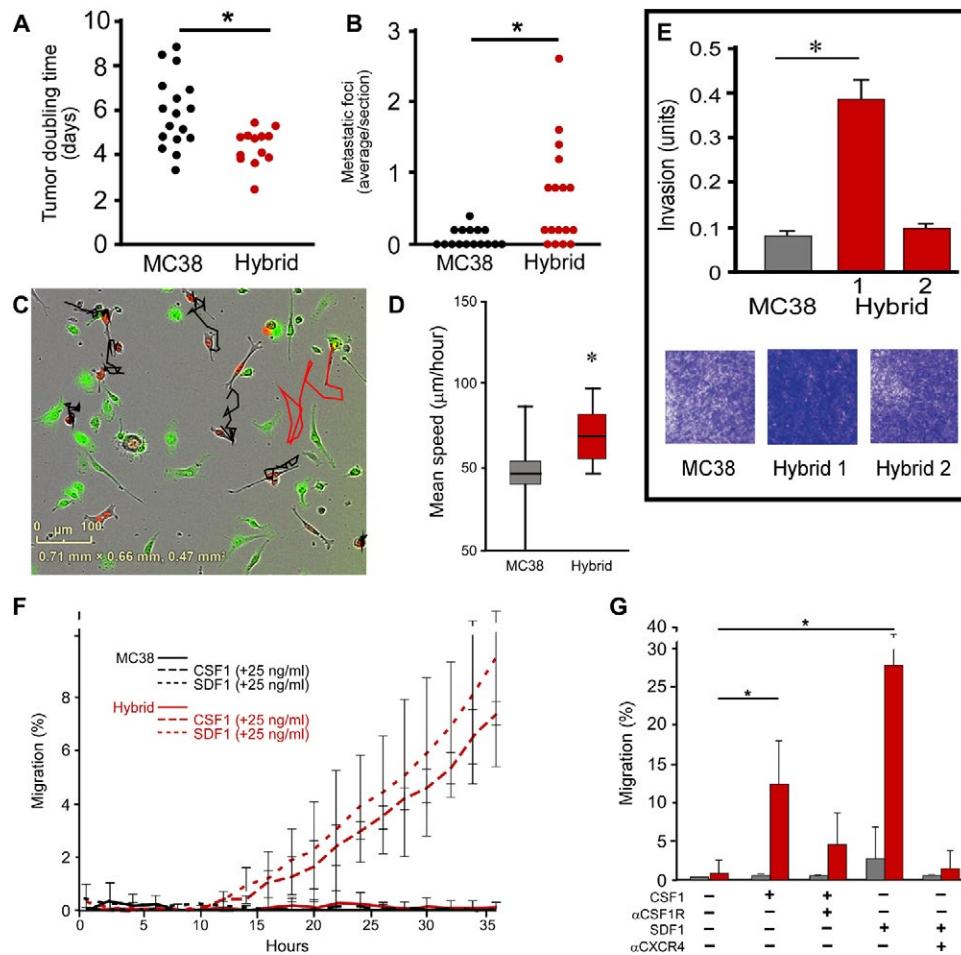
Fig. 2. In vitro–derived fusion hybrid characterization. (**A**) Proliferation analysis of MC38 cells and MC38-derived hybrids injected into flanks of immunocompetent syngeneic mice ($n = 13$ mice, each from two different hybrid isolates). Each data point reflects tumor growth in a single mouse. (**B**) Analysis of metastatic seeding potential of hybrids and MC38 cells injected into spleens and area analyzed in H&E-stained tissue sections of the liver [$n = 15$ mice injected with MC38 cells (three different hybrid clones), $n = 17$ mice injected with hybrids, with each data point reflecting metastatic tumors analyzed in the liver]. (**C**) Static portrayal of migration tracks from unfused MC38s (black) and a MC38-derived fusion hybrid (red) generated from live-imaged cocultures. Images reflect representative images. (**D**) The mean speed of hybrids (red bar) relative to MC38s (gray bar) is statistically significant (*$P < 1.1 \times 10^{-9}$). (**E**) In vitro invasion assay of MC38 cells and MC38-derived hybrids in Matrigel invasion chambers, stained with crystal violet after 15 hours. Data reflect the average of triplicate samples in biologic replicates. (**F**) A representative data set evaluating chemotaxis toward CSF1 and SDF1 ligands. Hybrid cell chemotaxis toward CSF1 and SDF1 is statistically significant relative to unfused MC38 cells after 24 hours ($P < 0.05$). Three independent experiments of triplicates or quadruplicates were conducted for each ligand. Multiple hybrid clones were assessed. (**G**) Incubation of cells with blocking antibodies to CSF1R and CXCR4 reduces migration of hybrids toward ligands. $P < 0.05$ and $P < 0.01$, respectively (hybrid, red bar; MC38, gray bar).

factor–independent adhesive preferences for select ECM molecules, most notably, fibronectin (fig. S3C). MΦs, by contrast, exhibited enhanced adhesion to collagen XXIII, vitronection (the ECM component), and more uniform adhesion across all MEMA conditions relative to unfused cancer cells (fig. S3C). Fusion hybrids exhibited a combination of adhesion biases, reflecting properties of both parental cells, potentially providing a broader adhesive affinity in different microenvironments. Further analysis, using hierarchical clustering, distinguished hybrids from unfused cancer cells with respect to adhesion on independent microenvironments (fig. S3C).

To extend these observations and determine whether MΦ fusion provided cancer cells with a selective proliferative or survival advantage, we directly analyzed effects of >90 different cytokines, chemokines, and soluble factors on unfused MC38s and hybrids (not shown). A number of growth factors induced differential influence on MC38 cells as compared to hybrids, including transforming growth factor (TGFβ1-3), which induced dose-dependent suppression of MC38 proliferation but showed no effect on hybrids (fig. S4, A, C, and D). Likewise, a moderate, dose-dependent growth-suppressing effect of hepatocyte growth factor (HGF) was apparent on MC38 cells but not on hybrids (fig. S4B). More strikingly, hybrids were resistant to tumor necrosis factor–α (TNF-α) that profoundly inhibited proliferation of MC38 cells (fig. S4E). Resistance of hybrids to cytokine concentrations that suppressed MC38 growth indicates that MΦ fusion influences selective cellular phenotypes and altered cancer cell responses to microenvironmental factors to yield adhesive, proliferative, and potentially survival advantages.

Fusion hybrids acquire MΦ-associated phenotypes
To determine whether cell fusion provides a mechanism by which neoplastic cells acquire MΦ phenotypes consistent with tumor promotion, we evaluated MΦ attributes up-regulated in hybrids identified
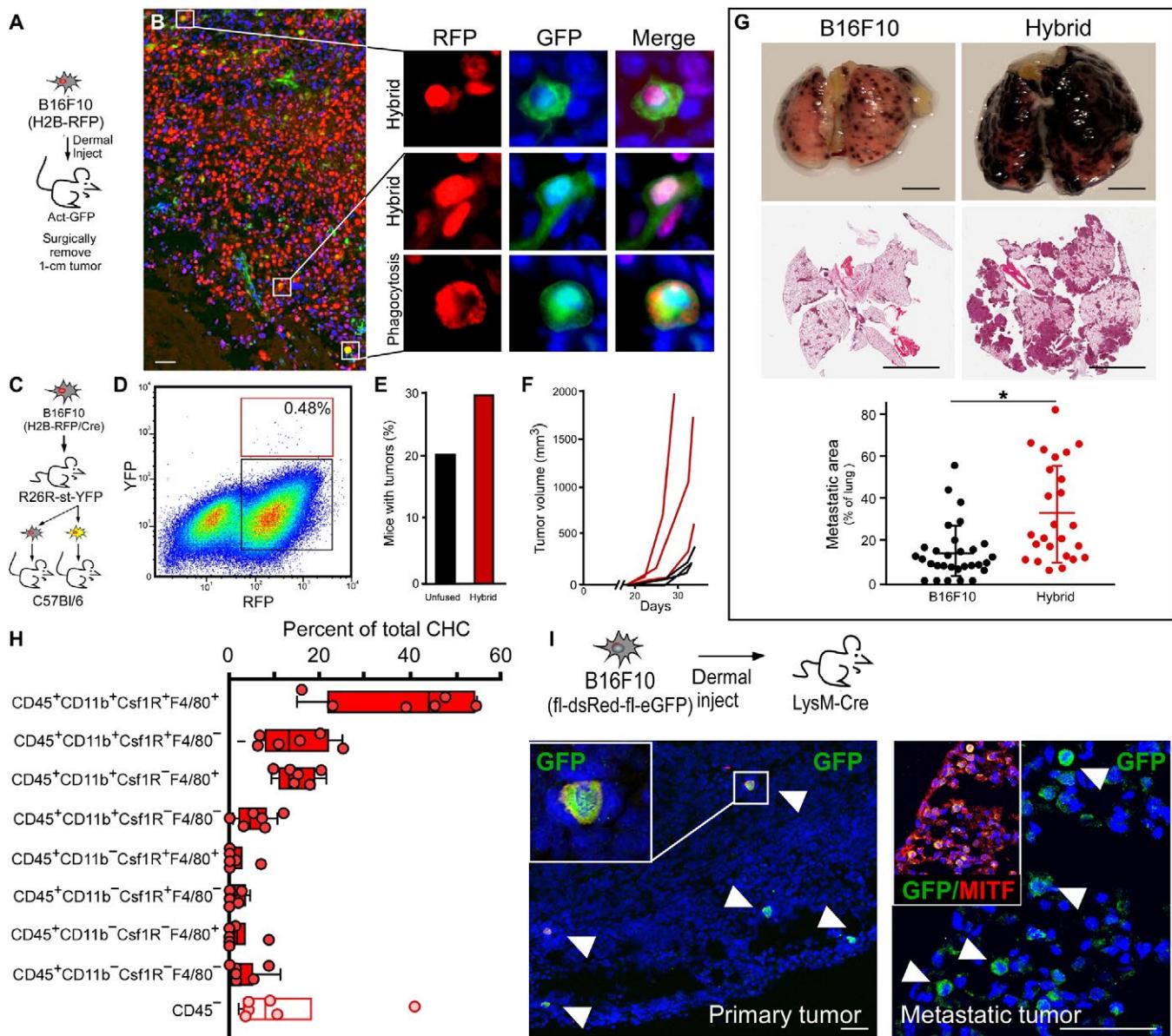
Fig. 3. B16F10 in vivo–derived fusion hybrids. (A) B16F10 (H2B-RFP) cells (5×10^4 cells) intradermally injected into GFP-expressing mice ($n = 12$, two hybrid clones) were harvested at ~1.0 cm at study end point. (**B**) Fluorescence analyses of tumor sections for RFP (red) and GFP (green) reveal double-positive hybrids and phagocytosed cancer cells with different nuclear morphology. Scale bar, 25 μm. (**C**) B16F10 (H2B-RFP/Cre) cells injected (5×10^4 cells) into R26R-stop-YFP transgenic mice ($n = 8$). (**D**) Representative FACS plot of hybrid and unfused cancer cells from a dissociated tumor, for example, hybrids (red box) and unfused (black box) cancer cells ($n = 6$ single tumor analyses, $n = 2$ pooled tumor analyses, $n = 13$ mice). (**E**) Three hundred FACS-isolated cells were injected into wild-type secondary recipient mice ($n = 19$ unfused, $n = 19$ hybrids) analyzed for tumor growth at 40 days, and (**F**) 3000 FACS-isolated cells were injected into syngeneic recipient mice ($n = 3$ MC38 injected mice, black lines; $n = 3$ hybrid injected mice, red lines) and temporally monitored for growth. (**G**) B16F10 (H2B-RFP) or MΦ–B16F10–derived hybrid cells tail vein–injected into wild-type mice ($n = 12$ mice). Macroscopic view of lungs and H&E of a tissue section. Quantification of tumor area. (**H**) Flow analyses of in vivo–derived B16F10 fusion hybrids from a primary tumor. RFP/GFP coexpressing cells analyzed for cell surface MΦ identity. All boxes represent hybrid populations. Open box denote hybrids that have lost CD45 expression ($n = 6$ mice each). (**I**) B16F10 (fl-dsRed-fl-eGFP) cells intradermally injected into LysM-Cre mice ($n = 4$) were harvested at ~1 cm. Primary tumor or metastatic lung tumors stained with antibodies to GFP (green) and the tumor protein microphtalmia-associated transcription factor (MITF, red). Scale bar, 25 μm.

by GO pathway analysis (table S2); behaviors shared by both MΦ and fusion hybrids included migration, invasion, and response to paracrine stimuli (table S3). To determine whether hybrids harbored enhanced functional motility, we evaluated in vitro–derived MC38-derived fusion hybrids and unfused MC38 cells for migratory capacity in live-imaged MΦ and neoplastic cell cocultures. Using

TrackMate analysis to calculate the mean speed of cells, hybrid clones exhibited increased motility compared to nearby unfused MC38 cells (Fig. 2, C and D). Moreover, when evaluated for invasive properties using a Boyden invasion assay, MC38-derived hybrids exhibited enhanced migration and invasion activity, relative to unfused cancer cells (Fig. 2E); these results were consistent with invasive properties

displayed by B16F10-derived hybrids (fig. S1B). Notably, we evaluated two independent hybrid isolates for each cell type; these displayed varying degrees of invasion, supporting our hypothesis that cell fusion can yield heterogeneous clonal outgrowths (Fig. 2E and fig. S1B).

GO genes involved in "response to stimulus" expressed at high levels in MΦs were also up-regulated in MΦ–cancer cell fusion hybrids. In particular, fusion hybrids harbored elevated expression of the MΦ-associated gene colony-stimulating factor 1 receptor (*Csf1R*), a significant recruitment, differentiation, and survival molecule for MΦs (*26*) implicated in regulating prometastatic MΦ effector functions (*27*). To determine whether acquisition of MΦ-associated receptor gene expression translated to functional ligand-mediated migration, we analyzed hybrids and unfused cancer cells in Transwell chemotaxis assays coupled to live imaging (IncuCyte Chemotaxis, Essen BioScience). Under these conditions, fusion hybrids migrated toward the ligand CSF1or stromal cell–derived factor-1 (SDF1) at multiple concentrations (shown at 25 ng/ml), whereas unfused MC38 cancer cells were incapable of responding to either chemoattractant; in contrast, B16F10 cancer cell hybrids exhibited decreased migratory responses (Fig. 2, F and G, and fig. S1C). Notably, the presence of a ligand did not change proliferative dynamics of either fusion hybrids or unfused cancer cells (not shown); however, incubation with blocking antibodies to CSF1R or CXCR4 reduced chemotactic responses of fusion hybrids (Fig. 2G). Some hybrid clones expressed both CSF1R and its main ligand, CSF1.

Tumor cell fusion hybrids are generated in vivo
In vitro–derived fusion hybrids allowed for in-depth functional analyses of acquired MΦ behaviors, and FISH analyses of human tumors indicated that cell fusion occurs in vivo. These studies provided only partial insight into the physiologic relevance of fusion hybrids in human cancer. Therefore, to extend the relevance and functional significance of fusion in enhanced tumor heterogeneity and acquired behaviors for tumor progression, we investigated cell fusion in mouse models of cancer. MC38 cancer cells were injected into the flanks of R26R-YFP Cre reporter mice; fusion hybrids were identified as RFP$^+$YFP$^+$ cells detected among unfused tumor cells (RFP$^+$) using immunohistochemical analyses of primary tumors. Orthotopic injection of MC38 cancer cells into the cecum resulted in pervasive peritoneal seeding, limiting the utility of this model. B16F10 melanoma cells injected intradermally into recipient mice readily developed 1.0-cm tumors (Fig. 3A), and subsequent analysis of primary B16F10 tumors revealed presence of RFP$^+$/GFP$^+$ fusion hybrids in actin-GFP recipient mice (Fig. 3B) and RFP$^+$/YFP$^+$ fusion hybrids in R26R-YFP Cre reporter mice (fig. S5, A and B). Using the latter system and FACS analyses, primary tumors were dissociated to single cells to determine the extent of cell fusion (YFP$^+$/RFP$^+$) in primary tumors. Hybrids represented a rare neoplastic cell population within the primary tumor (representative experiment: hybrids, 0.48% of RFP$^+$ cells; but overall hybrids range from 0.03 to 0.69% of RFP$^+$ cells from $n = 6$ individually analyzed tumors; Fig. 3D and fig. S6). To determine whether hybrids retained tumorigenicity, 300 FACS-isolated in vivo–derived hybrid cells were injected intradermally into secondary recipient mice ($n = 19$; Fig. 3C) and resulted in tumor growth, demonstrating that hybrid cells retained tumorigenicity (Fig. 3E). To assess tumor heterogeneity, we subsequently isolated and injected 3000 in vivo–derived fusion hybrids to facilitate temporal analyses of robust tumor growth properties in $n = 3$ mice. In vivo–derived fusion hybrids displayed different rates of tumor

growth (Fig. 3F), indicating that hybrid cells have heterogeneous growth capacity and resulted in different rates of tumor growth.

Hybrid cell metastatic potential was evaluated using an experimental metastases model. In vitro–derived hybrids were introduced into circulation by tail vein injection. Tumor cells that trafficked to the lungs and grew as metastatic foci were identified macroscopically by their pigmented appearance and microscopically on tissue section by hematoxylin and eosin (H&E) staining. Metastatic tumor area was quantified; hybrid cells showed markedly greater metastatic burden than injected unfused tumor cells (Fig. 3G).

To determine whether in vivo–derived fusion was generated from MΦ fusion partners, primary tumors were dissociated into single cells, and fusion hybrids coexpressing RFP and GFP were analyzed for cell surface MΦ antigen expression (Fig. 3H and fig. S7). Identification of discrete populations of hybrids with MΦ-associated surface identity points to a MΦ fusion partner. Further, to establish the MΦ as a fusion partner, B16F10 cells harboring an fl-dsRed-fl-eGFP allele were orthotopically injected into LysM-Cre transgenic mice (Fig. 3I). Analyses of primary tumor and lung metastases revealed MITF-expressing tumor cells harboring Cre-mediated GFP expression (Fig. 3I). Collectively, these data indicate that hybrid cells are of MΦ-tumor fusion origin, develop spontaneously in vivo, retain tumorigenic capacity, and exhibit accelerated tumor progression properties.

MΦ–tumor cell fusion hybrids are enriched in circulation
Detectable fusion hybrids in both primary and metastatic sites supported the possibility that fused neoplastic cells readily disseminate from primary to distant sites. To explore this, we collected blood from GFP$^+$ mice with established isogenic RFP$^+$ B16F10 tumors (Fig. 4A). Peripheral blood was subjected to flow cytometry for quantification of circulating tumor cells (CTCs) as defined by their RFP expression. RFP$^+$/GFP$^+$ fusion hybrids (or CHC) were easily detectible within the total RFP$^+$ population, representing 90.1% of the tumor cells in circulation, markedly outnumbering unfused RFP$^+$/GFP^{-ve} CTCs (Fig. 4A). Imaging of individual CHCs confirmed their fusion identity and morphologically distinguished them from MΦs that had phagocytosed or adhered to cancer cells (Fig. 4A).

The classical definition of CTCs in human cancer is a circulating cell expressing a tumor antigen [typically epithelial cell adhesion molecule (EPCAM) or cytokeratin for epithelial cancers] and not expressing the pan-leukocyte antigen CD45 (*28*, *29*). MΦs normally express CD45; therefore, we reasoned that MΦ–cancer cell fusion hybrids would also express this cell surface epitope. The majority of RFP$^+$/GFP$^+$ fusion hybrids expressed CD45, while unfused RFP$^+$ cancer cells largely did not (Fig. 4B and fig. S8). The classical isolation approaches for CTCs exclude any CD45-expressing cells and therefore exclude the novel CHC population from routine analyses. The presence of both classical CTCs and novel CHCs highlights heterogeneity of tumor cells in circulation destined to seed metastatic tumors. Moreover, presence of CD45-expressing CHCs in our mouse models prompted us to investigate the presence of this unique hybrid population in human cancer patients.

Hybrids in humans correlate with disease stage and patient survival
To evaluate the biological significance of hybrids in humans, we first determined whether hybrids between blood cells and epithelial-derived cancer cells were detectable. To accomplish this, we exploited a disease scenario that supports identification of hybrids harboring properties of peripheral mononuclear blood cells and epithelial

cells (*30*)—specifically, the analysis of tumor biopsies from female cancer patients who had previously received a sex-mismatched bone marrow transplant and subsequently developed a secondary solid tumor. In these patients, only the male donor–transplanted hematopoietic cells should contain a Y chromosome; therefore, identification of Y chromosome–positive nuclei in cytokeratin-positive cells within the tumor biopsy could indicate fusion between a peripheral mononuclear blood cell and an epithelial tumor cell. We identified tumor epithelia by a pathologic review of H&E-stained tumor biopsies, and then we probed tissue sections with pan-cytokeratin antibodies and interrogated with Y chromosome FISH probes to identify cellular products consistent with fusion between neoplastic cells and transplanted male hematopoietic cells (Fig. 5 and fig. S9). In a biopsy from a female patient with pancreatic ductal adenocarcinoma (PDAC), neoplastic cell nuclei containing a Y chromosome were detectible throughout regions of the tumor (Fig. 5, A to E, and fig. S9, A and B), as well as in premalignant regions of pancreatic intraepithelial neoplasia (PanIN;

fig. S9C). Confocal microscopy confirmed that Y chromosomes were located in nuclei of cytokeratin-positive epithelial tumor cells (see higher magnifications in Fig. 5). For these studies, tumor specimens from seven patients were examined, and all contained evidence of fusion by these criteria. Y chromosome–positive epithelial tumor cells were not unique to PDAC, as fusion hybrids were detected in other solid tumors from female recipients of sex-mismatched transplantation, including renal cell carcinoma, head and neck squamous cell carcinoma, and lung adenocarcinoma (fig. S9, D to F). Control tissue staining from female and male tissue samples were carried as controls for Y chromosome detection (fig. S10, A and B). These observations were consistent with previous case reports of cell fusion in human cancer using a variety of other detection methods (*11*, *17*, *31*, *32*).

To examine hybrid cells in circulation from human patients, we analyzed the peripheral blood from a sex-mismatched bone marrow–transplanted female cancer patient. CHCs that coexpressed CD45, a pan-leukocyte marker, and EPCAM, an epithelial marker, were
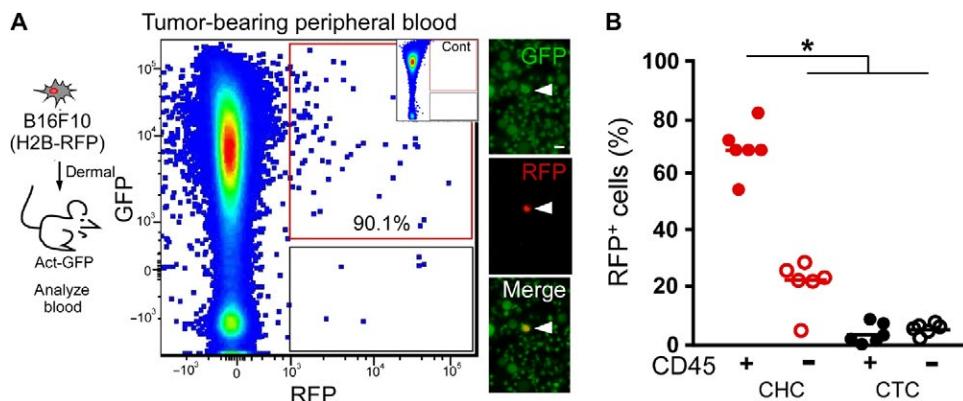
Fig. 4. Murine CTCs. (**A**) B16F10 (H2B-RFP) cells (5×10^4 cells) intradermally injected into a syngeneic GFP-expressing recipient mouse. Blood collected at time of tumor resection and analyzed by flow cytometry for GFP and RFP expression. RFP$^+$GFP$^+$ cells were detectible in presorted cell preparations by immunofluorescence. Scale bar, 50 μm. We analyzed GFP-expressing blood by flow cytometry as a negative control for (A) inset. (**B**) Percentages of fusion hybrids (RFP$^+$/GFP$^+$) and unfused CTCs (RFP$^+$/GFP$^-$) expressing the leukocyte antigen CD45 (*P < 0.000002).
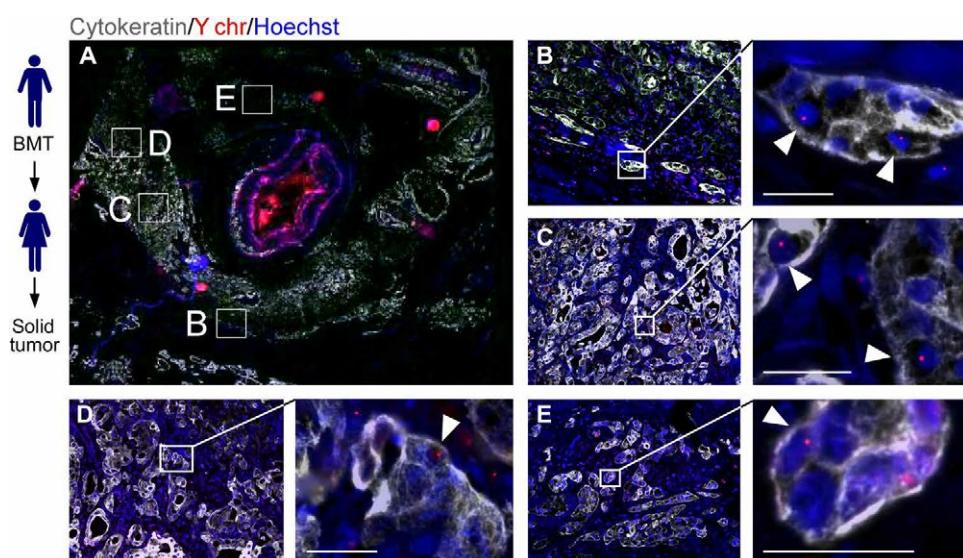
Fig. 5. Cell fusion in human tumors. Solid tumors from women ($n = 7$) with previous sex-mismatched bone marrow transplantation (BMT) permits analysis of cell fusion. (**A**) PDAC tumor section with cytokeratin (gray), the Y chromosome (Y chr, red), and Hoechst (blue) detection revealed areas of cytokeratin-positive cells with Y chr–positive nuclei (white arrowheads). Boxed representative areas are enlarged in (**B**) to (**E**). Scale bars, 25 μm.

detected (Fig. 6A). Both of these CHCs and leukocytes expressed the Y chromosome (Fig. 6A and fig. S10B). To determine whether CHCs expressed MΦ markers, analogous to fusion hybrids found in the murine tumor model, we identified MΦ epitope expression by immunohistochemical (Fig. 6B) and by flow cytometric analyses (Fig. 6C and fig. S11) was performed. A variety of MΦ epitopes were expressed on CK$^+$/CD45$^+$ CHCs, including CD163, CD68, CSFR1, and CD66b (Fig. 6B). Similarly, flow cytometric analyses revealed that CHCs from three different PDAC patients expressed MΦ epitopes, including CD14, CD16, CD11c, and CD163 (Fig. 6C). CTCs analyzed from the same patients had low expression levels of CD16. These results indicate that MΦ–tumor cell hybrids are the predominant tumor cell in circulation, although other leukocyte-tumor cell hybrids with discrete MΦ surface antigen expression most likely exist.

To explore the presence of CHCs in pancreatic cancer patients diagnosed at various tumor stages, node-negative, node-positive, or metastatic, we collected peripheral blood and performed in situ antibody staining (CD45, and CK) on isolated cells. Digital image analyses allowed validation of double-positive expression of CD45 and CK on CHCs (Fig. 6D) while excluding doublets or clusters of cells that could register as double-positive cells by flow cytometry. We determined that the number of CHCs expressing CD45$^+$/CK$^+$ significantly correlated with advanced disease (Fig. 6E). Notably, CHC enumeration revealing high expression in the blood provided a prognostic indicator of overall survival, regardless of disease stage (Fig. 6F). Conventionally defined CTCs (CD45^{-ve}/CK$^+$) did not correlate with stage or survival (Fig. 6, E and G) and were detected at quantities an order of magnitude lower than CHCs in metastatic disease. These findings identify a novel population of tumor cells in circulation, a population previously overlooked and excluded from routine analyses, which has a biologic function and correlation to clinically relevant disease status in human cancer patients.

DISCUSSION

Cell fusion between immune and neoplastic cells initiating tumorigenesis and affecting progression is an untested, century-old hypothesis (*9*, *10*) that has been only circumstantially examined (*10*, *12*, *17*, *31*, *33*–*37*). Reports of cells located in tumors containing components of both immune and neoplastic cells are increasingly frequent (*1*, *10*–*13*, *16*, *17*, *19*, *35*, *38*–*44*), although without strong evidence of etiologic mechanism or physiologic relevance. Early in vitro studies revealed that cell fusion hybrids displayed reduced cell doubling times relative to their genetic burden (*45*). More recently, it was suggested that tumor cell fusions gain shorter cell cycling times compared to either of their parental cells (*46*, *47*). These divergent views add to the controversy of whether cell fusion provides a selective advantage to evolving tumors. Moreover, description of MΦ gene expression in metastatic cancer cells as evidence that fusion propagates aggressive metastatic spread of cancer (*31*, *42*) is presented without substantial proof that these cells arise from a fusion event, further diminishing enthusiasm for this mechanism. Despite this, MΦ–neoplastic cell fusion does provide an intriguing mechanism for how neoplastic cells rapidly gain discrete cellular behaviors to facilitate metastases and to propagate intratumoral heterogeneity. Before this study, experimental results demonstrating in vivo tumor cell fusion with MΦs or investigating a function role for cell fusion in tumor progression were undetermined. Here, we demonstrate that cell fusion occurs spontaneously in a number of systems. Cell fusion can contribute to the generation of diverse neoplastic clones with altered phenotypes, implicating it as a mechanism for gain of intratumoral heterogeneity. This finding may reveal insight into diverse tumor cell pathophysiology that underlies treatment resistance, progression, and posttreatment tumor recurrence in human cancer.

We present a systematic analysis of MΦ–cancer cell fusion and provide evidence that hybrids impart physiologically relevant and functionally significant aspects contributing to tumor evolution. Together, our in vitro and in vivo murine data indicate that neoplastic cells fuse spontaneously with leukocytes and myeloid cells (that is, MΦs, neutrophils, or dendritic cells) and produce heterogeneous cancer hybrid clones. Despite their diversity, hybrid clones retain MΦ genotypes with functional phenotypes, thereby bestowing MΦ-like behaviors on neoplastic cells. Fusion hybrids express functional levels of CSF1R, which is relevant to cancer progression exemplified by the association of CSF1 overexpression in lung cancer with increased tumor cell proliferation and invasion (*48*), by the inhibition of CSF1R with decreased tumor metastasis (*49*) and by late-stage metastatic breast carcinomas frequently acquiring CSF1R expression (*38*). In human tumors, the mechanism by which tumor cells gain chemotactic responsive receptor expression remains unclear, but multiple mechanisms likely underlie transcriptional changes. Our data indicate that cell fusion could play a role in the acquisition of migratory/chemotactic functional behavior and may have important clinical implications considering hybrids' potential response to clinically relevant CSF1R inhibitors (*27*, *50*, *51*).

Data presented here indicate that MΦ–cancer cell hybrids are differentially responsive to microenvironment-derived regulatory forces. Specific extracellular conditions provide a selective adhesive and/or growth advantage to hybrids but not to unfused neoplastic cells. Given that genotypic and phenotypic diversity provides selective advantages to the fittest neoplastic clones, continued cell fusion may underlie adaptation, survival, and growth of dominant neoplastic clones within the evolving tumor microenvironment during tumor progression. Considering this possibility, cell fusion provides a previously underappreciated mechanism by which neoplastic cells gain phenotypic diversity, increasing opportunities for highly fit subclones to overcome selection pressure and drive tumor progression.

These data indicate that tumor-initiating hybrid populations can acquire behaviors allowing for navigation of the metastatic cascade—from the primary tumor, to survival in circulation, to seeding of ectopic sites, and to propagation of metastatic foci. Further, in vivo–derived hybrid cells were readily detected in peripheral blood. CHCs outnumbered conventionally isolated CTCs in both mice and humans. Moreover, in patients with pancreatic cancer, CHCs directly correlated with tumor stage and inversely correlated with overall survival—highlighting an exciting prognostic opportunity for development of this novel cell population as a liquid biomarker for disease status in discrete cancers where conventional biomarkers (for example, CTCs, cell-free DNA, exosomes, and proteins) have not demonstrated efficacy.

Unlike in murine models, the etiology of human CHCs, while consistent with cell fusion, cannot be conclusively determined. It is possible that tumor cells can gain expression of leukocyte- and MΦ-associated proteins by an undetermined mechanism or that CD45-expressing blood cells transdifferentiate into epithelial cancer cells. Despite these unexplored caveats, the CHC is a population of tumor cells previously overlooked and understudied. Our initial investigations indicate that this novel cell population has exciting potential.
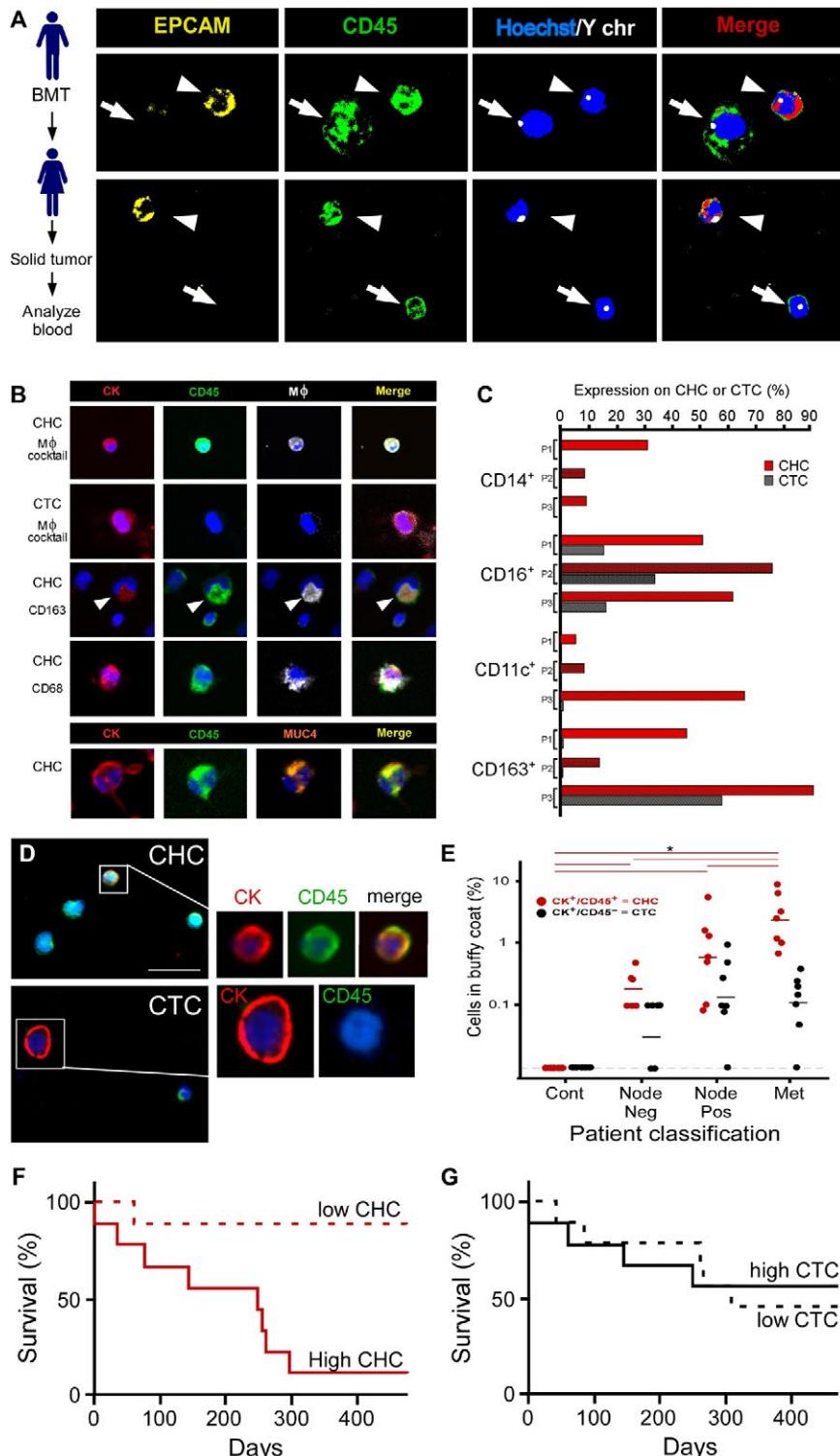
Fig. 6. Human CTCs. (**A**) Sex-mismatched bone marrow–transplanted (BMT) patient who acquired a solid tumor (PDAC). Peripheral blood was analyzed for the presence of cell fusion. Two panels displaying cell fusion hybrids (arrowheads) that costain for EPCAM (yellow) and CD45 (green) and have a Y chromosome (white dot) in their nuclei (blue). Arrows denote leukocytes. (**B**) CHCs and CTCs analyzed from $n = 4$ patients with PDAC. CHCs (CK$^+$/CD45$^+$) also express MΦ proteins (cocktail: CD68, CD163, CD66b, and CSF1R), while CTCs (CK$^+$/CD45^{-ve}) do not. CHCs also express the tumor-specific protein MUC4. (**C**) CHCs and CTCs analyzed by flow cytometry for CD14, CD16, CD11c, and CD163 expression or the cancer-specific protein MUC4 ($n = 4$ patients). (**D**) Human pancreatic cancer patient peripheral blood analyzed for cytokeratin$^+$ (red) and CD45$^+$ (green) expression using in situ analyses and digital scanning. (**E**) CK$^+$/CD45$^+$ and CK$^+$/CD45$^-$ cells quantified in patient blood across cancer stages [analysis of variance (ANOVA), *$P < 0.023$]. (**F** and **G**) Kaplan-Meier curve of dichotomized biomarkers based on median value (CHC and CTC) was associated with statistically significant increased risk of death for CHCs ($P = 0.0029$) but not for CTCs ($P = 0.95$).

Identification of functionally significant properties of this unique tumor population, a chimera of MΦs and neoplastic cells, offers opportunities for understanding the dynamic interaction between neoplastic cells and diverse infiltrating immune cell populations. Elevated CHCs relative to CTCs in peripheral blood might suggest that hybrids are immune privileged—a trait bestowed by their leukocyte identity. This scenario could have implications on immune-mediated therapeutic strategies for cancer treatment. Therefore, understanding how hybrids respond to immune therapies, such as inhibitors or agonists to costimulatory and/or coinhibitory receptors, offers an important area of future investigation. Acquisition of functional myelomonocytic receptors on hybrids indicates that they may be vulnerable to targeted therapies such as CSF1/CSF1R blockade, now being investigated in clinical trials (*51*). Alternatively, these therapies may inhibit MΦ–neoplastic cell fusion. The presence of tumor cells with acquired MΦ phenotypes supports a cell fusion mechanism in the propagation of intratumoral heterogeneity, introduces a functionally significant aspect of tumor progression and evolution, identifies an unappreciated CHC population, and uncovers a new area of tumor cell biology.

METHODS
Human samples and ethics statement
All human blood and tissue samples were collected and analyzed with approved protocols in accordance with the ethical requirements and regulations of the Oregon Health & Science University (OHSU) institutional review board. Informed consent was obtained from all subjects. Peripheral blood was obtained from cancer patients diagnosed with PDAC at various stages and treated at OHSU, as well as from healthy controls. Identification and acquisition of solid tumor or peripheral blood biopsies from female patients who previously received a gender-mismatched bone marrow transplantation was conducted by screening of the Center for International Bone Marrow Transplant Registry.

FISH and immunohistochemical analyses of human solid tumors and peripheral blood cells
Analyses of human solid tumors
The presence of cell fusion between Y chromosome–containing blood cells and host tumor epithelium was evaluated by dual FISH and immunohistochemical analyses. X and Y chromosome FISH probes were hybridized to 5-μm formalin-fixed paraffin-embedded primary human tumor sections using CEP X (DXZ1 locus) and CEP Y (DYZ1 locus) probes (Abbott Molecular) following the manufacturer's protocols. Briefly, the tissue was treated with Retrievagen A solutions (BD Biosciences) and Tissue Digestion Kit II reagents (Kreatech) and then hybridized with a probe at 80°C for 5 min and 37°C for 12 hours. Tissue sections were permeabilized with graded detergent washes at 24°C and then processed for immunohistochemical staining. Tissue was incubated with antibodies to pan-cytokeratin (Fitzgerald) and counterstained with Hoechst dye (1 μg/ml). Two slides were analyzed for each tumor section. Slides were digitally scanned and quantified by two independent investigators. Areas with Y chromosome positivity were analyzed by confocal microscopy. H&E staining was conducted on adjacent sections.

In situ analyses of human peripheral blood analyses
Patient peripheral blood was collected in heparinized Vacutainer tubes (BD Biosciences), and then, lymphocytes and peripheral mononu-clear cells were isolated using density centrifugation and LeucoSep Centrifuge Tubes (Greiner Bio-One) according to the manufacturer's protocol. Cells were then prepared for antibody staining. Briefly, cells were adhered to poly-D-lysine–coated slides, fixed, and permeabilized before staining for CD45 and cytokeratin expression using antibodies to CD45 (eBioscience) and human pan-cytokeratin (Fitzgerald). Phenotypes of CHCs were evaluated with additional antibody staining, including to CD66b (BD Pharmingen), CD68 (Abcam), CD163 (Neomarkers), CSF1R (Abcam), and EPCAM (1:200; US Biological). Tissue was developed with appropriate fluorescent-conjugated secondary antibodies [anti-mouse Cy3 (Jackson ImmunoResearch), goat anti-guinea pig 488 (Invitrogen), goat anti-guinea pig 555 (Invitrogen), anti-rabbit A647 (1:500; Thermo Fisher Scientific), and anti-mouse Cy5 (1:500; Jackson ImmunoResearch)] and then was stained with Hoechst (1 μg/ml). Slides were digitally scanned with a Leica DM6000 B microscope or a Zeiss AxioObserverZ1 microscope and analyzed using Ariol or Zeiss Zenblue software.

To determine whether circulating $CD45^+/CK^+$ or $CD45^+/EPCAM^+$ cells were cell fusion products, patient peripheral blood was subjected to FISH/immunohistochemical analyses as described for solid tissues (see above). Processed peripheral blood was interrogated with Y chromosome FISH (DYZ1 locus) probes (Abbott Molecular) following the manufacturer's protocols. Briefly, cells were hybridized with a probe at 42°C for 16 to 20 hours and then subjected to graded detergent washes. Cells were then subjected to antibody staining with anti-CD45 conjugated to fluorescein isothiocyanate (FITC) (1:100; BioLegend) and EPCAM (1:100, US Biological) and processed with anti-rabbit AF647 secondary antibodies (1:250; JacksonImmuno Research). Cells were imaged on a Zeiss AxioOberverZ1 microscope. Images were postprocessed to rule out nonspecific staining. Briefly, CZI files were opened using ZEN 2.3 Lite (Blue Edition) and saved as single-channel TIF files [four channels per CZI: EPCAM (white), Y chromosome (red), CD45 (green), and 4′,6-diamidino-2-phenylindole (DAPI; blue)]. Single-channel TIF files were loaded into MATLAB as UINT8 matrices containing RGB information at each pixel. To create binary images, pixel intensity thresholds were set for each channel image separately: Any pixel with a value above the threshold was turned ON (that is, maximum intensity), and the remaining pixels were turned OFF (that is, zero intensity). Two binary channel images were reassigned colors (EPCAM: white → yellow, Y chromosome: red → white); all binary channel images were then overlaid.

Quantification of CHCs in patient blood
Manual quantification by three independent investigators of randomly selected regions containing 2000 cells evaluated CD45 and cytokeratin status of Hoescht$^+$ cells. Percentages of CHCs in the buffy coat correlate with disease stage with significance determined by overall ANOVA post-test [$P < 6.3 \times 10^{-8}$; P values for no nodal-met (0.00035), nodal-met (0.05), and no nodal-nodal (0.15)], while none of the conventional CTCs (that is, $CD45^{-ve}$) comparisons across stage were statically significant [P values for no nodal-met (0.31) and nodal-met (0.9)]. Survival analysis was conducted on 18 of 20 pancreatic patients (2 patients were lost to follow-up, 9 patients have high levels, and 9 patients have low levels) to correlate CHCs or CTCs with time to death using Kaplan-Meier curve and log-rank test using dichotomized biomarkers based on a median value. High $CK^+/CD45^+$ (median, >0.808) was associated with a statistically significant increased risk of death (log-rank test, $P = 0.0029$) with a hazard ratio of 8.31, but high $CK^+/CD45^{-ve}$ (median, >0.101) did not have a statistically significant effect on time to death (log-rank test, $P = 0.95$).

Flow cytometric analyses of fusion hybrids in peripheral blood

For flow cytometric analysis, patient blood was collected, as described above. Red blood cell (RBC) lysis was performed by a 1-min incubation in 0.2% NaCl, followed by addition of the equivalent volume of 1.6% NaCl. Cells were washed and resuspended in FACS buffer [phosphate-buffered saline (PBS), 1.0 mM EDTA, and 5% fetal bovine serum (FBS)]. Cells were incubated in PBS containing LIVE/DEAD Fixable Aqua (1:500; Invitrogen) with Fc Receptor Binding Inhibitor (1:200; eBioscience). Cells were then incubated in FACS buffer for 30 min with CD45-APC (1:25; Thermo Fisher Scientific), CD11c-APCeF780 (1:100; Thermo Fisher Scientific), CD14-BV785 (1:100; BioLegend), CD163-PECy7 (1:100; BioLegend), EPCAM-FITC (1:100; Abcam), or cytokeratin-PE (1:500; Abcam). A BD LSRFortessa cell analyzer was used for analyses. A gating scheme established with single-color controls is provided in fig. S10. Data reflect analyses from $n = 3$ patients with PDAC.

Mice

All mouse experiments were performed in accordance to the guidelines issued by the Animal Care and Use Committee at OHSU or the Fred Hutchinson Cancer Center, using approved protocols. Mice were housed in a specific pathogen–free environment under strictly controlled light cycle conditions, fed a standard Rodent Laboratory Chow (#5001 PMI Nutrition International), and provided with water ad libitum. The following strains were used in the described studies: C57BL/6J (JAX #000664), Gt(ROSA)26Sor$^{tm(EYFP)Cos}$/J (R26R-stop-YFP; JAX#006148) (21), Tg(act-EGFP)Y01Osb (Act-GFP; JAX #006567) (20), and B6.129P2-Lyz2tm1(cre)Lfo/J (LysM-Cre; JAX#004781) (52). Mice of both genders were randomized and analyzed at 8 to 10 weeks of age. When possible, controls were littermates housed in the same cage as experimental animals.

Cell culture

MC38 mouse intestinal epithelial cancer cells were provided by J. Schlom [National Cancer Institute (NCI)], and B16F10 mouse melanoma cells were obtained from the American Type Culture Collection. Validation of cell lines was confirmed by polymerase chain reaction and by functional metastasis assay for the latter. Cell lines, both derived from C57BL/6J mice, were cultured in Dulbecco's modified Eagle's medium (DMEM) + 10% serum (Life Technologies). Stable cancer cell lines, MC38 (H2B-RFP), MC38 (H2B-RFP/Cre, B16F10 (H2B-RFP), and B16F10 (H2B-RFP/Cre), were generated by retroviral transduction using pBABE-based retroviruses, and polyclonal populations were selected by antibiotic resistance and flow-sorted for bright fluorescence as appropriate. B16F10 (fl-dsRed-fl-eGFP) cells were generated by stably expressing a pMSCV-LoxP-dsRed-LoxP-eGFP-PURO construct (Addgene #32702) into the parental B16F10 cells. Primary MΦ derivation was conducted from the bone marrow of R26R-stop-YFP or Act-GFP mice. To elicit MΦs, cells were cultured for 6 days in DMEM + 15% serum supplemented with sodium pyruvate, nonessential amino acids (Life Technologies), and CSF1 (25 ng/ml; PeproTech).

Cell fusion hybrid generating cocultures were established in MΦ-derivation media without CSF1 for 4 days. MC38 or B16F10 cells and MΦs were coseeded at a 1:2 ratio at low density. Hybrid cells were FACS-isolated for appropriate fusion markers on a Becton Dickinson InFlux or FACSVantage SE cell sorters (BD Biosciences). FACS plots are representative of at least 20 independent MC38 or B16F10 hybrid isolates (technical replicates). Low-passage hybrid isolates were established; functional experiments were conducted on passage 8 to 20 hybrid isolates. Live imaging of cocultured cells was performed using an IncuCyte ZOOM automated microscope system and associated software (Essen BioScience). Technical triplicates generated 36 movies that covered 77.4 mm^2 and were screened for hybrid generation and division. Movie contains a fusion event; a total of 21 video clips were captured containing fusion events.

EdU-labeling and karyotype analysis
During hybrid generation

Cultured cells were fixed in 4% formaldehyde in PBS and processed for immunohistochemical analyses with antibodies against GFP (1:500; Life Technologies) or RFP (1:1000; Allele Biotechnology). EdU labeling and detection were performed according to the manufacturer's directions (Life Technologies). Briefly, MΦ DNA was labeled with 10 μM EdU supplemented in media for 24 hours before hybrid generation coculture. EdU (10 μM) was used for the determination of S-phase indices as well. Biologic and technical replicates ($n = 6$) were conducted and screened for biparental hybrids.

For karyotype analyses

Chromosome spreads from cells in S phase were prepared using standard protocols, from cells treated for >12 hours with Colcemid (100 ng/ml; Life Technologies) to induce mitotic arrest. DNA was visualized by staining with DAPI; X and Y chromosomes were identified using fluorescently labeled nucleotide probes (ID Labs) as directed by the manufacturer. Images of stained fixed cells and chromosome spreads were acquired using a 40× 1.35 UApo oil objective on a DeltaVision-modified inverted microscope (IX70; Olympus) using SoftWorx software (Applied Precision) and represented maximum intensity projections of deconvolved z-stacks unless otherwise indicated. Experiments were replicated eight times. Each biologic replicate was analyzed in an independent experiment. A minimum of $n = 20$ cells were analyzed in each experiment. Chromosomes were counted manually by two independent investigators.

Gene expression analysis

Microarray analysis was performed with Mouse 430.2 gene chips (Affymetrix) at the OHSU Gene Profiling Shared Resource, and data were analyzed using GeneSifter software (Geospiza) to identify relative expression differences between cell types (replicates: MΦ, $n = 3$; MC38, $n = 3$; hybrids, $n = 5$ independent isolates) and produce GO analyses. GO category enrichment was calculated using the GOstats R package (53) and visualized using functions from the GOplot R package (54).

Code availability

The source code used to generate figures and corresponding tables is available for download from our public repository (55).

Polymerase chain reaction

DNA was extracted from frozen formalin-fixed melanoma primary tumor and lymph node sections by 40 min of incubation in lysis buffer [25 mM NaOH and 0.2 mM EDTA (pH 12)] at 95°C, followed by neutralization with equal volumes of neutralization buffer [40 mM tris-HCl (pH 5)]. RFP primers were as follows: 5′-CAGTTCCAG-TACGGCTCCAAG-3′ (forward) and 5′- CCTCGGGGTACATCCG-CTC-3′ (reverse). Actin primers were as follows: 5′-GAAGTACC-CCATTGAACATGGC-3′ (forward) and 5′-GACACCGTCCC-CAGAATCC-3′ (reverse). Reactions were run with a 60°C annealing temperature.

Microenvironment arrays

Recombinant proteins (R&D Systems and Millipore) were diluted to desired concentrations in print buffer (ArrayIt), and pairwise combinations of ECM proteins and growth factors or cytokines were made in a 384-well plate. A Q-Array Mini microarray printer (Genetix) was used to draw from the 384-well plate and print protein combinations onto Nunc 8-well chambered cell culture plates (Thermo Fisher Scientific). Each combination was printed in quintuplicate in each array, and arrays were dried at room temperature. Printed MEMAs were blocked for 5 min using 0.25% (w/v) F108 copolymer (Sigma-Aldrich) in PBS and then rinsed with PBS and media before plating cells. Cells were trypsinized, filtered to exclude cell clumps, and counted; 10^5 cells were plated on each array in 2 ml of DMEM + 2.5% serum and incubated for 30 min in a humidified tissue culture incubator. Unbound cells were gently removed, and fresh media were added. After 12 hours, the arrays were fixed with 4% formaldehyde in PBS for 10 min and stained with DAPI. Adhesion was measured as relative cellular preference: the number of cells occupying a given microenvironment condition relative to the average cell number over all occupied microenvironmental spots across the entire MEMA for each sample. Five replicate samples each for MC38 cells and MΦ and five independent MC38-derived hybrid isolates were analyzed. Standard two-tailed t tests were performed with $P < 0.05$ reported as significant. Error bars represent SEM.

In vitro–derived hybrid proliferation

For phenotypic profiling growth responsiveness to cytokines and soluble factors, 95 different cytokines or soluble signaling molecules were distributed at high, medium, and low concentrations in 384-well plates, in 25 μl of RPMI 1640 (Life Technologies) supplemented with 1% FBS, and 25 μl of a suspension of hybrid or MC38 cells (1.2×10^4 cells/ml) in DMEM + 4% FBS was added to each well. Ninety-nine wells of each plate were left cytokine-free, and no cells were added to 2 of these wells, which served to provide measurements of background signal. Plates were cultured in a humidified incubator for 72 hours, after which 5 μl of MTS reagent was added to each well. Two hours later, absorbance at 490 nm was read with a 384-well plate reader. For each plate, absorbance values for each cytokine-treated well were normalized to the mean absorbance of the cytokine-free wells on that plate and expressed in terms of SDs from the cytokine-free mean. Three independent hybrid isolates and three MC38 replicates were analyzed. Cytokines or factors that showed a potential differential effect on growth of MC38 and hybrid cells were retested in 96-well plates. In these experiments, 2.5×10^4 hybrid or MC38 cells were plated in the presence of three different concentrations for each soluble factor or in media alone (DMEM + 2.5% FBS), in triplicate for each condition. Plates were imaged every 2 hours for 90 hours, and then, cell viability was assessed.

Chemotaxis assay

Chemotaxis assays were performed using IncuCyte Chemotaxis Cell Migration Assay (Essen BioScience) with at least three technical replicates of triplicate samples. Briefly, 1000 cancer cells were plated media for 20 hours. CSF1 or SDF1 ligand (25 ng/ml) was added to the bottom well, and cells were incubated at 37°C for at least 36 hours with live imaging. The neutralizing antibodies to the CSF1R (eBioscience), CXCR4 (BioLegend), and isotype control antibody were added to the top and bottom wells (2.5 ng/μl). Migration was quantified by measuring the phase-contrast area of the top and bottom wells for each

time point using IncuCyte ZOOM software. Triplicates of each condition were performed, and the means and SDs were calculated. $P < 0.02$ for hybrids treated with CSF1 or SDF1 relative to hybrids without CSF1 or SDF1 by unpaired t test. Two independent hybrid isolates were analyzed. Technical octupulicates (MC38) or sextuplicates (B16F10) with biologic quadruplicates or triplicates were analyzed. For inhibitor studies, technical duplicates with biologic triplicates were analyzed.

Migration analysis

From IncuCyte live imaging of cocultured MΦs and cancer cells, 24- to 48-hour image series containing a cancer-MΦ fusion event was cropped and exported as two separate uncompressed audio video interleave (AVI) files: one containing only the red channel for TrackMate analysis and another containing both red and green channels with a sizing legend. Red-channel AVI files were imported into FIJI and converted to 8-bit image series with a mean filter of 1.5 pixels applied. TrackMate analysis was then performed on nuclei with an estimated diameter of 10 pixels and a tolerance of 17.5. Using the Linear Assignment Problem Tracker, settings for tracking nuclei were as follows: 75.0-pixel frame-to-frame linking and 25.0-pixel and two-frame gap track segment gap closing. Tracks segments were not allowed to split or merge. Using the analysis function in TrackMate, track statistics were exported to an Excel file, and tracks containing 11 or fewer frames were excluded from the analysis. A total of 9 hybrid cells and 536 unfused cells were analyzed with a $P < 1.1 \times 10^{-9}$ by unpaired t test. Error bars represent SD.

Boyden chamber invasion assay

In vitro invasion assay was performed, as described previously (56). Briefly, cellular invasion was measured in a growth factor–reduced Matrigel invasion chamber with 8-μm pores (#354483, Corning). Cells (3×10^5) in a medium containing 0.1% FBS were placed into each Boyden chamber. The medium containing 10% FBS was placed in the lower chamber to facilitate chemotaxis. Invasion assays were run for 15 hours, and then cells that passed through the Matrigel membrane were stained with 0.09% crystal violet/10% ethanol. After extraction by elution buffer [1:2:1 acetate buffer (pH 4.5)/ethanol/ deionized water], the stain was measured at 560 nm. Representative images of invaded cells were taken by an Axio Zoom.V16 microscope (Zeiss). The assay was run in triplicate, in biologic replicate.

In vivo analyses of in vitro–derived cell fusion hybrids

For tumor growth, 8-to 12-week-old C57BL/6J mice (Jackson Laboratories) were injected with 5×10^4 cells (MC38 and MC38-derived hybrids) or 5×10^5 cells (B16F10 and B16F10-derived hybrids) subcutaneously or intradermally, respectively. Length (L) and width (W) of palpable tumors were measured three times weekly with calipers until tumors reached a maximum diameter of 2 cm. Tumors were surgically removed in survival surgery, or animals were sacrificed during tumor removal in accordance with OHSU Institutional Animal Care and Use Committee guidelines. Animals were observed for at least 6 months for detection of tumor growth. For each tumor, volume (V) was calculated by the formula $V = \frac{1}{2} (L \times W^2)$; volume doubling time for each tumor was extracted from a curve fit to a plot of log tumor volume over time. Curves with R^2 values of less than 0.8 were excluded from analysis, as were tumors with six or fewer dimension measurements; these exclusion criteria were established in response to the unanticipated early ulceration of some

tumors, which precluded accurate measurements of length and width ($P < 0.05$, by Mann-Whitney U test). For growth of tumor at metastatic sites, 1×10^6 MC38 cells were injected into the spleen. Livers were analyzed 3 weeks later for tumor burden by H&E stain. Hybrids formed metastatic foci more readily with a $P < 0.008$ by Mann-Whitney U test. MC38 ($n = 17$) and MC38-derived hybrids ($n = 13$) were injected in four different technical replicate experiments. For B16F10 cells, 2.5×10^5 cells were retro-orbitally injected, and lungs were analyzed 16 days after injection. Melanin-marked tumor metastasis was visualized. Tumor burden was analyzed on paraffin-embedded tissue sections located every 100 μm apart through the entire lung ($n = 5$ tissue sections per lung). Metastatic foci areas were measured using an Aperio ImageScope V12.3.0.5056 to outline metastatic tumors and quantify area. A nonparametric t test was performed. Duplicate studies of B16F10 and B16F10-derived hybrids ($n = 12$ mice) were analyzed.

In vivo–derived cell fusion hybrids
For isolation of in vivo–derived hybrids or assessment of CTCs, 5×10^5 B16F10 (H2B-RFP with or without Cre) cells were injected intradermally into R26R-YFP or actin-GFP mice, respectively. Once tumors reached 1 to 2 cm^3 in diameter, they were surgically removed for immunohistochemical analyses or for FACS/flow analyses.

For demonstration that tumor cells can fuse with myeloid cells, 5×10^5 B16F10 (fl-dsRed-fl-eGFP) cells were injected intradermally into 6- to 8-week-old LysM-Cre transgenic mice. When tumors reached 1 cm^3, primary tumors and lungs were removed for immunohistochemical analyses.

Immunohistochemical analysis of in vivo–derived tumors
B16F10 (H2B-RFP, Cre) primary tumors in Act-GFP or R26R-stop-YFP mice were fixed in 10% buffered formalin, frozen in optimum cutting temperature (OCT), and 5-μm sections were obtained. Tumors from R26R-stop-YFP mice were incubated with antibodies for GFP (1:500; Life Technologies) followed by detection with fluorescent secondary antibody (1:500, Alexa Fluor 488; Jackson ImmunoResearch). Nuclei were counterstained with Hoechst (1 μg/ml). Slides were digitally scanned with a Leica DM6000 B microscope and analyzed using Ariol software. Confocal images were acquired with a FluoView FV1000 confocal microscope (Olympus).

B16F10 (fl-dsRed-fl-eGFP) primary tumors and lungs from LysM-Cre mice were fixed in 4% paraformaldehyde for 2 hours at 20°C, washed, and cryopreserved in 30% sucrose for 16 hours at 4°C and then embedded in OCT. Primary tumors were stained as described in the paragraph above. Lung sections were cut to 8-μm thickness, baked for 30 min at 37°C, then subjected to antigen retrieval under standard conditions (R&D Systems, CTS016), blocked with DAKO Protein Block Serum-Free (Agilent, X090930-2), and incubated for 16 hours at 4°C with primary antibodies [anti-MITF (1:500; Abcam, ab12039), anti-dsRed (1:250; Clontech, 632496), and anti-GFP (1:1000; Abcam, ab13970)] in background-reducing antibody diluent (Agilent, S302281-2). Fluorescent-tagged secondary antibodies were applied, and then sections were mounted in a ProLong Gold antifade reagent (Molecular Probes, P36934). Antibody specificity was determined by immunostaining healthy lungs of nontumor-bearing mice and performing secondary antibody only controls.

FACS isolation and flow cytometric analyses of fusion hybrids
Tumors were diced and digested for 30 min at 37°C in DMEM + Collagenase A (2 mg/ml; Roche) + DNase (Roche) under stirring conditions. Digested tumor cells were filtered through a 40-μm filter

and washed with PBS. For FACS isolation, hybrid and unfused cells were isolated by direct fluorescence on a Becton Dickinson InFlux sorter. For flow cytometric analysis, blood was collected retro-orbitally using heparinized microhematocrit capillary tubes (Fisher) into K$_2$EDTA-coated tubes (BD Biosciences). RBC lysis was performed as described above. Cells were washed and resuspended in FACS buffer (PBS, 1.0 mM EDTA, and 5% FBS). Cells were incubated in PBS containing LIVE/DEAD Fixable Aqua (1:500; Invitrogen) with Fc Receptor Binding Inhibitor (1:200; eBioscience). Cells were then incubated in FACS buffer for 30 min with CD45-PeCy7 (1:8000; BioLegend), CSF1R-BV711 (1:200; BioLegend), F4/80-APC (1:400; BioLegend), and CD11b-AF700 (1:200; eBioscience). A BD LSRFortessa FACS machine was used for analyses. A statistical significance of $P < 2.2 \times 10^{-6}$ by unpaired t test was determined for CD45$^+$ hybrid CTCs relative to CD45$^-$ hybrid, CD45$^+$ unfused, and CD45$^-$ unfused CTCs. Technical duplicates of $n = 5$ or 6 mice were analyzed.

Tumorigenic analyses of FACS-isolated in vivo–derived hybrids
A total of 100 or 3000 FACS-isolated hybrids and unfused B16F10 cells were injected intradermally into C57BL/6J mice. For experiments with 100 cells, technical octuplicates with biologic duplicates, triplicates, or quadruplicates were performed, depending on the number of hybrids isolated form the primary tumor, for a total of $n = 16$ mice analyzed. For experiments with 3000 cells injected, technical triplicates were performed.

Statistical analyses and graphical displays
Dot plots, bar charts, and line charts were generated in GraphPad Prism or Excel. GraphPad Prism and Excel were also used for statistical analyses of these data, including ensuring that data met assumptions of the tests used and comparisons of variance between groups when appropriate. Microsoft Excel was used to perform two-tailed t tests. A three-dimensional scatterplot was generated in R using the rgl package. Flow cytometry data were prepared for display using FlowJo software. Microarray gene expression data were displayed as a heatmap prepared using Genesifter software. Heatmap of MEMA data was generated in R using the standard heatmap function and default parameters.

REFERENCES AND NOTES
1. G. H. Heppner, Tumor heterogeneity. *Cancer Res.* **44**, 2259–2265 (1984).
2. A. Marusyk, K. Polyak, Tumor heterogeneity: Causes and consequences. *Biochim. Biophys. Acta* **1805**, 105–117 (2010).

3. A. Z. Rizvi, J. R. Swain, P. S. Davies, A. S. Bailey, A. D. Decker, H. Willenbring, M. Grompe, W. H. Fleming, M. H. Wong, Bone marrow-derived cells fuse with normal and transformed intestinal stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6321–6325 (2006).

4. P. S. Davies, A. E. Powell, J. R. Swain, M. H. Wong, Inflammation and proliferation act together to mediate intestinal cell fusion. *PLOS ONE* **4**, e6530 (2009).

5. M. Alvarez-Dolado, R. Pardal, J. M. Garcia-Verdugo, J. R. Fike, H. O. Lee, K. Pfeffer, C. Lois, S. J. Morrison, A. Alvarez-Buylla, Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. *Nature* **425**, 968–973 (2003).

6. F. D. Camargo, R. Green, Y. Capetanaki, K. A. Jackson, M. A. Goodell, Single hematopoietic stem cells generate skeletal muscle through myeloid intermediates. *Nat. Med.* **9**, 1520–1527 (2003).

7. N. Terada, T. Hamazaki, M. Oka, M. Hoki, D. M. Mastalerz, Y. Nakano, E. M. Meyer, L. Morel, B. E. Petersen, E. W. Scott, Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416**, 542–545 (2002).

8. X. Wang, H. Willenbring, Y. Akkari, Y. Torimaru, M. Foster, M. Al-Dhalimy, E. Lagasse, M. Finegold, S. Olson, M. Grompe, Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422**, 897–901 (2003).

9. A. Carter, Cell fusion theory: Can it explain what triggers metastasis? *J. Natl. Cancer Inst.* **100**, 1279–1281 (2008).

10. J. M. Pawelek, Tumour-cell fusion as a source of myeloid traits in cancer. *Lancet Oncol.* **6**, 988–993 (2005).

11. T. Dittmar, K. S. Zänker, Tissue regeneration in the chronically inflamed tumor environment: Implications for cell fusion driven tumor progression and therapy resistant tumor hybrid cells. *Int. J. Mol. Sci.* **16**, 30362–30381 (2015).

12. A. E. Powell, E. C. Anderson, P. S. Davies, A. D. Silk, C. Pelz, S. Impey, M. H. Wong, Fusion between intestinal epithelial cells and macrophages in a cancer context results in nuclear reprogramming. *Cancer Res.* **71**, 1497–1505 (2011).

13. R. Lazova, G. S. LaBerge, E. Duvall, N. Spoelstra, V. Klump, M. Sznol, D. Cooper, R. A. Spritz, J. T. Chang, J. M. Pawelek, A melanoma brain metastasis with a donor-patient hybrid genome following bone marrow transplantation: First evidence for fusion in human cancer. *PLOS ONE* **8**, e66731 (2013).

14. D. L. Adams, S. S. Martin, R. K. Alpaugh, M. Charpentier, S. Tsai, R. C. Bergan, I. M. Ogden, W. Catalona, S. Chumsri, C.-M. Tang, M. Cristofanilli, Circulating giant macrophages as a potential biomarker of solid tumors. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3514–3519 (2014).

15. C. R. Cogle, N. D. Theise, D. Fu, D. Ucar, S. Lee, S. M. Guthrie, J. Lonergan, W. Rybka, D. S. Krause, E. W. Scott, Bone marrow contributes to epithelial cancers in mice and humans as developmental mimicry. *Stem Cells* **25**, 1881–1887 (2007).

16. M. Grimm, O. Feyen, J. F. Coy, H. Hofmann, P. Teriete, S. Reinert, Analysis of circulating CD14+/CD16+ monocyte-derived macrophages (MDMs) in the peripheral blood of patients with oral squamous cell carcinoma. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **121**, 301–306 (2016).

17. G. S. LaBerge, E. Duvall, Z. Grasmick, K. Haedicke, J. Pawelek, A melanoma lymph node metastasis with a donor-patient hybrid genome following bone marrow transplantation: A second case of leucocyte-tumor cell hybridization in cancer metastasis. *PLOS ONE* **12**, e0168581 (2017).

18. I. Shabo, H. Olsson, X. F. Sun, J. Svanvik, Expression of the macrophage antigen CD163 in rectal cancer cells is associated with early local recurrence and reduced survival time. *Int. J. Cancer* **125**, 1826–1831 (2009).

19. Y. Yilmaz, R. Lazova, M. Qumsiyeh, D. Cooper, J. Pawelek, Donor Y chromosome in renal carcinoma cells of a female BMT recipient: Visualization of putative BMT-tumor hybrids by FISH. *Bone Marrow Transplant.* **35**, 1021–1024 (2005).

20. M. Okabe, M. Ikawa, K. Kominami, T. Nakanishi, Y. Nishimune, 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).

21. S. Srinivas, T. Watanabe, C.-S. Lin, C. M. William, Y. Tanabe, T. M. Jessell, F. Costantini, Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).

22. M. J. Bissell, W. C. Hines, Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nat. Med.* **17**, 320–329 (2011).

23. D. Hoshino, K. M. Branch, A. M. Weaver, Signaling inputs to invadopodia and podosomes. *J. Cell Sci.* **126** (Pt. 14), 2979–2989 (2013).

24. J. Massagué, TGFβ in cancer. *Cell* **134**, 215–230 (2008).

25. C.-H. Lin, J. K. Lee, M. A. LaBarge, Fabrication and use of MicroEnvironment microArrays (MEArrays). *J. Vis. Exp.*, 4152 (2012).

26. C. J. Sherr, C. W. Rettenmier, The fms gene and the CSF-1 receptor. *Cancer Surv.* **5**, 221–232 (1986).

27. D. G. DeNardo, D. J. Brennan, E. Rexhepaj, B. Ruffell, S. L. Shiao, S. F. Madden, W. M. Gallagher, N. Wadhwani, S. D. Keil, S. A. Junaid, H. S. Rugo, E. S. Hwang, K. Jirström, B. L. West, L. M. Coussens, Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy. *Cancer Discov.* **1**, 54–67 (2011).

28. T. Fehm, A. Sagalowsky, E. Clifford, P. Beitsch, H. Saboorian, D. Euhus, S. Meng, L. Morrison, T. Tucker, N. Lane, B. M. Ghadimi, K. Heselmeyer-Haddad, T. Ried, C. Rao,
J. Uhr, Cytogenetic evidence that circulating epithelial cells in patients with carcinoma are malignant. *Clin. Cancer Res.* **8**, 2073–2084 (2002).

29. E. Racila, D. Euhus, A. J. Weiss, C. Rao, J. McConnell, L. W. M. M. Terstappen, J. W. Uhr, Detection and characterization of carcinoma cells in the blood. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4589–4594 (1998).

30. A. D. Silk, C. E. Gast, P. S. Davies, F. D. Fakhari, G. E. Vanderbeek, M. Mori, M. H. Wong, Fusion between hematopoietic and epithelial cells in adult human intestine. *PLOS ONE* **8**, e55572 (2013).

31. A. Lorico, D. Corbeil, J. M. Pawelek, R. Alessandro, Transmission of information in neoplasia by extracellular vesicles. *Biomed. Res. Int.* **2015**, 289567 (2015).

32. R. Lazova, A. Chakraborty, J. M. Pawelek, Leukocyte-cancer cell fusion: Initiator of the warburg effect in malignancy? *Adv. Exp. Med. Biol.* **714**, 151–172 (2011).

33. R. Halaban, J. Nordlund, U. Francke, G. Moellmann, J. M. Eisenstadt, Supermelanotic hybrids derived from mouse melanomas and normal mouse cells. *Somatic Cell Genet.* **6**, 29–44 (1980).

34. R. S. Kerbel, A. E. Lagarde, J. W. Dennis, T. P. Donaghue, Spontaneous fusion in vivo between normal host and tumor cells: Possible contribution to tumor progression and metastasis studied with a lectin-resistant mutant tumor. *Mol. Cell. Biol.* **3**, 523–538 (1983).

35. M. Lizier, A. Anselmo, S. Mantero, F. Ficara, M. Paulis, P. Vezzoni, F. Lucchini, G. Pacchiana, Fusion between cancer cells and macrophages occurs in a murine model of spontaneous *neu*+ breast cancer without increasing its metastatic potential. *Oncotarget* **7**, 60730–60806 (2016).

36. X. Lu, Y. Kang, Cell fusion as a hidden force in tumor progression. *Cancer Res.* **69**, 8536–8539 (2009).

37. I. Shabo, K. Midtbö, H. Andersson, E. Åkerlund, H. Olsson, P. Wegman, C. Gunnarsson, A. Lindström, Macrophage traits in cancer cells are induced by macrophage-cancer cell fusion and cannot be explained by cellular interaction. *BMC Cancer* **15**, 922 (2015).

38. A. Patsialou, Y. Wang, J. Pignatelli, X. Chen, D. Entenberg, M. Oktay, J. S. Condeelis, Autocrine CSF1R signaling mediates switching between invasion and proliferation downstream of TGFβ in claudin-low breast tumor cells. *Oncogene* **34**, 2721–2731 (2015).

39. Y. Mo, O. Giricz, C. Hu, K. Dahlman, S. Bhattacharyya, H. Nguyen, B. Matusow, T. Bhagat, Y. Yu, R. Shellooe, E. Burton, J. Habets, J. Greally, P. Kenny, J. Sosman, G. Bollag, B. L. West, A. Verma, Integrated epigenomic profiling reveals widespread demethylation in melanoma, and reveals aberrant CSF-1 receptor expression as a regulator of malignant growth and invasion inhibited by PLX3397. *Cancer Res.* **73**, Abstract A26 (2013).

40. W. Sheng, O. O. Ogunwobi, T. Chen, J. Zhang, T. J. George, C. Liu, Z. H. Fan, Capture, release and culture of circulating tumor cells from pancreatic cancer patients using an enhanced mixing chip. *Lab Chip* **14**, 89–98 (2014).

41. G. A. Clawson, E. Kimchi, S. D. Patrick, P. Xin, R. Harouaka, S. Zheng, A. Berg, T. Schell, K. F. Staveley-O'Carroll, R. I. Neves, P. J. Mosca, D. Thiboutot, Circulating tumor cells in melanoma patients. *PLOS ONE* **7**, e41052 (2012).

42. G. A. Clawson, G. L. Matters, P. Xin, C. McGovern, E. Wafula, C. dePamphilis, M. Meckley, J. Wong, L. Stewart, C. D'Jamoos, N. Altman, Y. Imamura Kawasawa, Z. Du, L. Honaas, T. Abraham, "Stealth dissemination" of macrophage-tumor cell fusions cultured from blood of patients with pancreatic ductal adenocarcinoma. *PLOS ONE* **12**, e0184451 (2017).

43. G. Rappa, J. Mercapide, A. Lorico, Spontaneous formation of tumorigenic hybrids between breast cancer and multipotent stromal cells is a source of tumor heterogeneity. *Am. J. Pathol.* **180**, 2504–2515 (2012).

44. M.-H. Xu, X. Gao, D. Luo, X.-D. Zhou, W. Xiong, G.-X. Liu, EMT and acquisition of stem cell-like properties are involved in spontaneous formation of tumorigenic hybrids between lung cancer and bone marrow-derived mesenchymal stem cells. *PLOS ONE* **9**, e87893 (2014).

45. J. R. McGill, P. J. Lalley, R. J. Leach, T. J. Johnson, D. D. Von Hoff, Chromosomal influence on hybrid cell proliferation. *Cell Prolif.* **25**, 345–355 (1992).

46. M. Q. Islam, L. da S. Meirelles, N. B. Nardi, P. Magnusson, K. Islam, Polyethylene glycol-mediated fusion between primary mouse mesenchymal stem cells and mouse fibroblasts generates hybrid cells with increased proliferation and altered differentiation. *Stem Cells Dev.* **15**, 905–919 (2006).

47. J. Xue, Y. Zhu, Z. Sun, R. Ji, X. Zhang, W. Xu, X. Yuan, B. Zhang, Y. Yan, L. Yin, H. Xu, L. Zhang, W. Zhu, H. Qian, Tumorigenic hybrids between mesenchymal stem cells and gastric cancer cells enhanced cancer proliferation, migration and stemness. *BMC Cancer* **15**, 793 (2015).

48. J. Y. Hung, D. Horn, K. Woodruff, T. Prihoda, C. LeSaux, J. Peters, F. Tio, S. L. Abboud-Werner, Colony-stimulating factor 1 potentiates lung cancer bone metastasis. *Lab. Invest.* **94**, 371–381 (2014).

49. J. B. Mitchem, D. J. Brennan, B. L. Knolhoff, B. A. Belt, Y. Zhu, D. E. Sanford, L. Belaygorod, D. Carpenter, L. Collins, D. Piwnica-Worms, S. Hewitt, G. M. Udupi, W. M. Gallagher, C. Wegner, B. L. West, A. Wang-Gillam, P. Goedegebuure, D. C. Linehan, D. G. DeNardo, Targeting tumor-infiltrating macrophages decreases tumor-initiating cells, relieves immunosuppression, and improves chemotherapeutic responses. *Cancer Res.* **73**, 1128–1141 (2013).

50. S. F. Ngiow, K. M. Meeth, K. Stannard, D. S. Barkauskas, G. Bollag, M. Bosenberg, M. J. Smyth, Co-inhibition of colony stimulating factor-1 receptor and BRAF oncogene in mouse models of BRAF^{V600E} melanoma. *Oncoimmunology* **5**, e1089381 (2016).

51. B. Ruffell, L. M. Coussens, Macrophages and therapeutic resistance in cancer. *Cancer Cell* **27**, 462–472 (2015).

52. B. E. Clausen, C. Burkhardt, W. Reith, R. Renkawitz, I. Förster, Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8**, 265–277 (1999).

53. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).

54. W. Walter, F. Sánchez-Cabo, M. Ricote, GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914 (2015).

55. J. Burkhart, *Cell Fusion Potentiated Acquisition of Macrophage-like Behavior in Cancer cells and Its Contribution* (Oregon Health & Science University, 2016).

56. D. F. Seals, E. F. Azucena Jr., I. Pass, L. Tesfay, R. Gordon, M. Woodrow, J. H. Resau, S. A. Courtneidge, The adaptor protein Tks5/Fish is required for podosome formation and function, and for the protease-driven invasion of cancer cells. *Cancer Cell* **7**, 155–165 (2005).

CANCER

Mechanism of FACT removal from transcribed genes by anticancer drugs curaxins

Han-Wen Chang[1], Maria E. Valieva[2,3], Alfiya Safina[4], Răzvan V. Chereji[5], Jianmin Wang[6], Olga I. Kulaeva[1], Alexandre V. Morozov[7], Mikhail P. Kirpichnikov[2,8], Alexey V. Feofanov[2,8], Katerina V. Gurova[4], Vasily M. Studitsky[1,2]*

Human FACT (facilitates chromatin transcription) is a multifunctional protein complex that has histone chaperone activity and facilitates nucleosome survival and transcription through chromatin. Anticancer drugs curaxins induce FACT trapping on chromatin of cancer cells (c-trapping), but the mechanism of c-trapping is not fully understood. Here, we show that in cancer cells, FACT is highly enriched within the bodies of actively transcribed genes. Curaxin-dependent c-trapping results in redistribution of FACT from the transcribed chromatin regions to other genomic loci. Using a combination of biochemical and biophysical approaches, we have demonstrated that FACT is bound to and unfolds nucleosomes in the presence of curaxins. This tight binding to the nucleosome results in inhibition of FACT-dependent transcription in vitro in the presence of both curaxins and competitor chromatin, suggesting a mechanism of FACT trapping on bulk nucleosomes (n-trapping).

INTRODUCTION

Histone chaperone FACT (facilitates chromatin transcription) is involved in DNA transcription (*1–6*), replication (*7–10*) and repair (*11–15*), cell differentiation, and cancer development (*16–19*) [reviewed in (*20*)]. Human FACT (hFACT) is a heterodimer composed of two proteins: SPT16 (suppressor of Ty16) and SSRP1 (structure-specific recognition protein 1) (*21*). FACT preferentially interacts with the histone H2A/H2B dimer and also binds the H3/H4 tetramer and DNA (*21–26*). By interacting with different targets in the nucleosome, yeast FACT has the ability to reorganize the nucleosomal structure (*24, 27, 28*); this activity is likely important for transcription initiation (*24, 28–30*). hFACT facilitates transcription through chromatin by RNA polymerase II (Pol II) in vitro and nucleosome survival during this process (*1, 2, 5*) by transiently interacting with the DNA binding surface of the H2A/H2B dimers (*2, 5, 6*); it also facilitates nucleosome survival during transcription in cancer cells (*31*). In cancer and stem cells, the expression of FACT is up-regulated so that it correlates with the malignancy of tumor cells (*16–19, 32–34*).

Recently, FACT has been reported as the potential target for anticancer compounds curaxins (*17, 33–35*) that intercalate into DNA without causing DNA damage (*17, 36, 37*). After treatment of cancer cells with curaxins, FACT is redistributed within the cell nuclei and becomes tightly bound to chromatin [c-trapped (*17*)]; there is a correlation between the ability to cause c-trapping and cytotoxicity of various curaxins (*37*). Treatment of cancer cells with curaxins is also accompanied by activation of p53, inhibition of nuclear factor κB–(NF-κB) and hypoxia-inducible factor 1α (HIF1α)–dependent transcription, and cancer cell death (*17, 34*). However, the mechanism of c-trapping upon treatment by curaxins is not fully understood.

In this work, the mechanism of c-trapping was analyzed using a combination of single-particle Förster resonance energy transfer (spFRET), biochemical, and genomic approaches. Genomic data suggest that in cancer cells, curaxins induce redistribution of FACT from transcribed chromatin to other genomic loci. Our in vitro studies indicate that curaxins induce FACT binding to nucleosomes that, in turn, results in nucleosome unfolding. Together, our data suggest that curaxins create multiple "false targets" for FACT, thus inducing FACT redistribution from transcribed genes to other chromatin regions.

RESULTS

Experimental rationale

FACT facilitates transcription in vitro by interacting with the histone surfaces transiently exposed during Pol II transcription (*5*). Because curaxins are DNA intercalators (*36*) and induce FACT relocation (c-trapping), we hypothesized that curaxins could induce changes in nucleosome structure and thus create high-affinity sites for FACT binding and induce genome-wide redistribution of FACT. This hypothesis makes the following predictions: (i) Curaxins create multiple sites for FACT binding genome-wide and therefore deplete FACT from highly transcribed genes. (ii) FACT is more tightly associated with bulk nucleosomes in the presence of curaxins. (iii) Curaxins inhibit FACT action during transcription through a nucleosome only in the presence of competitor chromatin. Below, we systematically evaluate these predictions. Throughout the study, we used curaxin CBL0137. This curaxin has been selected as our primary experimental model because, in preclinical studies, it can strongly inhibit growth of various, including drug-resistant, cancers (*35, 38*). Some in vitro data were confirmed with a structurally similar curaxin, CBL0100.

Curaxins induce redistribution of FACT from transcribed chromatin to other genomic loci

hFACT facilitates Pol II transcription through chromatin in vitro (*1, 2, 5*), raising the possibility that curaxins affect association of FACT with Pol II–transcribed chromatin in cancer cells. To obtain initial clues on colocalization of FACT and transcribing Pol II, we

[1]Fox Chase Cancer Center, Philadelphia, PA 19111, USA. [2]Biology Faculty, Lomonosov Moscow State University, 119992 Moscow, Russia. [3]Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany. [4]Department of Cell Stress Biology, Roswell Park Cancer Institute, Buffalo, NY 14263, USA. [5]Eunice Kennedy Shriver National Institute for Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [6]Department of Bioinformatics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA. [7]Department of Physics and Astronomy and Center for Quantitative Biology, Rutgers University, Piscataway, NJ 08854, USA. [8]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 117997 Moscow, Russia.
*Corresponding author. Email: vasily.studitsky@fccc.edu

have analyzed immunofluorescence of SSRP1 and transcribing Pol II using anti-SSRP1 and anti–Pol II S2-P antibodies, respectively, before and after curaxin treatment (fig. S1). Because FACT is an abundant protein localized in the nuclei, it was difficult to detect colocalization of the small fraction of FACT bound to chromatin in the absence of curaxins with transcribing Pol II using an immunofluorescence approach (fig. S1). However, after curaxin treatment, Pol II and FACT were localized within different chromatin regions (fig. S1).

To evaluate the effect of curaxins on FACT association with transcribed genes in more detail, the distribution of FACT in the genome of HT1080 cancer cells was studied using chromatin immunoprecipitation with anti-SSRP1 antibodies followed by next-generation sequencing (ChIP-seq) before or after treatment with curaxin CBL0137 (19). FACT association with transcribed and nontranscribed genomic regions was notably reduced or increased, respectively, upon treatment with CBL0137 (Fig. 1, A and B). The total amount of FACT in CBL0137-treated cells was not changed [fig. S2 and (17)]. The total amount of chromatin-bound FACT was substantially increased (figs. S2 and S3). The SSRP1 ChIP-seq data were compared with nascent RNA sequencing (RNA-seq) data (37) to determine FACT occupancy in the transcribed regions of the genome before and after curaxin treatment. FACT is highly enriched on genes actively transcribed in untreated cells, especially on the 5% of most actively transcribed genes ($n = 653$; Fig. 1C and fig. S4). Consistently, there is a strong correlation between the extent of FACT enrichment and levels of transcription for the 500 highly transcribed genes in untreated cells (Fig. 1, D and E, and figs. S5 and S6). The depletion of SSRP1 upon CBL0137 treatment preferentially occurs at actively transcribed genes (Fig. 1 and figs. S4 to S6). Thus, general redistribution of FACT from transcribed to nontranscribed chromatin regions occurs upon treatment of the cells with curaxin CBL0137. It would be expected that global changes in the chromatin structure occur soon after curaxin treatment, and this is indeed the case (36).

FACT and curaxins strongly, synergistically, and reversibly uncoil nucleosomal DNA

The effect of curaxins on the nucleosome structure and FACT-nucleosome interaction was studied using the spFRET in solution in vitro. Precisely positioned nucleosomes assembled on a DNA template containing the 603 nucleosome positioning sequence (NPS) (39, 40) and fluorescent labels Cy3 and Cy5 at the +13 and +91 positions on nucleosomal DNA (from the 5′ end of nontemplate DNA strand of nucleosomal DNA; Fig. 2A), respectively, were gel purified (fig. S7). The donor fluorophore (Cy3) was excited by a 514.5-nm wavelength laser (28). The nucleosome complexes were diffused in solution, and only one particle of the nucleosome complex traversed the focal volume of the microscope at any given time. The fluorescence intensities of both donor (Cy3) and acceptor (Cy5) dyes were then measured, and the proximity ratios (E_{PR}) that reveal the changes of distances between labeled DNA sites through changes in FRET efficiency were calculated (28). The data obtained using the fluorescently labeled nucleosome are typically described by two Gaussian peaks: a minor peak that corresponds to histone-free DNA in the solution and a major peak representing the intact nucleosome (Fig. 2, B and C). The signals obtained from control nucleosomes are stable during the experimental time periods (fig. S8).

FACT used in our studies has been purified from insect cells and therefore is likely phosphorylated and unable to bind nucleosomes

(22, 25, 26, 41, 42); this is the state of bulk FACT complexes in cancer cells (6, 22, 26). Adding either CBL0137 or FACT alone to nucleosomes results in only minimal changes of E_{PR} (Fig. 2B). In contrast, adding CBL0137 and FACT together results in a marked increase of the peak corresponding to the histone-free DNA and a corresponding decrease of the peak corresponding to the intact nucleosome (Fig. 2C). The data indicate that in the presence of CBL0137, FACT binds to nucleosomes, resulting in FACT trapping on nucleosomes (n-trapping). Thus, the data suggest that FACT binds to nucleosomes only in the presence of curaxins and induces a strong, partially reversible nucleosome unfolding or uncoiling of the nucleosomal DNA (Fig. 2D). Because FACT alone does not bind nucleosomes, the affinity of FACT to nucleosomes is strongly increased in the presence of curaxins. It should be noted that the term n-trapping was used previously to describe the interaction of FACT with the hexasomes (a nucleosome missing one H2A/H2B dimer) (36).

The changes in E_{PR} are partially reversed by adding an excess of unlabeled histone-free DNA after the formation of FACT-nucleosome complex unfolded in the presence of curaxins to remove FACT (and likely some curaxins bound to nucleosomal DNA) from the complexes (28). The reversibility indicates that in the presence of CBL0137 and FACT, the nucleosome is either unfolded or the nucleosomal DNA is uncoiled without notable accompanying dissociation of the majority of core histones from the nucleosomal DNA. Any histone loss from the nucleosomes after FACT-dependent uncoiling is likely to be irreversible (43), especially in the presence of an excess competitor DNA. Even if nucleosomes would reform after the reversal of FACT binding, they will preferably form on the competitor DNA. Consistently, our previous experiments have shown that nucleosome reorganization by yeast FACT is accompanied by only minimal (typically 5 to 10%) loss of the dimers (28). The reversibility is incomplete, and the low E_{PR} peak is slightly increased as compared to the control nucleosomes (Fig. 2C), likely because some nucleosomes are displaced by FACT in the presence of CBL0137 (see below) and some nucleosomes contain DNA that is partially uncoiled from the octamer in the presence of curaxins.

Next, the effect of curaxins on FACT-nucleosome interaction was analyzed in vitro using a gel shift assay. The nucleosome cores were assembled on the ^{32}P-labeled 147–base pair (bp) DNA fragment containing the 603 NPS. Gel-purified nucleosomes contain less than 5% histone-free DNA (fig. S7). After incubation in the presence of FACT and/or curaxins CBL0137 or CBL0100, the samples were incubated either with or without an excess of unlabeled DNA competitor and analyzed by native polyacrylamide gel electrophoresis (PAGE) (Fig. 3A). FACT, competitor histone-free DNA, or curaxins alone minimally affect the nucleosomes (Fig. 3, B and C). Nucleosomes are slightly more efficiently disrupted after incubation with CBL0100 than after incubation with CBL0137 (Fig. 3, B and C), probably because CBL0100 is a more potent DNA intercalator than CBL0137 (17, 36).

In contrast, nucleosomes incubated in the presence of both FACT and curaxins are very unstable, with approximately 70% of the templates disrupted during gel electrophoresis in the absence of competitor DNA. This effect is largely reversed by adding the excess of competitor DNA after the reaction (Fig. 3, B and C), indicating that nucleosomes incubated in the presence of FACT and curaxins are stable in solution, but unstable during electrophoresis. Apparently, removal of FACT from the nucleosomes in the presence of competitor DNA in solution results in recovery of nearly intact nucleosomes that are stable during electrophoresis. Thus, consistent with the results
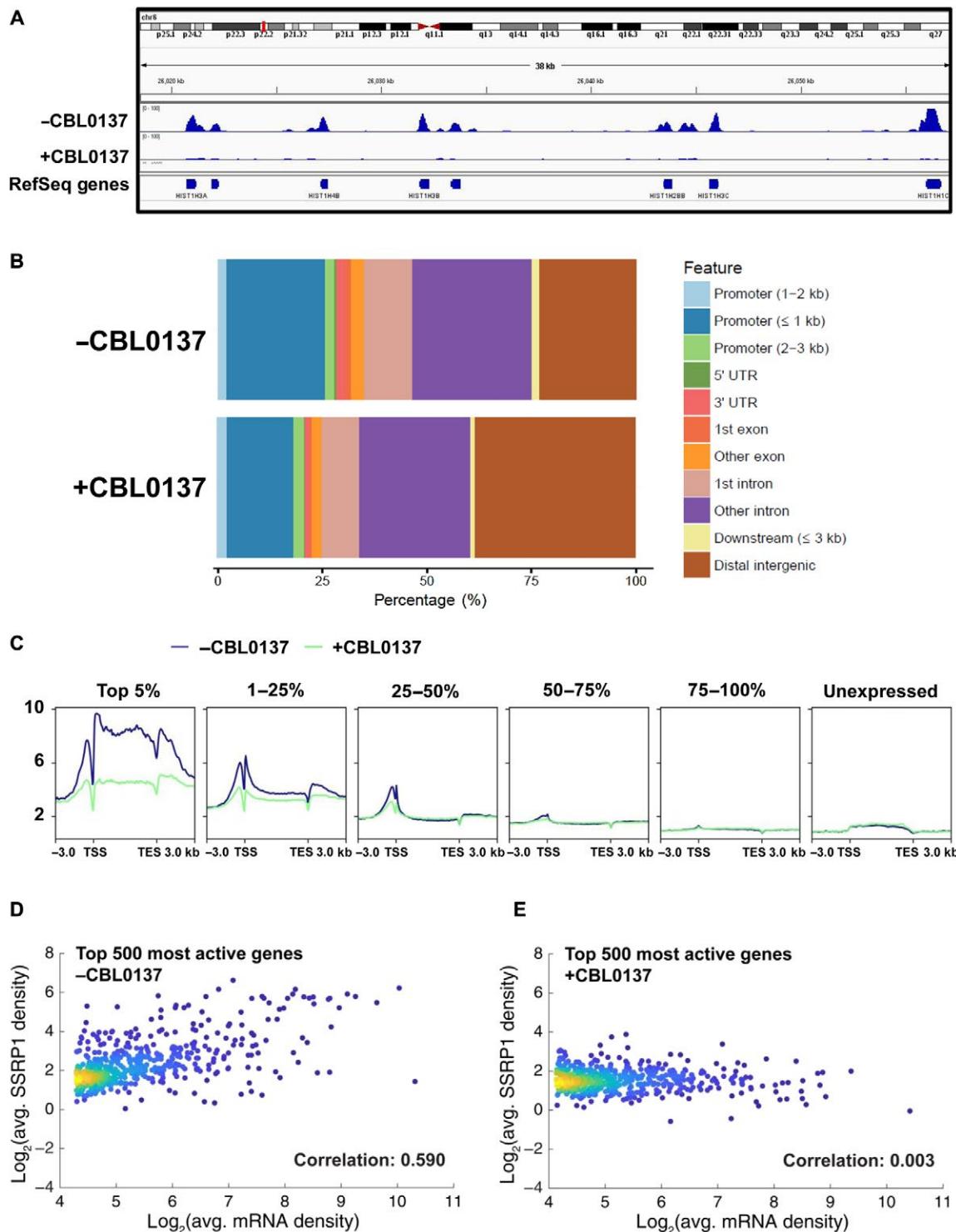
Fig. 1. FACT subunit SSRP1 is preferentially enriched on highly transcribed genes, and curaxins remove SSRP1 from the gene bodies. (**A**) Integrated genome views of the SSRP1 distributions at selected regions of chromosome 6 of the human genome. (**B**) Distribution of SSRP1 peaks in control and CBL0137-treated cells in relation to genome annotation features. (**C**) SSRP1 protein is preferentially enriched on highly transcribed genes. Average SSRP1 occupancy near the transcription start sites (TSSs), in the transcribed regions, and in transcription end sites (TESs) of the genes in HT1080 cells (incubated in the absence or presence of 3 μM curaxin CBL0137) was determined using ChIP-seq. The genes are grouped on the basis of the RNA-seq data. The corresponding heatmaps are shown in fig. S4. (**D** and **E**) Curaxins preferentially remove SSRP1 from actively transcribed genes. Density scatter plots representing average SSRP1 densities over gene bodies against the level of transcription of the corresponding genes, quantified by the average level of nascent transcripts determined in HT1080 cells by NET-seq (native elongating transcript sequencing). Analysis of all (fig. S5) or top 500 most actively transcribed genes (D and E) in the absence (D) or presence (E) of the curaxin. For the top 500 most active genes, the Pearson correlation coefficient is markedly decreased (from 0.590 to 0.003) upon treatment with curaxin.
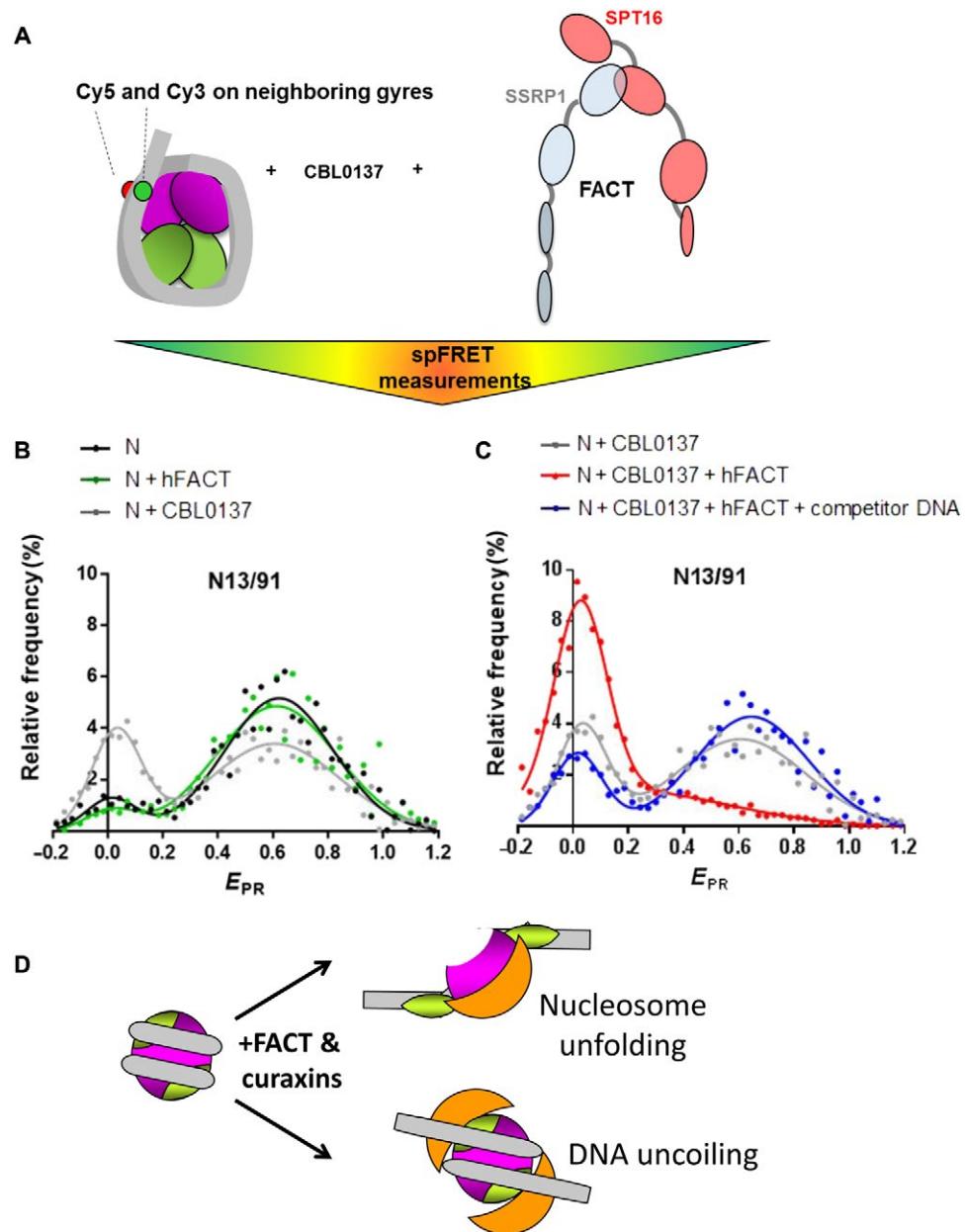
Fig. 2. FACT and curaxins strongly and synergistically affect the nucleosomal structure: Analysis by spFRET. (**A**) Experimental approach. The mononucleosomes contained the single pair of Cy3 and Cy5 dyes on the nucleosomal DNA (the positions of Cy3 and Cy5 are shown by green and red circles, respectively). spFRET from nucleosomes was measured in the absence or presence of curaxin CBL0137, FACT, and/or competitor DNA. (**B** and **C**) Typical frequency distributions of FRET efficiencies (E_{PR}). Analysis by spFRET microscopy. Sample sizes and other numerical parameters are shown in table S1. (B) Only minor changes in nucleosome structure are detected in the presence of either CBL0137 or FACT. (C) FACT and CBL0137 together induce marked and partially reversible uncoiling of the nucleosomal DNA. The uncoiling is partially reversed by subsequent addition of an excess of competitor DNA, resulting in removal of FACT from the complex. (**D**) Possible changes in the nucleosome structure in the presence of FACT and curaxins: nucleosome unfolding or uncoiling of nucleosomal DNA from the histone octamer.

of the spFRET studies (Fig. 2), biochemical studies suggest that nucleosomes are destabilized in the presence of FACT and curaxins, but the majority of these nucleosomes are recovered in the presence of an excess of unlabeled DNA competitor that removes FACT from the FACT-nucleosome complexes (Fig. 3).

Together, the data suggest that in the presence of curaxins, FACT binds to nucleosomes (n-trapped), causing a considerable change in the nucleosome structure that is partially reversible and is not accompanied by the loss of core histones (Figs. 2 and 3). The FACT-nucleosome complex formed in the presence of curaxins can be disrupted in the presence of competitor DNA or during electrophoresis.

Curaxins and competitor nucleosomes synergistically inhibit FACT-dependent transcription in vitro
To evaluate the effect of n-trapping on FACT-dependent transcription, a well-established in vitro Pol II transcription system containing the purified yeast Pol II and nucleosomes or corresponding DNA templates containing the 603 NPS was used [Fig. 4A
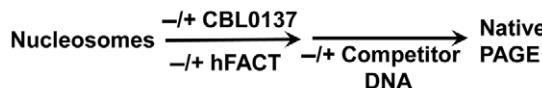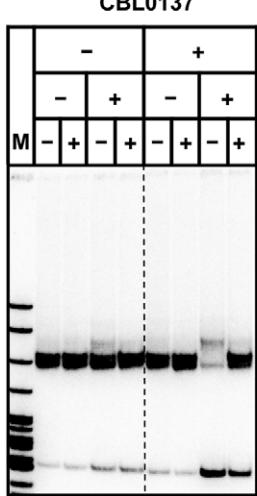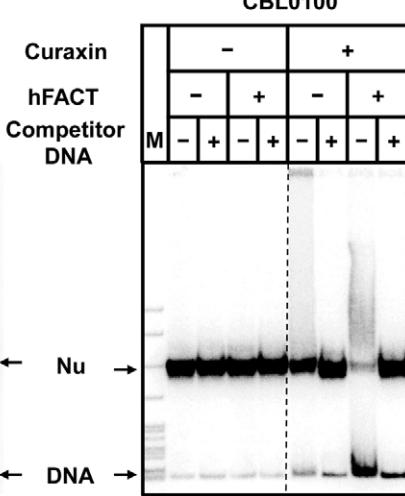
Fig. 3. Nucleosomes are reversibly destabilized in the presence of curaxins and FACT. (**A**) Experimental approach. DNA-labeled nucleosomes were incubated in the presence of FACT, unlabeled competitor DNA, CBL0137 (**B**), or CBL0100 (**C**) and analyzed by native PAGE. Note that in the presence of both FACT and curaxins, nucleosomes are unstable (as reflected in release of histone-free DNA); this effect is reversed in the presence of DNA competitor that partially removes FACT from the nucleosomes (see Fig. 2).

(*39*, *40*, *44*)]. Authentic Pol II elongation complexes (ECs) were assembled and immobilized on Ni-NTA beads and ligated to DNA or nucleosomal templates (*43*, *45*). The transcription was continued in the presence of incomplete combination of nucleotide triphosphates (NTPs), RNA was pulse labeled in the presence of α-^{32}P–labeled guanosine 5′-triphosphate (GTP), and the ECs were stalled. The Pol II ECs were then washed, eluted from the beads, and further transcribed in the presence of all unlabeled NTPs, FACT, curaxin (CBL0137 or CBL0100), and/or competitor core nucleosomes. The pulse-labeled RNA transcripts were purified and analyzed by denaturing PAGE (Fig. 4, B and C).

In the presence of the curaxins, the pattern of Pol II pausing is altered (Fig. 4, B and C), likely due to curaxin intercalation in DNA that affected the DNA structure in the nucleosome (*36*). The differences between CBL0137- and CBL0100-induced Pol II pausing patterns (Fig. 4, B and C) likely occur due to the different structures of the curaxins. In particular, CBL0100 is a more potent DNA intercalator than CBL0137 (*17*, *36*); consistently, it has a stronger effect on the pausing. The +15 nucleosome-specific pausing is partially relieved in the presence of the curaxins, suggesting that they destabilize some DNA-histone interactions in the nucleosome. However, curaxins do not notably affect the catalytic activity of Pol II (fig. S9). In agreement with the previously published results (*5*), FACT facilitates Pol II transcription through the nucleosome, relieving nearly all nucleosomal pausing (Fig. 4, B and C). Curaxins do not considerably inhibit FACT action during transcription through the nucleosome (Fig. 4, B and C), suggesting that FACT redistribution upon curaxin treatment in vivo does not occur because of direct in-
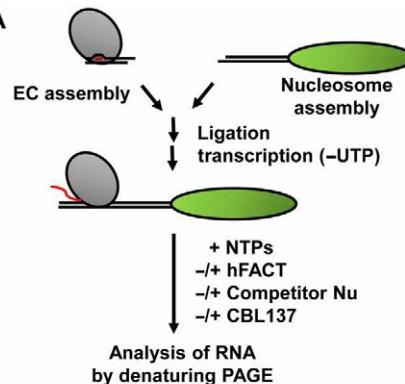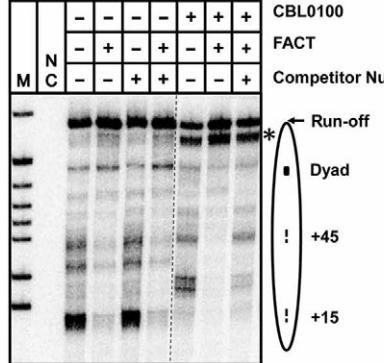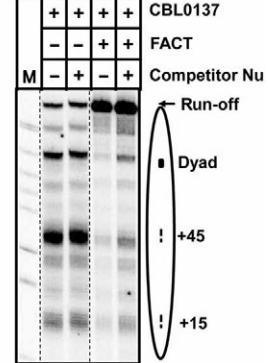
Fig. 4. Curaxins and competitor nucleosomes synergistically inhibit FACT-dependent transcription in vitro. (**A**) Experimental approach. The assembled Pol II EC-119 was ligated to the 603 DNA or nucleosome (*43*, *58*). The RNA was pulse labeled in the presence of a subset of NTPs and [α-^{32}P]GTP, and Pol II was stalled at position −83 (*43*, *45*, *58*). Then, transcription was resumed by adding all unlabeled NTPs, FACT (to 0.2 μM), unlabeled competitor nucleosomes, and CBL0100 (**B**) or CBL0137 (**C**). Note that without competitor nucleosomes, curaxins moderately stimulate transcription through chromatin but do not inhibit FACT action. UTP, uridine 5′-triphosphate.

hibition of Pol II or FACT by curaxins. Therefore, we next studied the effect of curaxin-induced n-trapping on FACT action during Pol II transcription.

In the control experiments in the absence of curaxins and FACT, free nontranscribed mononucleosomes present in the reaction minimally affect the nucleosomal pausing pattern and FACT activity; the curaxins themselves do not inhibit FACT action (Fig. 4, B and C). However, the action of FACT during Pol II transcription through the nucleosome is inhibited in the presence of both curaxins and excess of competitor mononucleosomes added together (Fig. 4, B and C).

Curaxin treatment results in FACT-induced nucleosome unfolding and n-trapping under the same experimental conditions (Figs. 2C and 3), suggesting that FACT-induced unfolding of competitor mononucleosomes and FACT n-trapping cause the inhibition of FACT-dependent transcription, likely after redistribution of FACT to competitor nucleosomes present in excess. To further evaluate this possibility, transcription was conducted in the presence of an excess of biotinylated competitor nucleosomes, the nucleosomes were immobilized on magnetic beads, and the amount of FACT remaining in supernatant was quantified (fig. S10). In the presence of CBL0137, less than 30% of FACT remained in the solution, confirming that

FACT n-trapping on the competitor nucleosomes causes the inhibition of FACT-dependent transcription.

Mechanism of curaxin action in cancer cells: N-trapping

Together, our data suggest that FACT is preferentially associated with transcribed genes in cancer cells, most likely interacting with nucleosomes that are already destabilized by transcribing Pol II complexes. The association of FACT with bodies of transcriptionally active genes is considerably decreased upon curaxin treatment, preferentially on the highly transcribed genes (Fig. 1). FACT interacts with and unfolds nucleosomes in the presence of curaxins (n-trapping; Figs. 2 and 3). Destabilization of nucleosomes, caused by genome-wide CBL0137 binding, makes possible redistribution of FACT to multiple chromatin loci (Fig. 1), which thus compete with transcribed regions for FACT and deplete FACT from transcribed genes. Our in vitro data suggest that curaxins likely uncoil the nucleosomal DNA and thus affect the Pol II transcription through chromatin (Fig. 4). Curaxins inhibit FACT action during Pol II transcription through the nucleosome only in the presence of competitor mononucleosomes, suggesting that the FACT complexes are n-trapped on the excess of transcriptionally silent nucleosomes destabilized by the drug (Fig. 4).

The data suggest a model describing the effect of curaxins on FACT-dependent transcribed genes (Fig. 5). In cancer cells, FACT is associated with gene bodies, especially with the highly transcribed genes, and participates in transcription through chromatin and nucleosome survival by transiently interacting with histone H2A/H2B dimers within the transcribed nucleosomes (2, 5). Curaxins intercalate into DNA of nucleosomes present in various genomic locations, destabilize the nucleosomes, and, together with FACT, cause reversible, partial uncoiling of nucleosomal DNA. FACT tightly binds to the unfolded nucleosomes and thus becomes n-trapped on the vast excess of nontranscribed chromatin that is present in cancer cells; as a result, FACT is depleted from the transcribed regions of the genes.

DISCUSSION

We observed a strong decrease of FACT association with transcribed genes after curaxin treatment. This should result in inhibition of FACT-dependent Pol II transcription. Indeed, n-trapping of FACT results in the inhibition of FACT-dependent Pol II transcription in vitro, suggesting a plausible mechanism for curaxin action in cancer cells through their effect on transcription. Curaxins also alter the structure of nucleosomes, suggesting that, together with FACT, they could act by perturbing global chromatin structure in cancer cells. Last, a recent study has demonstrated that curaxins can induce accumulation of Z-DNA and FACT trapping on the Z-DNA (z-trapping) in cancer cells (36); this mechanism could additionally contribute to the curaxin action.

The process of nucleosome unfolding by FACT in the presence of curaxins is adenosine 5′-triphosphate (ATP) independent and is partially reversed after removal of FACT with an excess of competitor DNA. Only nonphosphorylated (bacterially expressed) hFACT can interact with nucleosomes in the absence of curaxins (the structural state of the nucleosomes in the complex is unknown) (22, 41). In contrast, phosphorylated hFACT purified from insect cells [and likely representing the majority of hFACT complexes in cancer cells (6, 25, 26)] does not interact with intact nucleosomes
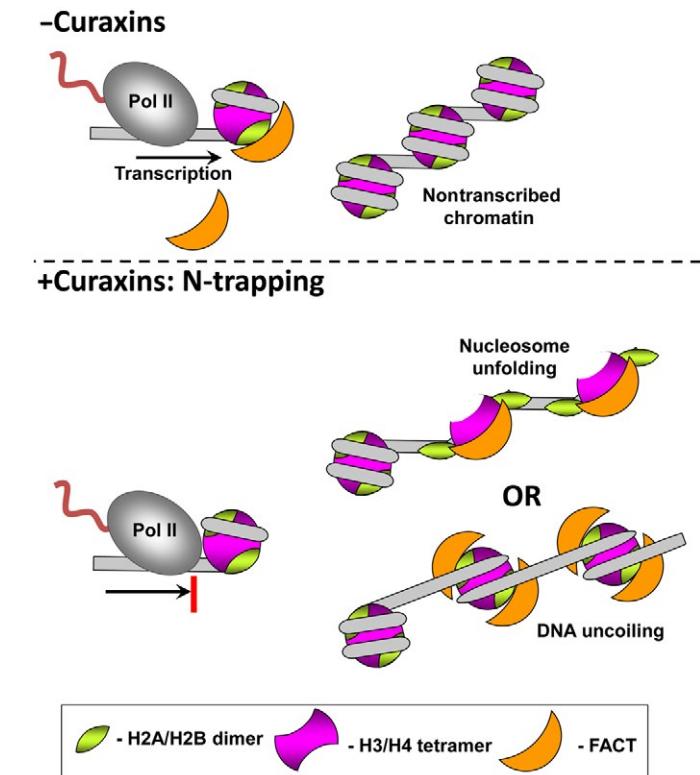
Fig. 5. Mechanism of curaxin action in cancer cells: N-trapping. In the cells without curaxin treatment (−Curaxins), FACT loosely interacts with the transiently exposed DNA binding surface of the H2A/H2B dimer and thus facilitates transcription through the nucleosome barrier. In the presence of curaxins (+Curaxins), FACT causes genome-wide nucleosome unfolding and is tightly trapped on the unfolded nucleosomes (n-trapping). Because the nontranscribed chromatin is present in vast excess, FACT is trapped primarily within inactive chromatin regions. Thus, curaxin-induced n-trapping reduces density of FACT on transcribed regions, resulting in loss of FACT from the transcribed genes and likely affects chromatin dynamics and/or transcript elongation process.

(6, 22, 25, 26, 41, 42). In contrast, yeast FACT can induce reversible nucleosome unfolding even in the absence of curaxins, but only in the presence of Nhp6 protein (28); this activity is likely important for DNA replication and transcription initiation (7, 10, 46). In the complexes of nucleosomes with yeast or human FACT in the absence of curaxins, the interactions occur through multiple binding surfaces on the histone octamer (including on H2A/H2B dimers and H3/H4 tetramers) and FACT (including C-terminal tails of Spt16 and SSRP1 subunits, and mid-AID domain of Spt16) (22–26, 28). It is possible that this multipoint FACT-nucleosome interaction is a prerequisite for nucleosome unfolding that occurs in the presence of curaxins.

The difference in the nucleosome unfolding activities of yeast and human FACT are likely explained by the requirement for an additional protein that is involved in nucleosome unfolding in the case of yeast FACT. FACT subunits Spt16 and SSRP1 (Pob3 in yeast) are highly conserved between yeast and human; additional yeast Nhp6 protein is required for nucleosome unfolding by yeast FACT (7, 21, 46) and by phosphorylated hFACT (47). Yeast Nhp6 is an HMGB-like DNA binding protein that supports various functions of yeast FACT (7, 46). Similar to phosphorylated hFACT, yeast FACT does not induce nucleosome unfolding in the absence of Nhp6 protein

(*24*, *28*, *48*), suggesting that the lack of Nhp6-like protein in hFACT results in additional requirements for curaxins to induce nucleosome unfolding. The ability of hFACT to unfold nucleosomes in the presence of curaxins suggests that hFACT has an intrinsic, conserved nucleosome unfolding activity. However, hFACT can unfold nucleosomes only if they are destabilized, e.g., by curaxins or by DNA damage (*25*). Alternatively, there could be an unknown Nhp6-like human protein that facilitates interaction of hFACT with nucleosomes.

Previously, the c-trapping of FACT was positively correlated with the curaxin toxicity for types of cancer that involved several signaling pathways related to cell death (*17*, *34*). Transcriptions of cell survival–related genes such as NF-κB– and HIF1α-dependent genes were inhibited in the curaxin-treated cancer cells (*17*). During the c-trapping, curaxins also activate the CK2-p53 pathway involved in cell death (*17*), suggesting that curaxin-induced FACT-nucleosome complex could serve as a signal inducing the cell death signaling cascade. Because the c-trapping of FACT occurs during a very short time period (within 1 min) after adding curaxins to cancer cells (*36*), c-trapping therefore is a very early step, acting upstream of and possibly inducing the cell death signaling pathways.

In summary, we have shown that hFACT interacts with and unfolds bulk nucleosomes in the presence of curaxins (n-trapping). This, in turn, results in redistribution of FACT from the bodies of transcribed genes to other genomic regions, strongly affecting the function of FACT in cancer cells.

MATERIALS AND METHODS
Cells
HT1080 cells were obtained from the laboratory of A. Gudkov (Roswell Park Cancer Institute) and authenticated using short tandem repeat polymerase chain reaction (PCR) (100% match with American Type Culture Collection cell line). Cells were maintained in Dulbecco's modified Eagle's medium supplemented with 5% fetal bovine serum and antibiotic/antimycotic solution. HT1080 cells constantly expressing green fluorescent protein–tagged SSRP1 or mCherry-tagged H2B were described previously (*36*).

Analysis of colocalization of FACT and transcribing Pol II using immunofluorescence
Cells were plated in 35-mm plastic dished with no 1.5 glass inserts. Cells were fixed in 4% paraformaldehyde for 10 min, washed three times with phosphate-buffered saline, and blocked in 3% bovine serum albumin (BSA) and 0.1% Triton X-100. Antibody staining was done in 0.5% BSA and 0.1% Triton X-100. Antibody against phosphorylated RNA polymerase (RBP1) at Ser^2 was purchased from Abcam (clone H5, catalog no. ab24758) and diluted 1:200. Antibody against SSRP1 was purchased from BioLegend (clone 10D1, catalog no. 609702) and diluted 1:500. Images were taken using Zeiss Axio Observer A1 inverted microscope with N-Achroplan 100×/1.25 oil lens, Zeiss MRC5 camera, and AxioVision Rel.4.8 software.

Immunoblotting
Cells were lysed in 1× cell culture lysis reagent buffer (Promega) supplemented with protease inhibitor cocktail (Roche). After 10 min of incubation on ice, lysates were spun down at 13,000 rpm for 10 min at +4°C. Supernatant was collected and used as soluble protein extract. Pellet was suspended in the same buffer and sonicated

for 10 cycles, 30 s on/30 s off on ice. Protein concentrations were measured using Bradford reagent (Bio-Rad). Electrophoresis and blotting were done using Bio-Rad Criterion running and blotting cameras and polyvinylidene difluoride (PVDF) membrane (Invitrogen). The following antibodies were used for the detection of SSRP1 (clone 10D1, catalog no. 609702, BioLegend), SPT16 (clone 8D2, catalog no. 607001, BioLegend), and β-actin (clone AC-74, catalog no. A2228, Sigma-Aldrich).

ChIP-seq and nascent RNA-seq
SSRP1 ChIP-seq data in the form of bed files are available at www. ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45393 (*19*). Nascent RNA-seq data done in HT1080 cells are available at GSE107595 (*37*). Raw reads that passed quality filter from Illumina RTA were mapped to human reference genome (hg19) after quality control check using FastQC. For both ChIP-seq and nascent strand RNA-seq data, the raw counts in gene body were generated using featureCounts from Subread R package with RefSeq gene annotation database (*49*, *50*). The quantification data were normalized using DESeq2 R package. To remove impact of overlapping genes on both data, genes with overlapping were removed from analysis. The big wiggle files, heatmaps, and profiles of SSRP1 binding around genes were generated using deepTools software (*51*, *52*).

Alignments of densities of ChIP-seq against the levels of nascent RNA-seq
Single-end SSRP1 ChIP-seq reads (*19*) were aligned to the human reference genome hg19 using Bowtie2 with parameters -X 1000 (to map sequences up to 1 kb with maximum accuracy). The aligned single-end reads were first analyzed using the SPP package (available at https://github.com/hms-dbmi/spp/) to estimate the average fragment length that resulted in the sonication process. This average length of 85 bp was used to extend the fragments from the 5′ end that was sequenced toward the 3′ end. The raw occupancy profiles were computed in MATLAB by stacking the extended (85 bp) reads and counting the number of reads that overlapped with each base pair. The raw occupancy profile for each chromosome was normalized by the chromosome average. For each RefSeq annotated gene, the average SSRP1 density was computed by averaging the normalized SSRP1 occupancy over the whole gene body. For counting the levels of transcriptions, single-end nascent RNA-seq reads (*37*) were aligned to the human transcriptome using TopHat2 with the default parameters. The raw nascent RNA-seq counts were computed in MATLAB by stacking all 5′ ends of the aligned reads, which correspond to the 3′ ends of the nascent transcripts. For each chromosome, the corresponding NET-seq profiles were normalized by the chromosome average. For each RefSeq annotated gene, the normalized nascent RNA-seq counts were averaged over the whole gene body.

Analysis of SSRP1 occupancy of various genomic regions in control and CBL0137-treated samples
MACS 2.0 (*53*) with default parameters for pair-end BAM files was used to identify peaks from ChIP-seq of SSRP1 (*19*). Heatmaps and profiles under all conditions were generated using deepTools on RPKM (reads per kilobase million) normalized coverage data using merged bam files of biological replicates. Tag densities for RefGenes body with 3 kb up- and downstream were plotted. Heatmaps of the unexpressed and top-expressed (5%) genes and quantiles of all genes

were prepared using nascent RNA-seq data for control untreated HT1080 cells (*37*).

DNA templates

All DNA templates were amplified by PCR and purified from gel electrophoresis using a gel extraction kit (Omega Bio-Tek), as described (*39*). For spFRET, DNA templates were assembled from fluorescence oligos [Cy3/Cy5-labeled pairs at +13 and +91 bp relative to 603 NPS boundary], as described (*28*). All sequence of primers and templates will be provided per request.

Protein purification

Yeast Pol II was purified as described (*45*, *54*). -H1 chicken erythrocyte chromatin, histone H2A/H2B dimer, and histone H3/H4 tetramer were purified as described (*44*, *55*, *56*). hFACT was purified as described (*2*).

Nucleosome assembly and purification

Nucleosomes were assembled as described (*44*). In short, NPS templates were mixed with purified chicken erythrocyte H2A/H2B dimers and H3/H4 tetramers or with -H1 chicken erythrocyte chromatin in the presence of salmon testes DNA as competitors in the buffer containing 2 M NaCl, 10 mM tris-HCl (pH 7.4), 0.1% NP-40, and 0.2 mM EDTA (pH 8). The samples were then dialyzed against buffers with progressively decreasing (2 M, 1.5 M, 1 M, 0.75 M, 0.5 M, and 10 mM) NaCl at 4°C for 2 hours at each step. For gel shift analysis of FACT binding and spFRET, nucleosomes were gel purified after assembly and analyzed as described (*57*).

Gel shift analysis of FACT binding

Core nucleosomes (final concentration, 15 nM) were incubated with hFACT (final concentration, 400 nM) in the transcription buffer [TB; 20 mM tris-HCl (pH 8.0), 5 mM MgCl$_2$, and 2 mM β-mercaptoethanol] containing 40 mM KCl and then mixed with CBL0137 (final concentration, 1 μM) or CBL0100 (final concentration, 2.5 μM) for 1 min. Curaxins were provided by Incuron Inc. The samples were analyzed by native PAGE as described (*57*). To reverse FACT binding, DNA competitor [salmon testes DNA (250 ng/μl)] was added to the loading buffer.

Single-particle Förster resonance energy transfer

Nucleosomes reconstituted in the presence of donor chromatin from chicken erythrocytes and fluorescently labeled DNA templates (containing 603 positioning sequence) were gel purified and used for spFRET measurements at a concentration of 0.5 nM, as described (*28*). Nucleosomes were incubated in the presence of hFACT (0.1 μM) and/or CBL0137 (5 μM) in the TB containing 150 mM KCl for 5 min at 25°C. spFRET measurements and raw data analysis were conducted as described (*28*).

In vitro transcription assay

The in vitro transcription assay with yeast Pol II was performed as described (*5*, *58*). In short, the ECs were assembled using purified yeast Pol II and DNA/RNA oligonucleotides. The assembled Pol II ECs were immobilized on Ni-NTA (nickel–nitrilotriacetic acid) resins (Qiagen), washed, and eluted from the beads. ECs and nucleosomal templates (or corresponding DNA fragments) were ligated by T4 ligase (Promega). Pol II was then progressed to position −83 using limited mixture of NTPs and α-^{32}P–labeled GTP. ECs were washed

from the unincorporated NTPs and eluted, and transcription was resumed in the presence of unlabeled NTPs, hFACT (final concentration, 0.1 μM), CBL0137 (1 μM) or CBL0100 (5 μM), and mononucleosomes (~0.18 μM) in the TB containing 150 mM KCl for 10 min. Transcription was terminated using phenol/chloroform extraction. RNA transcripts were purified and analyzed by denaturing PAGE.

Biotinylated nucleosome pull-down assay

The competitor core nucleosomes were assembled using 5′ end-biotinylated 147-bp DNA template. Transcription by Pol II in the presence or absence of curaxins, FACT, and competitor core nucleosomes (containing 50% of end-biotinylated DNA) was conducted as described above. After transcription, hydrophilic streptavidin magnetic beads (NEB) were added to the reaction mixture for 10 min at room temperature, and the supernatant was collected. The beads were resuspended in 1× TB40 buffer. The SDS loading buffer [1× SDS loading buffer containing 50 mM tris-HCl (pH 6.8), 2% SDS, 70 mM β-mercaptoethanol, and 10% glycerol] was added to the samples and heated at 99°C for 10 min. Electrophoresis and blotting were done using NuPAGE (4 to 12% bis-tris) (Invitrogen), NuPAGE MOPS SDS Running Buffer (Invitrogen), PVDF membrane (Invitrogen), and film developing solutions (SuperSignal West Femto Maximum Sensitivity Substrate, Thermo Fisher Scientific). The Flag-tagged SPT16 protein was detected using the DYKDDDDK Tag Monoclonal Antibody (FG4R) (Invitrogen). The data were quantified using OptiQuant software and normalized for total protein loading.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/11/eaav2131/DC1

Fig. S1. Analysis of colocalization of FACT and transcribing Pol II using immunofluorescence.
Fig. S2. Redistribution of FACT in HT1080 cells from nucleoplasm to chromatin.
Fig. S3. Redistribution of FACT in nucleus of HT1080-treated cells from nucleoplasm to chromatin.
Fig. S4. Heatmaps of SSRP1 occupancy in the vicinity of TSSs and TES of the genes in HT1080 cells.
Fig. S5. Analysis of the average SSRP1 densities over gene bodies against the levels of transcription of the corresponding genes.
Fig. S6. Curaxins preferentially remove SSRP1 from gene bodies of highly transcribed genes.
Fig. S7. Analysis of gel-purified nucleosomes by native PAGE.
Fig. S8. Typical frequency distributions of FRET efficiencies of the N13/91 nucleosomes.
Fig. S9. The catalytic activity of Pol II is minimally affected by curaxins.
Fig. S10. Trapping of FACT on immobilized competitor nucleosomes after Pol II transcription in the presence of CBL0137.
Table S1. Statistical data for spFRET analysis.

REFERENCES AND NOTES

1. G. Orphanides, G. LeRoy, C.-H. Chang, D. S. Luse, D. Reinberg, FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **92**, 105–116 (1998).
2. R. Belotserkovskaya, S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, D. Reinberg, FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090–1093 (2003).
3. P. B. Mason, K. Struhl, The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Mol. Cell. Biol.* **23**, 8323–8333 (2003).
4. A. Saunders, J. Werner, E. D. Andrulis, T. Nakayama, S. Hirose, D. Reinberg, J. T. Lis, Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science* **301**, 1094–1096 (2003).
5. F.-K. Hsieh, O. I. Kulaeva, S. S. Patel, P. N. Dyer, K. Luger, D. Reinberg, V. M. Studitsky, Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7654–7659 (2013).
6. P. Chen, L. Dong, M. Hu, Y.-Z. Wang, X. Xiao, Z. Zhao, J. Yan, P.-Y. Wang, D. Reinberg, M. Li, W. Li, G. Li, Functions of FACT in breaking the nucleosome and maintaining its integrity at the single-nucleosome level. *Mol. Cell* **71**, 284–293.e4 (2018).

7. J. Wittmeyer, T. Formosa, The *Saccharomyces cerevisiae* DNA polymerase α catalytic subunit interacts with Cdc68/Spt16 and with Pob3, a protein similar to an HMG1-like protein. *Mol. Cell. Biol.* **17**, 4178–4190 (1997).

8. B.C.-M. Tan, C.-T. Chien, S. Hirose, S.-C. Lee, Functional cooperation between FACT and MCM helicase facilitates initiation of chromatin DNA replication. *EMBO J.* **25**, 3975–3985 (2006).

9. T. Abe, K. Sugimura, Y. Hosono, Y. Takami, M. Akita, A. Yoshimura, S. Tada, T. Nakayama, H. Murofushi, K. Okumura, S. Takeda, M. Horikoshi, M. Seki, T. Enomoto, The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. *J. Biol. Chem.* **286**, 30504–30512 (2011).

10. C. F. Kurat, J. T. P. Yeeles, H. Patel, A. Early, J. F. X. Diffley, Chromatin controls DNA replication origin selection, lagging-strand synthesis, and replication fork rates. *Mol. Cell* **65**, 117–130 (2017).

11. D. M. Keller, X. Zeng, Y. Wang, Q. H. Zhang, M. Kapoor, H. Shu, R. Goodman, G. Lozano, Y. Zhao, H. Lu, A DNA damage–induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol. Cell* **7**, 283–292 (2001).

12. D. M. Keller, H. Lu, p53 serine 392 phosphorylation increases after UV through induction of the assembly of the CK2 hSPT16 SSRP1 complex. *J. Biol. Chem.* **277**, 50206–50213 (2002).

13. N. M. Krohn, C. Stemmer, P. Fojan, R. Grimm, K. D. Grasser, Protein kinase CK2 phosphorylates the high mobility group domain protein SSRP1, inducing the recognition of UV-damaged DNA. *J. Biol. Chem.* **278**, 12710–12715 (2003).

14. K. Heo, H. Kim, S. H. Choi, J. Choi, K. Kim, J. Gu, M. R. Lieber, A. S. Yang, W. An, FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. *Mol. Cell* **30**, 86–97 (2008).

15. J. L. Charles Richard, M. S. Shukla, H. Menoni, K. Ouararhni, I. N. Lone, Y. Roulland, C. Papin, E. Ben Simon, T. Kundu, A. Hamiche, D. Angelov, S. Dimitrov, FACT assists base excision repair by boosting the remodeling activity of RSC. *PLOS Genet.* **12**, e1006221 (2016).

16. H. Garcia, D. Fleyshman, K. Kolesnikova, A. Safina, M. Commane, G. Paszkiewicz, A. Omelian, C. Morrison, K. Gurova, Expression of FACT in mammalian tissues suggests its role in maintaining of undifferentiated state of cells. *Oncotarget* **2**, 783–796 (2011).

17. A. V. Gasparian, C. A. Burkhart, A. A. Purmal, L. Brodsky, M. Pal, M. Saranadasa, D. A. Bosykh, M. Commane, O. A. Guryanova, S. Pal, A. Safina, S. Sviridov, I. E. Koman, J. Veith, A. A. Komar, A. V. Gudkov, K. V. Gurova, Curaxins: Anticancer compounds that simultaneously suppress NF-κB and activate p53 by targeting FACT. *Sci. Transl. Med.* **3**, 95ra74 (2011).

18. F.-K. Hsieh, O. I. Kulaeva, I. V. Orlovsky, V. M. Studitsky, FACT in cell differentiation and carcinogenesis. *Oncotarget* **2**, 830–832 (2011).

19. H. Garcia, J. C. Miecznikowski, A. Safina, M. Commane, A. Ruusulehto, S. Kilpinen, R. W. Leach, K. Attwood, Y. Li, S. Degan, A. R. Omilian, O. Guryanova, O. Papantonopoulou, J. Wang, M. Buck, S. Liu, C. Morrison, K. V. Gurova, Facilitates chromatin transcription complex is an "accelerator" of tumor transformation and potential marker and target of aggressive cancers. *Cell Rep.* **4**, 159–173 (2013).

20. K. Gurova, H.-W. Chang, M. E. Valieva, P. Sandlesh, V. M. Studitsky, Structure and function of the histone chaperone FACT—Resolving FACTual issues. *Biochim. Biophys. Acta* **1861**, 892–904 (2018).

21. G. Orphanides, W.-H. Wu, W. S. Lane, M. Hampsey, D. Reinberg, The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**, 284–288 (1999).

22. D. D. Winkler, U. M. Muthurajan, A. R. Hieb, K. Luger, Histone chaperone FACT coordinates nucleosome interaction through multiple synergistic binding events. *J. Biol. Chem.* **286**, 41883–41892 (2011).

23. M. Hondele, T. Stuwe, M. Hassler, F. Halbach, A. Bowman, E. T. Zhang, B. Nijmeijer, C. Kotthoff, V. Rybin, S. Amlacher, E. Hurt, A. G. Ladurner, Structural basis of histone H2A–H2B recognition by the essential chaperone FACT. *Nature* **499**, 111–114 (2013).

24. D. J. Kemble, L. L. McCullough, F. G. Whitby, T. Formosa, C. P. Hill, FACT disrupts nucleosome structure by binding H2A-H2B with conserved peptide motifs. *Mol. Cell* **60**, 294–306 (2015).

25. Y. Tsunaka, Y. Fujiwara, T. Oyama, S. Hirose, K. Morikawa, Integrated molecular mechanism directing nucleosome reorganization by human FACT. *Genes Dev.* **30**, 673–686 (2016).

26. T. Wang, Y. Liu, G. Edwards, D. Krzizike, H. Scherman, K. Luger, The histone chaperone FACT modulates nucleosome structure by tethering its components. *Life Sci. Alliance* **1**, e201800107 (2018).

27. T. Formosa, The role of FACT in making and breaking nucleosomes. *Biochim. Biophys. Acta* **1819**, 247–255 (2013).

28. M. E. Valieva, G. A. Armeev, K. S. Kudryashova, N. S. Gerasimova, A. K. Shaytan, O. I. Kulaeva, L. L. McCullough, T. Formosa, P. G. Georgiev, M. P. Kirpichnikov, V. M. Studitsky, A. V. Feofanov, Large-scale ATP-independent nucleosome unfolding by a histone chaperone. *Nat. Struct. Mol. Biol.* **23**, 1111–1116 (2016).

29. S. Takahata, Y. Yu, D. J. Stillman, FACT and Asf1 regulate nucleosome dynamics and coactivator binding at the *HO* promoter. *Mol. Cell* **34**, 405–415 (2009).

30. T. Y. Erkina, A. Erkine, ASF1 and the SWI/SNF complex interact functionally during nucleosome displacement, while FACT is required for nucleosome reassembly at yeast heat shock gene promoters during sustained stress. *Cell Stress Chaperones* **20**, 355–369 (2015).

31. S. Carvalho, A. C. Raposo, F. B. Martins, A. R. Grosso, S. C. Sridhara, J. Rino, M. Carmo-Fonseca, S. F. de Almeida, Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res.* **41**, 2881–2893 (2013).

32. I. E. Koman, M. Commane, G. Paszkiewicz, B. Hoonjan, S. Pal, A. Safina, I. Toshkov, A. A. Purmal, D. Wang, S. Liu, C. Morrison, A. V. Gudkov, K. V. Gurova, Targeting FACT complex suppresses mammary tumorigenesis in *Her2/neu* transgenic mice. *Cancer Prev. Res.* **5**, 1025–1035 (2012).

33. A. Safina, H. Garcia, M. Commane, O. Guryanova, S. Degan, K. Kolesnikova, K. V. Gurova, Complex mutual regulation of facilitates chromatin transcription (FACT) subunits on both mRNA and protein levels in human cells. *Cell Cycle* **12**, 2423–2434 (2013).

34. N. V. Maluchenko, H. W. Chang, M. T. Kozinova, M. E. Valieva, N. S. Gerasimova, A. V. Kitashov, M. P. Kirpichnikov, P. G. Georgiev, V. M. Studitsky, Inhibiting the pro-tumor and transcription factor FACT: Mechanisms. *Mol. Biol.* **50**, 599–610 (2016).

35. C. Burkhart, D. Fleyshman, R. Kohrn, M. Commane, J. Garrigan, V. Kurbatov, I. Toshkov, R. Ramachandran, L. Martello and K. V. Gurova, Curaxin CBL0137 eradicates drug resistant cancer stem cells and potentiates efficacy of gemcitabine in preclinical models of pancreatic cancer. *Oncotarget* **5**, 11038–11053 (2014).

36. A. Safina, P. Cheney, M. Pal, L. Brodsky, A. Ivanov, K. Kirsanov, E. Lesovaya, D. Naberezhnov, E. Nesher, I. Koman, D. Wang, J. Wang, M. Yakubovskaya, D. Winkler, K. Gurova, FACT is a sensor of DNA torsional stress in eukaryotic cells. *Nucleic Acids Res.* **45**, 1925–1945 (2017).

37. E. Nesher, A. Safina, I. Aljahdali, S. Portwood, E. S. Wang, I. Koman, J. Wang, K. V. Gurova, Role of chromatin damage and chromatin trapping of FACT in mediating the anticancer cytotoxicity of DNA-binding small molecule drugs. *Cancer Res.* **78**, 1431–1443 (2018).

38. T. A. Barone, C. A. Burkhart, A. Safina, G. Haderski, K. V. Gurova, A. A. Purmal, A. V. Gudkov, R. J. Plunkett, Anticancer drug candidate CBL0137, which inhibits histone chaperone FACT, is efficacious in preclinical orthotopic models of temozolomide-responsive and -resistant glioblastoma. *Neuro Oncol.* **19**, 186–196 (2017).

39. V. A. Bondarenko, L. M. Steele, A. Ujvari, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse, V. M. Studitsky, Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* **24**, 469–479 (2006).

40. A. Thåström, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, J. Widom, Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* **288**, 213–229 (1999).

41. Y. Tsunaka, J. Toga, H. Yamaguchi, S.-i. Tate, S. Hirose, K. Morikawa, Phosphorylated intrinsically disordered region of FACT masks its nucleosomal DNA binding elements. *J. Biol. Chem.* **284**, 24610–24621 (2009).

42. M. E. Valieva, N. S. Gerasimova, K. S. Kudryashova, A. L. Kozlova, M. P. Kirpichnikov, Q. Hu, M. V. Botuyan, G. Mer, A. V. Feofanov, V. M. Studitsky, Stabilization of nucleosomes by histone tails and by FACT revealed by spFRET microscopy. *Cancers* **9**, pii: E3 (2017).

43. M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, V. M. Studitsky, Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription. *Mol. Cell* **9**, 541–552 (2002).

44. D. A. Gaykalova, O. I. Kulaeva, V. A. Bondarenko, V. M. Studitsky, Preparation and analysis of uniquely positioned mononucleosomes. *Methods Mol. Biol.* **523**, 109–123 (2009).

45. W. Walter, M. L. Kireeva, V. Tchernajenko, M. Kashlev, V. M. Studitsky, Assay of the fate of the nucleosome during transcription by RNA polymerase II. *Methods Enzymol.* **371**, 564–577 (2003).

46. N. K. Brewster, G. C. Johnston, R. A. Singer, A bipartite yeast SSRP1 analog comprised of Pob3 and Nhp6 proteins modulates transcription. *Mol. Cell. Biol.* **21**, 3491–3502 (2001).

47. L. L. McCullough, Z. Connell, H. Xin, V. M. Studitsky, A. V. Feofanov, M. E. Valieva, T. Formosa, Functional roles of the DNA-binding HMGB domain in the histone chaperone FACT in nucleosome reorganization. *J. Biol. Chem.* **293**, 6121–6133 (2018).

48. H. Xin, S. Takahata, M. Blanksma, L. McCullough, D. J. Stillman, T. Formosa, yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement. *Mol. Cell* **35**, 365–376 (2009).

49. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

50. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

51. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

52. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

53. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

54. M. L. Kireeva, N. Komissarova, D. S. Waugh, M. Kashlev, The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* **275**, 6530–6536 (2000).

55. R. H. Simon, G. Felsenfeld, A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. *Nucleic Acids Res.* **6**, 689–696 (1979).

56. C. von Holt, W. F. Brandt, H. J. Greyling, G. G. Lindsey, J. D. Retief, J. d. A. Rodrigues, S. Schwager, B. T. Sewell, Isolation and characterization of histones. *Methods Enzymol.* **170**, 431–523 (1989).

57. V. M. Studitsky, Preparation and analysis of positioned nucleosomes. *Methods Mol. Biol.* **119**, 17–26 (1999).

58. O. I. Kulaeva, D. A. Gaykalova, N. A. Pestov, V. V. Golovastov, D. G. Vassylyev, I. Artsimovitch, V. M. Studitsky, Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nat. Struct. Mol. Biol.* **16**, 1272–1278 (2009).

Citation: H.-W. Chang, M. E. Valieva, A. Safina, R. V. Chereji, J. Wang, O. I. Kulaeva, A. V. Morozov, M. P. Kirpichnikov, A. V. Feofanov, K. V. Gurova, V. M. Studitsky, Mechanism of FACT removal from transcribed genes by anticancer drugs curaxins. *Sci. Adv.* **4**, eaav2131 (2018).

CANCER IMMUNOLOGY

Combination cancer immunotherapy targeting PD-1 and GITR can rescue CD8+ T cell dysfunction and maintain memory phenotype

Bei Wang*, Wen Zhang*, Vladimir Jankovic, Jacquelynn Golubov, Patrick Poon, Erin M. Oswald, Cagan Gurer, Joyce Wei, Ilyssa Ramos, Qi Wu, Janelle Waite, Min Ni, Christina Adler, Yi Wei, Lynn Macdonald, Tracey Rowlands, Susannah Brydges, Jean Siao, William Poueymirou, Douglas MacDonald, George D. Yancopoulos, Matthew A. Sleeman, Andrew J. Murphy, Dimitris Skokos†

Most patients with cancer do not develop durable antitumor responses after programmed cell death protein 1 (PD-1) or programmed cell death ligand 1(PD-L1) checkpoint inhibition monotherapy because of an ephemeral reversal of T cell dysfunction and failure to promote long-lasting immunological T cell memory. Activating costimulatory pathways to induce stronger T cell activation may improve the efficacy of checkpoint inhibition and lead to durable antitumor responses. We performed single-cell RNA sequencing of more than 2000 tumor-infiltrating CD8+ T cells in mice receiving both PD-1 and GITR (glucocorticoid-induced tumor necrosis factor receptor–related protein) antibodies and found that this combination synergistically enhanced the effector function of expanded CD8+ T cells by restoring the balance of key homeostatic regulators CD226 and T cell immunoreceptor with Ig and ITIM domains (TIGIT), leading to a robust survival benefit. Combination therapy decreased CD8+ T cell dysfunction and induced a highly proliferative precursor effector memory T cell phenotype in a CD226-dependent manner. PD-1 inhibition rescued CD226 activity by preventing PD-1–Src homology region 2 (SHP2) dephosphophorylation of the CD226 intracellular domain, whereas GITR agonism decreased TIGIT expression. Unmasking the molecular pathways driving durable antitumor responses will be essential to the development of rational approaches to optimizing cancer immunotherapy.

INTRODUCTION

After the initial clinical successes obtained using programmed cell death protein 1 (PD-1) and cytotoxic T lymphocyte–associated antigen 4 (CTLA-4) antibody (Ab) treatments in patients with cancer, the number of immunotherapy agents in clinical development is expanding rapidly with goals of improving the limited response rate and generating more durable responses (1, 2). Recent elegant studies using infectious disease–related mouse models demonstrated that PD-1 blockade induces an incomplete rescue of exhausted T cells (T_{Exs}), thus failing to restore T_{Exs} into T effector memory cells (T_{EMs}), which are required for an effective immune memory (3, 4). Although PD-1 and/or programmed cell death ligand 1 (PD-L1) inhibition affects the transcriptional and epigenetic landscape of T_{Exs}, these cells are not fully reprogrammed when antigen concentration remains high and fail to acquire a memory phenotype upon antigen clearance (3). Designing additional strategies to induce more robust T cell activation and/or target other immunoregulatory pathways could contribute to the generation of sustainable long-term responses.

Combination treatments aimed at PD-1 inhibition and activation of costimulatory receptor GITR [glucocorticoid-induced tumor necrosis factor receptor (TNFR)–related protein or TNFRSF18] to induce a stronger T cell activation are currently being evaluated in early-phase clinical trials for patients with metastatic melanoma and other solid tumors (5, 6). Although the intrinsic properties of each Ab and the selection of the appropriate indications for use play an important role in the clinical outcome, emerging data from a small clinical study reported limited clinical activity for anti-GITR mono-

therapy and potentially promising data for the combination therapy (7). In preclinical studies in which monotherapy with anti-GITR or anti–PD-1 Ab has limited efficacy (e.g., in large or poorly immunogenic murine tumors) (6, 8), combination therapy was able to achieve long-term survival in mouse models of ovarian and breast cancer (9). However, the effect of this synergism on T cell dysfunction and the underlying molecular mechanisms remain unknown. Here, we sought to establish an experimental approach to identify the mechanisms of antitumor synergism between anti-GITR and anti–PD-1 Abs. We genetically profiled more than 2000 tumor-infiltrating CD8+ T cells in the murine MC38 colon adenocarcinoma model using single-cell T cell receptor (TCR) and transcriptome sequencing. Our systematic approach demonstrated that combination immunotherapy rescued T cell dysfunction while promoting a memory phenotype and also revealed the molecular pathways driving durable antitumor responses in the MC38 and RENCA tumor models. Identification of these pathways provides a rational basis for optimizing existing combination immunotherapies and improving tumor responsiveness.

RESULTS

Anti–PD-1 and anti-GITR Ab combination therapy synergistically rejects established tumors and reinvigorates dysfunctional intratumoral CD8+ T cells

To examine the antitumor effect of PD-1 and GITR combination immunotherapy, we first used a widely studied and poorly immunogenic murine MC38 colon adenocarcinoma model (Fig. 1A). Although variable reductions in tumor volume and modestly prolonged survival have been reported in this model (10), monotherapy with anti–PD-1 or anti-GITR [DTA-1 rat immunoglobulin G2b (IgG2b)] Ab is not effective at inducing complete and durable tumor regression
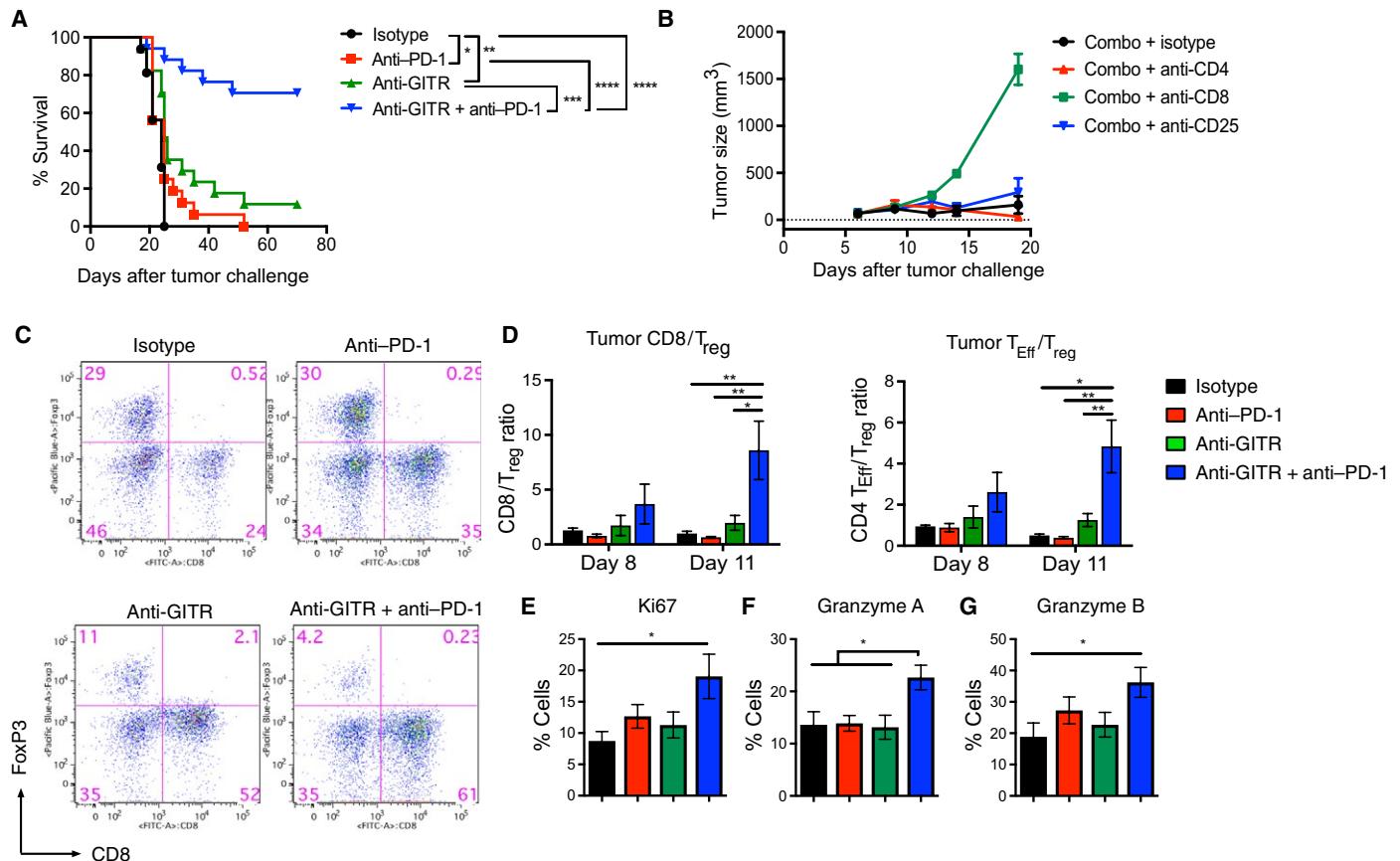
Fig. 1. Anti-GITR and anti–PD-1 combination synergistically rejects established murine tumors and reinvigorates intratumoral dysfunctional T cells. (**A**) Survival curve of MC38 tumor–bearing mice treated with anti-GITR and/or anti–PD-1 Ab treatment (days 6 and 13). Results depict cumulative survival curves with indicated treatment (n = 16 to 17 mice per group). (**B**) $CD8^+$ T cell–dependent long-term tumor protection mediated by combination treatment. C57BL/6 mice were treated with anti-CD8–, anti-CD4–, or anti-CD25–depleting Ab before and during therapy with anti-GITR and anti–PD-1 Ab or control Ab. Data shown are average tumor growth curve upon treatment with different depletion Abs (n = 6 mice per group). (**C** and **D**) Combination treatment increases intratumoral T_{Eff}/T_{reg} ratio. (C) Representative FACS plots showing tumor T cell subsets on day 11 ($FoxP3^+$ versus $CD8^+$, cells pregated on live/single cells/$CD45^+/CD3^+$). FITC, fluorescein isothiocyanate. (D) Summary FACS result of intratumoral $CD8^+$ T cell/T_{reg} and $CD4^+$ T_{Eff}/T_{reg} ratio on days 8 and 11 after tumor challenge. Data are representative of three independent experiments (n = 7 mice per group). (**E** to **G**) Combination treatment reinvigorates intratumoral dysfunctional T cells. Tumors were harvested on days 11 and 12 after implantation, dissociated into single-cell suspension restimulated with phorbol 12-myristate 13-acetate/ionomycin with the presence of brefeldin A. Cells were fixed and permeabilized, followed by intracellular staining with Ki67 (E), granzyme A (F), and granzyme B (G). Data shown are percentages of positive cells (n = 8 to 9 mice per group). All error bars in figures show SEM. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, log-rank test (A) and one-way ANOVA with Tukey's test (C to G).

in established tumors (*8*, *11*, *12*). To determine the relative contribution of GITR agonism, we selected the DTA-1 mouse IgG2b isotype in this model to avoid maximum regulatory T cell (T_{reg}) depletion as seen with the mouse IgG2a isotype, allowing us to test the effect of the combination therapy. The Abs were administered on 6 and 13 days after tumor challenge when tumors were palpable. In line with previous studies (*9*, *12*, *13*), whereas anti-GITR or anti–PD-1 treatment alone exhibited little antitumor effect, combination therapy synergistically prolonged overall survival time of mice (~70% of mice were tumor free for >80 days; Fig. 1A). Cell depletion studies showed that tumor rejection was primarily dependent on the $CD8^+$ T cells, because removal of $CD8^+$ T cells abrogated the antitumor effect elicited by anti–PD-1/GITR Ab treatment (Fig. 1B and fig. S1). The antitumor effect was associated with a reduction in intratumoral T_{reg} and a robust increase in the effector $CD8^+$ and $CD4^+$ T cell (T_{Eff})–to-T_{reg} ratio, indicating a more immunostimulatory microenvironment (Fig. 1, C and D) and validating previous findings by other groups (*13*, *14*). In

addition, combination therapy increased proliferation (as judged by expression of Ki67) and granzyme A and B production by intratumoral $CD8^+$ T cells (Fig. 1, E to G).

Concurrent single-cell TCR and transcriptome analyses identified a unique gene signature of clonally expanded $CD8^+$ T cells upon combination immunotherapy

We developed and validated a new bioinformatics platform designated rpsTCR (for random priming sequencing TCR) to extract, reconstruct, and analyze TCR sequences after single-agent or combination treatment. Briefly, this platform uses random priming RNA sequencing (RNA-seq) data generated from sorted single cells to identify mouse T cell clones potentially associated with tumor reactivity across different treatments (fig. S2 and table S2) (*15*). Mice were treated as described in Fig. 1, and more than 2000 $CD8^+$ T cells were sorted from tumor-bearing mice at days 8 and 11 after tumor challenge (Fig. 2A). The rpsTCR platform was used to profile the TCR sequences of 1379

CD8[+] T cells. Our detection rates of TCRB-CDR3 (86%), TCRA-CDR3 (78.2%), and paired TCRB and TCRA (73.1%) were comparable with the reported detection rates using targeted TCR sequencing from single T cells (table S2) (15). At the early time point (day 8), very few clones of high-frequency T cells (defined as ≥ 3 T cells sharing identical TCR sequences) were detected in all treatment groups (fig. S3). By day 11, we identified two high-frequency T cell clones (representing 5.7% of sequenced single CD8[+] T cells) from isotype control samples—3 clones (20.3%) for anti-GITR samples and 6 clones (26.7%) for anti–PD-1 samples—and 10 clones (31.9%) for combi-

nation-treated samples. Between days 8 and 11, a significant clonal expansion of intratumoral CD8[+] T cells was induced by anti–PD-1 monotherapy, in agreement with published data on human melanoma patients (16) showing an increase in TCR clonal size after anti–PD-1 therapy with pembrolizumab. Our results extend these findings by showing that dual targeting of PD-1 and GITR further enhances intratumoral CD8[+] T cell TCR clonal expansion (fig. S3). Of note, anti-GITR and/or anti–PD-1 had no significant impact on peripheral (spleen) T cell clonality (fig. S3), consistent with patient data using pembrolizumab (16).
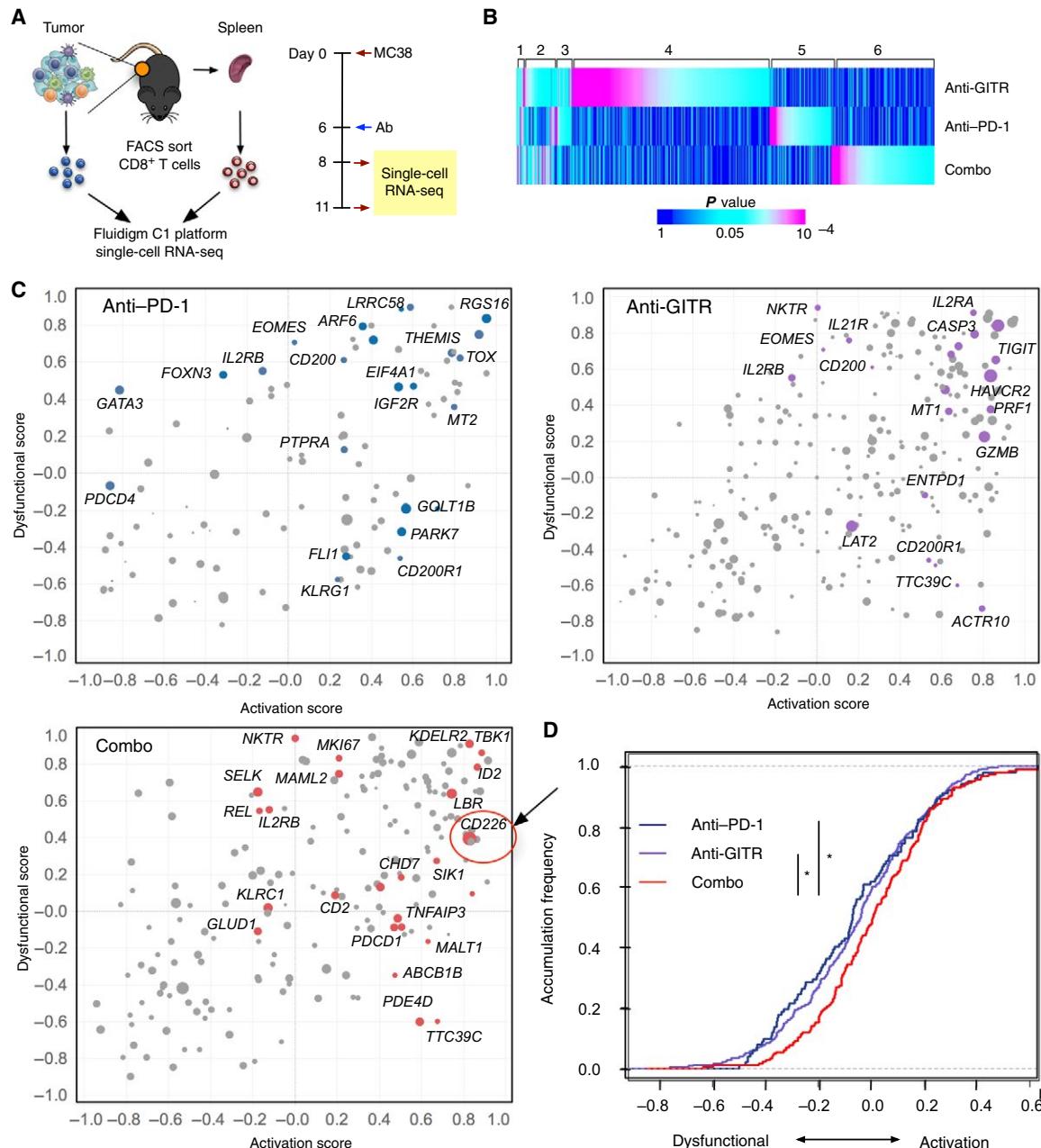
Fig. 2. Single-cell RNA-seq analysis of intratumoral CD8[+] T cells reveals unique gene profiles upon combination treatment. (A) Schematic of tumor immunotherapy and single-cell sorting study design. (B) Heat map of significantly changed genes in clonal expanded intratumoral CD8[+] T cells upon different treatment. Numbers indicate unique gene clusters (see table S3 for gene list). Color gradient indicates the P value. (C) Distribution of genes significantly affected by each treatment on the dysfunctional/activation plot. (D) KS plot of the values of the gene signatures on the dysfunction ↔ activation axis (KS, P < 0.05).

To evaluate the effect of combination therapy on the activation state of intratumoral CD8+ T cells, we performed gene expression pathway analysis on clonally expanded CD8+ T cells. Combination treatment synergistically integrated the pathways modulated by each single agent, resulting in a distinct transcriptional state associated with induction of adaptive immune response, cell cycle, and metabolic activity gene expression pathways (Table 1). We then applied to our data the uncoupled T cell dysfunction and activation gene expression module analysis, as described by Singer *et al.* (*17*). A total of 452 genes were significantly regulated by either single-agent or combination treatment compared with isotype control. These genes were plotted against the activation/dysfunctional score (Fig. 2B, *P* < 0.05). Genes shared or unique across treatment groups were sorted on the basis of *P* value and then grouped into six clusters (Fig. 2C). Selected genes differentially increased in each group and fold changes compared with isotype control treatment are listed in table S3. Up-regulation of certain T cell activation markers (i.e., *IL2rb*) was shared across treatment groups, whereas most of the differentially regulated gene signatures were nonoverlapping, consistent with distinct transcriptional states of expanded CD8+ T cell clones. GITR agonism specifically up-regulated genes involved in T cell activation/cytotoxicity (*Il2ra*, *Txk*, *Gzme*, *Gzmd*, and *Prf1*), in prosurvival function (*Tnfrsf1b*), and in immune checkpoint (*Havcr2*, *Tigit*, and *Entpd1*). PD-1 blockade stimulated induction of *Gata3*, *Tox*, *Mt2*, and *Pdcd4* (*18*), which are involved in T cell dysfunction and apoptosis, and promoted expression of *Themis* assisting TCR signaling to low-affinity ligands (*19*). *Cd226*, a co-stimulatory molecule that plays an important role in antitumor response (*20*), was the most up-regulated gene upon combination therapy, followed by other genes involved in signal transduction and T cell differentiation/activation (*Pde4d*, *Vav1*, *Mki67*, and *Id2*) (*21–23*). Furthermore, combination treatment, in contrast to monotherapy, selectively prevented the up-regulation of genes involved in T cell differentiation/dysfunction, such as *Eomes*, *Mt1*, *Mt2*, and *Cd200*.

This activation and dysfunctional module score allowed us to quantitatively compare the complex T cell activation profile among different treatment groups. The activation module had significantly higher scores in the combination versus monotherapy treatment groups, indicative of a more activated T cell state [Fig. 2D; Kolmogorov-Smirnov (KS) test, *P* < 0.05].

Overall, although single-agent therapy expanded intratumoral CD8+ T cell clones and modulated critical gene pathways, this was not sufficient for complete and long-lasting tumor rejection (Fig. 1A). Our findings suggest that a profound reprogramming of dysfunctional tumor-infiltrating T cells by combination therapy was required for tumor rejection and long-term response. This result is supported by a recent study showing that CD8+ T cells quickly become dysfunctional at early stage of tumor development and gradually evolve into a less flexible state (*24*).

Anti–PD-1 and anti-GITR combination therapy modulates distinct tumor-specific CD8+ T cell populations

To determine whether distinct cellular mechanisms underlie anti–PD-1– and anti-GITR–mediated tumor rejection, we profiled tumor-infiltrating T cells by high-dimensional flow cytometry and used well-validated data-driven unsupervised clustering approaches (*25, 26*). Our attempt to use published mutated MC38 tumor epitopes (*27*) to track MC38-specific T cell clones was not successful. This possibly reflects the different mutation status of tumor cell lines between laboratories, likely due to intrinsic genome instability of the tumor cells (*28*). As an alternative approach, we generated MC38 tumor cell variant expressing H-2K^b single-chain trimer of major histocompatibility complex class I with SIINFEKL peptide (as a surrogate tumor epitope) and β_2m [ovalbumin (OVA)–β_2m-K^b; fig. S4A]. This allowed us to track Ag (OVA/SIINFEKL)–specific CD8+ T cells, which can be identified by H-2K^b–SIINFEKL pentamer staining (fig. S4B). Consistent with the TCR clonality analysis (fig. S3), anti–PD-1 Ab and combination treatment induced significant clonal expansion of OVA-specific CD8+ T cells intratumorally. Only the combination therapy significantly increased the intratumoral density (cell number normalized to milligrams of tumor tissue). A fluorescence-activated cell sorter (FACS)–based method was used, which allowed us to sample substantially more cells than the capacity of single-cell RNA-seq; thus, we were able to identify significant clonal expansion induced by combination therapy in spleen (fig. S4B). Functionally, clonally expanded OVA-specific T cells produced higher levels of interferon-γ (IFN-γ) upon OVA peptide restimulation (fig. S4C).

Next, we profiled T cells from tumor-bearing mice treated with anti-GITR and/or anti–PD-1 Ab using T cell differentiation/activation markers (e.g., PD-1, TIM3, LAG3, KLRG1, CD244, and CD44) and T cell lineage transcription factors (e.g., Eomes and Tbet). We detected heterogeneity within the Ag-specific CD8+ T cells, reflecting different activation/differentiation states (Fig. 3A). At day 9, T cell populations between treatment groups were indistinguishable; however, a marked skew in the population was observed by day 12 in response to anti-GITR Ab alone or combination treatment (Fig. 3B). To gain a more comprehensive understanding of the phenotypes of the T cell populations that were affected by the treatment, we analyzed Ag-specific CD8+ T cells using unsupervised spanning-tree progression analysis of density-normalized events (SPADE) clustering. Fourteen distinct Ag-specific CD8 T cell clusters (>2% relative frequency) were identified (Fig. 3C). Among them, clusters 14, 12, and 7 were highly responsive to anti-GITR and anti–PD-1 treatment (Fig. 3D). On the basis of the high expression level of T cell activation/dysfunction markers (PD-1,

---

Table 1. Pathways specifically up-regulated in clonal expanded CD8+ T cells with Ab treatment (day 11). Genes specifically up-regulated in monotherapy or combination therapy compared with isotype control were analyzed using Illumina Gene Ontology Engine. Jak, Janus kinase; STAT, signal transducers and activators of transcription.

Treatment	Pathway	P value
Anti–PD-1 + Anti-GITR	Adaptive immune response	1.50×10^{-08}
	Cell cycle	5.80×10^{-07}
	Metabolism of lipids and lipoproteins	3.00×10^{-04}
Anti–PD-1	Lymphocyte activation	2.30×10^{-05}
	Gene targets for miR-124u (T cell activation)	1.80×10^{-05}
	Jak-STAT signaling pathway	3.00×10^{-04}
Anti-GITR	Glucose metabolism	8.20×10^{-10}
	Protein metabolism	7.50×10^{-08}
	The citric acid cycle and respiratory electron transport	3.90×10^{-06}

TIM3, LAG3, CD244, Eomes, Tbet, and KLRG1), these three clusters likely represent different stages of dysfunction (Fig. 3E and fig. S5), unlike cluster 10 representing a more naïve/quiescent phenotype. Cluster 14 is absent at the earlier time point and has the highest expression level of markers associated with dysfunctional T cells, suggesting a terminal stage of dysfunction. All treatment groups were associated with a relative decrease in dysfunctional T cell frequency, and there was no significant difference between treatment groups (Fig. 3D). To evaluate the dysfunctional state (plastic versus nonprogrammable) (24), we examined the expression level of PD-1, LAG3, CD38, CD101, and CD5 on intratumoral Ag-specific T cells. Using an unsupervised hierarchical clustering algorithm [cluster identification, characterization, and regression (CITRUS), Cytobank], we found that combination and anti-GITR Ab treatment significantly decreased a T cell cluster with a profile matching the nonprogrammable dysfunctional T cells (PD-1^{hi}LAG3^{hi}CD38^{hi}CD101^{hi}CD5^{low}; Fig. 3, F and G).

To further characterize the responsive T cell populations, we used additional differentiation and activation markers to phenotypically dissect the T cells (e.g., Sca1, CD95, CD127, CD122, CD226, and Ki67). Unique cell populations can be identified upon treatment on day 12 (Fig. 4, A and B, and fig. S6). SPADE clustering was conducted using this specific panel of markers (Fig. 4C) to allow for statistical analysis between groups for a single cluster. We observed a specific effect of the combination treatment that was driven primarily by the anti-GITR Ab in expanding a unique memory cell population (cluster 11; Fig. 4, C to F) that may confer long-lasting protective immunity. These cells express CD62L and CD44 (Fig. 4E), similar to memory cells in the spleen compartment, while up-regulating additional key markers (Sca1, CD95, CD127, CD122, and Eomes), characteristic of a memory precursor effector cell phenotype (MPEC) as previously described (29) (Fig. 4F). Overall, combination therapy drives the expansion of MPECs, which represent a very small fraction of CD8$^+$ T$_{Effs}$ expressing CD127/IL7R. MPECs have high propensity to survive
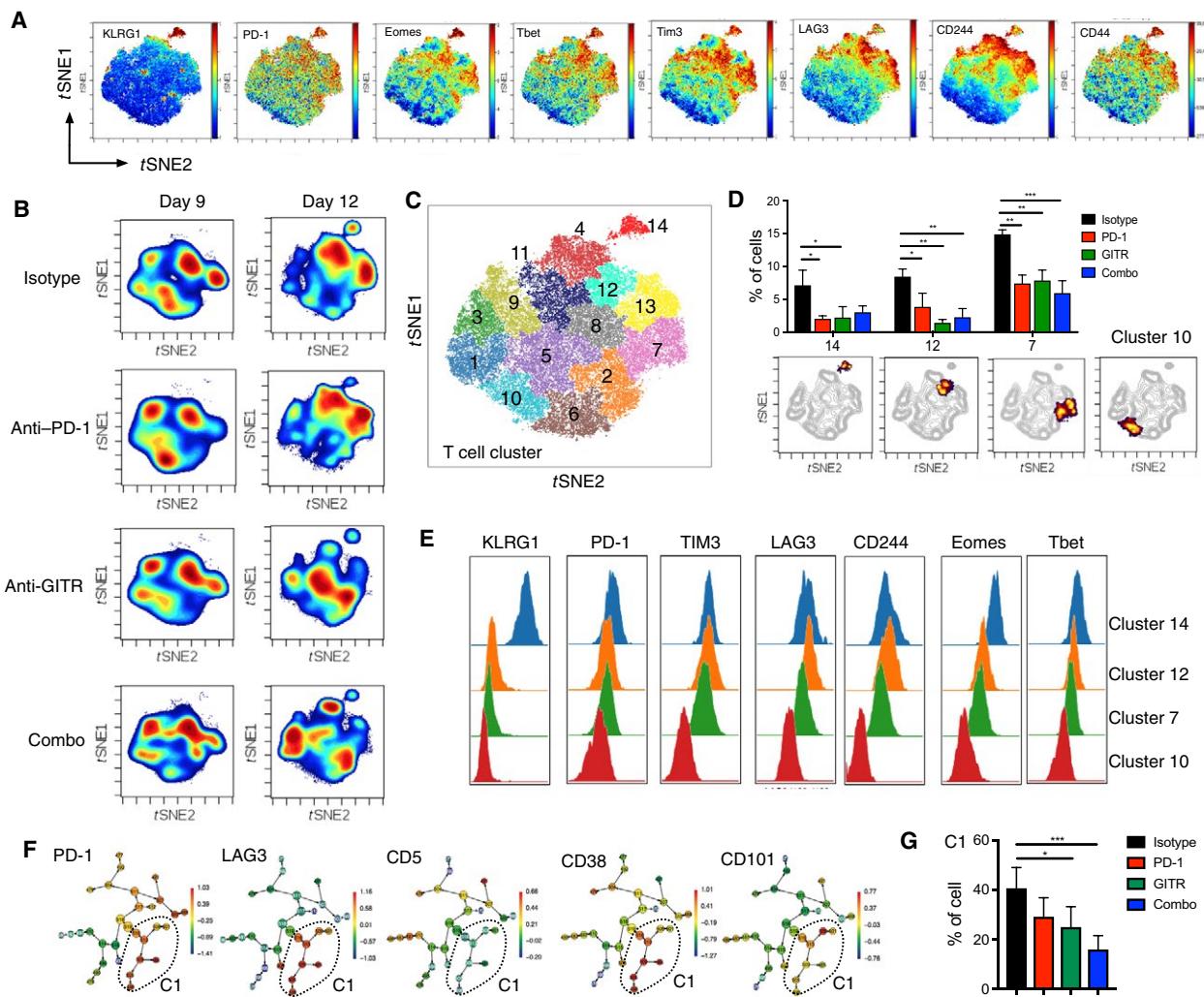
Fig. 3. Identification of combination treatment–responsive dysfunctional tumor-infiltrating CD8$^+$ T cell population. (**A**) Density viSNE plots of OVA-specific intratumoral CD8$^+$ T cells from each treatment group days 9 and 12 after tumor challenge. (**B**) viSNE plots of tumor-infiltrating T cells overlaid with the expression of indicated markers. (**C**) viSNE plot of intratumoral OVA-specific CD8$^+$ T cells overlaid with color-coded T cell clusters identified by SPADE. (**D**) Frequency of selected T cell clusters displayed on a per-mouse basis with means ± SEM. Indicated cluster was highlighted on viSNE plot below the bar graph. (**E**) Histogram displaying the expression level of indicated markers on T cell clusters. (**F**) CITRUS cluster result overlaid with indicated markers. Cluster that was significantly affected with Ab treatment is circled (C1). (**G**) Frequency of cluster C1 is shown with means ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001, one-way ANOVA test.
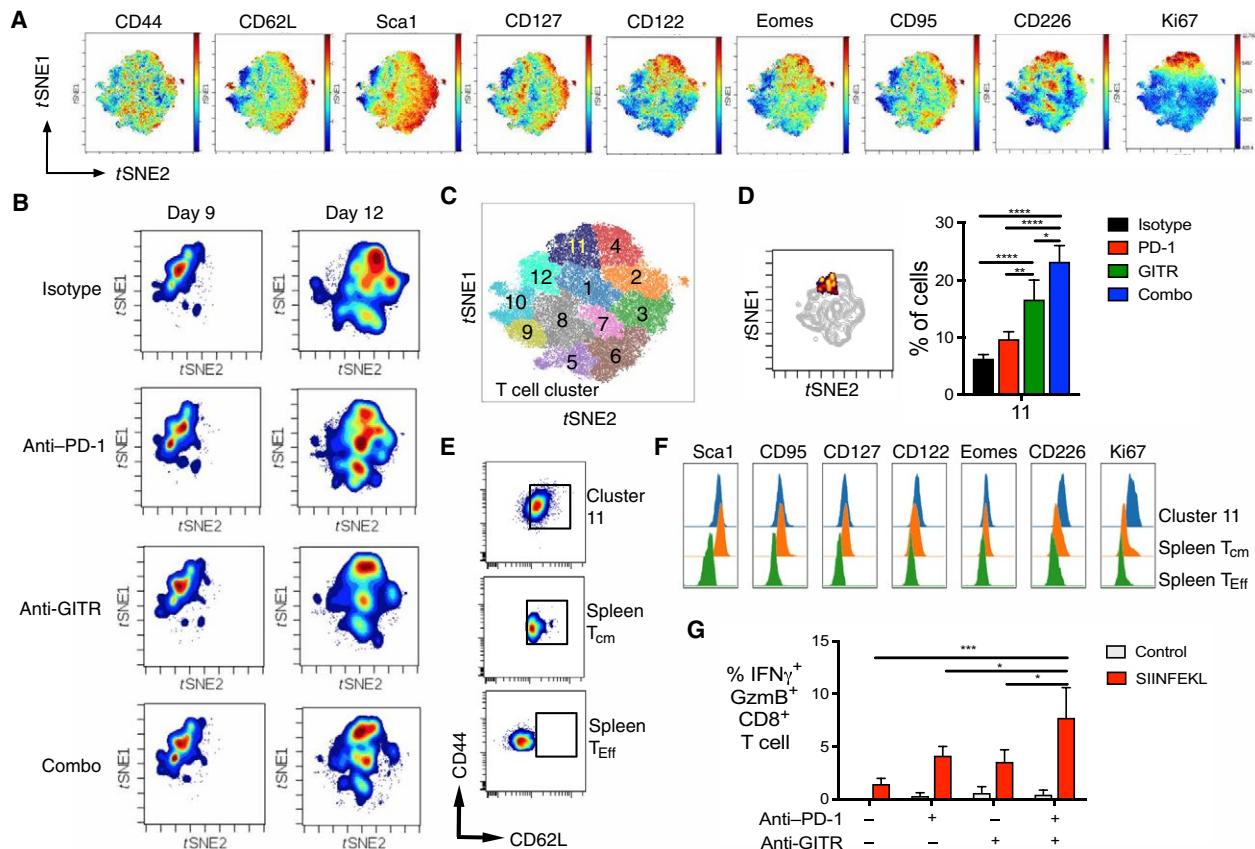
Fig. 4. Identification of combination treatment–responsive effector/memory tumor-infiltrating CD8⁺ T cell population. (**A**) viSNE plots of tumor-infiltrating T cells overlaid with the expression of selected markers. (**B**) Density viSNE plots of OVA-specific tumor-infiltrating CD8⁺ T cells from each treatment group days 9 and 12 after tumor challenge. (**C**) viSNE plot of MC38 infiltrating OVA-specific CD8⁺ T cells overlaid with color-coded T cell clusters identified by SPADE. (**D**) Frequency of selected T cell clusters (means ± SEM). Indicated cluster was highlighted on viSNE plot below the bar graph. (**E**) FACS plots of CD44 and CD62L expression on cluster 11 compared with spleen T cell subsets. T_{cm} central memory T cells. (**F**) Histogram displaying the expression level of selected markers on T cell clusters. (**G**) Frequency of cytokine production of CD8⁺ T cells upon restimulation in vitro with OVA peptide. GzmB, granzyme B. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, one-way ANOVA with Tukey's test).

and become functional memory T cells, which are essential to confer protective immunity (*3*). These cells also express high levels of CD226 and Ki67, suggesting an activated and highly proliferative state (Fig. 4F). Further, these intratumoral OVA-specific CD8⁺ T cells from cluster 11 showed increased IFN-γ, and granzyme B production upon Ag-specific restimulation followed single-agent treatment, but only combination treatment significantly enhanced their function (Fig. 4G).

In agreement with previous findings, we showed that although anti–PD-1 alone induces a less dysfunctional tumor-infiltrating lymphocyte (TIL) phenotype (Fig. 3), it is associated with low T_{EMs} (*30, 31*) and lack of long-term survival (Fig. 1) (*32*). We demonstrated that combination therapy can promote a memory CD8⁺ T cell phenotype, indicating that a costimulatory agonist Ab can synergize with PD-1 checkpoint blockade, leading to long-term antitumor responses. We further investigated the mechanism by which this specific combination therapy effect occurred by functional analysis of the identified up-regulated genes.

CD226 expression is induced on tumor-specific CD8⁺ T cells upon combination therapy

CD226, a costimulatory molecule with a well-validated role in the development of antitumoral immune response (*31*), was identified as the most up-regulated gene upon combination therapy (Fig. 1, B and

C, and table S3). We performed a *Cd226* expression analysis on different subsets of intratumoral CD8⁺ T cells (total, clonally expanded, or nonexpanded) across treatment groups (Fig. 5A), which revealed that *Cd226* mRNA levels were significantly increased by combination treatment on clonally expanded T cells (fold change, 10.7), whereas this difference was diluted in bulk/total CD8⁺ T cells (fold change, 3.5) and nonexpanded CD8⁺ T cells (not significant). Further, *Cd226* mRNA levels were significantly increased by combination treatment on clonally expanded CD8⁺ T cells in comparison to anti–PD-1 (fold change, 6.5) and anti-GITR (fold change, 9.2). Using the MC38-OVA-β₂m-K^b model, we found that protein levels of CD226 were highest on spleen OVA-specific CD8⁺ T cells after anti–PD-1 treatment (Fig. 5B) and were further elevated by combination treatment. The same treatment had no significant effect on the CD226 levels of nonspecific CD8⁺ T cells. This dataset suggests that anti–PD-1 treatment could play a dominant role in driving the increase of CD226 on Ag-specific T cells, providing key information on the mode of action of anti–PD-1 in antitumor immunity.

CD226 is a substrate for dephosphorylation by PD-1–SHP2

Next, we investigated a potential association between PD-1 signaling and CD226. The costimulatory receptor CD28 was previously identified as a target for PD-1–SHP2 dephosphorylation (*32*). Experiments using a cell-free reconstitution system in which the cytoplasmic domain

of PD-1 was bound to the surface of large unilamellar vesicles (LUVs) that mimic the plasma membrane of T cells demonstrated that CD28 is preferred over the TCR as a target for SHP2-mediated dephosphorylation. Using the same system, we examined whether CD226 could be another target for dephosphorylation by the PD-1–SHP2 complex. We reconstituted different components (CD3, CD226, etc.) involved in T cell signaling on the liposomes (Fig. 5C) (*33*). The sensitivity of each component in response to PD-1 titration on the LUVs was measured by phosphotyrosine (pY) Western blots. We confirmed previous published data (*32*) showing that CD28, but not inducible T cell costimulator (ICOS), is a substrate for desphosphorylation by PD-1–SHP2 (Fig. 5D). We found that CD226 was efficiently dephosphorylated by PD-1–SHP2 in a dose-dependent manner (Fig. 5D) in the presence of

LFA-1. Although LFA-1 facilitates T cell activation by lowering the amounts of antigen necessary for T cell activation (*33*), cross-linking of LFA-1 induces tyrosine phosphorylation of DNAM-1 (CD226) (*34*). Although additional experiments are needed to fully characterize the basis of CD226 sensitivity to PD-1–SHP2–induced dephosphorylation, we demonstrated a biochemical association between PD-1 signaling and CD226 dephosphorylation using a cell-free biochemical system. To validate these findings in a more physiologically relevant system involving primary cells from the tumor microenvironment, we isolated Ag-specific intratumoral CD8[+] T cells from MC38-OVA tumor–bearing mice (day 11) previously treated with anti–PD-1 Ab or isotype control Ab. Consistent with the cell-free in vitro assay, we found that CD8[+] TILs from the PD-1 Ab–treated group showed high CD226
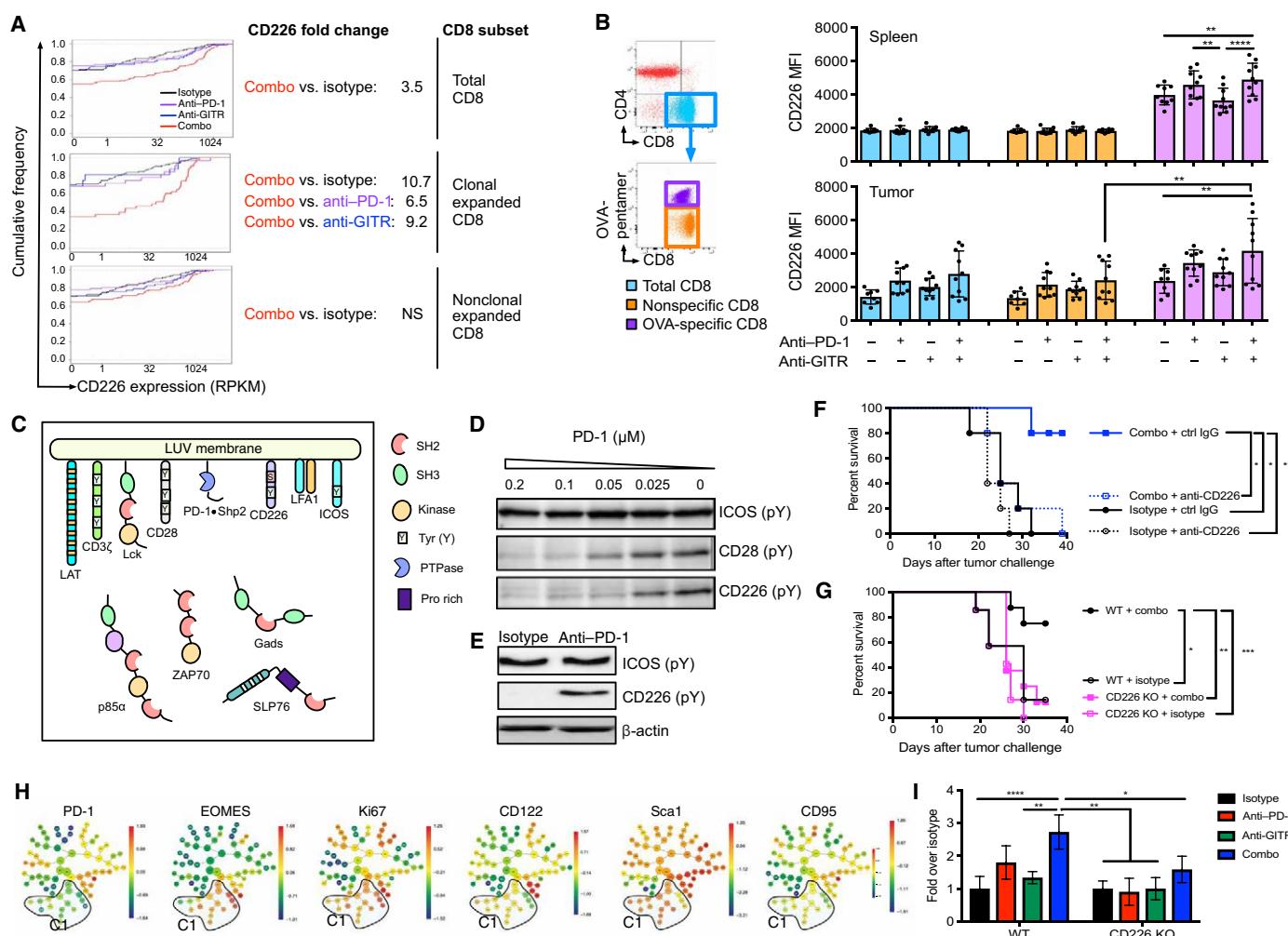
Fig. 5. CD226 signaling pathway is essential in mediating antitumor response induced by combination treatment in MC38 tumor model. (**A**) Cumulative distribution function plots show the expression of *CD226* in total, clonal expanded, or nonexpanded CD8[+] T cells. Fold changes of *CD226* expression level are indicated for each subset. RPKM, reads per kilobase million. (**B**) FACS analysis of CD226 expression [mean fluorescence intensity (MFI)] on spleen/tumor CD8[+] T cell populations, gating strategy indicated in FACS plots. Data show one representative experiment of two independent experiments (*n* = 10 mice per group). (**C**) Schematic shows LUVs. (**D**) Western blot shows phosphorylation status of ICOS, CD28, and CD226 with increasing PD-1 concentrations. (**E**) Western blot shows phosphorylation status of ICOS and CD226 from intratumoral OVA-specific CD8[+] T cells purified from mice treated with isotype or anti–PD-1 Ab. (**F**) MC38 tumor–bearing mice were treated with CD226-blocking Ab or isotype IgG before immunotherapy with anti-GITR and anti–PD-1 or isotype IgGs. The percentages of survival are shown here. Data show one representative experiment of three independent experiments (*n* = 5 mice per group). (**G**) CD226 KO mice or WT littermates were challenged with MC38 tumor cells and treated with either anti-GITR and anti–PD-1 Ab or isotype Abs on days 6 and 13 after tumor implantation. Data show one representative experiment of two independent experiments (*n* = 7 to 8 mice per group). (**H**) CITRUS cluster result overlaid with indicated markers. Cluster C1 that was significantly affected with Ab treatment is circled. (**I**) Frequency of T cell cluster C1 displayed on a per-mouse basis with means ± SEM. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, one-way ANOVA with Tukey's test.

phosphorylation upon PD-1 inhibition (Fig. 5E), confirming that CD226 is in an activated state. CD8[+] TILs from isotype control–treated mice showed a lack of CD226 phosphorylation, indicating that PD-1/PD-L1 pathway is a negative regulator of CD226 activity on intratumoral CD8[+] T cells. These data suggest that CD226 is an additional substrate for dephosphorylation by PD-1–SHP2.

Down-regulation of TIGIT expression upon GITR Ab treatment shifts the CD226/TIGIT pathway toward stronger costimulation

It has recently been shown that the strength of CD8[+] T cell response is affected by the overall balance between CD226 and the coinhibitory receptor TIGIT (31). Single-cell RNA-seq data indicated that anti-GITR Ab treatment increased TIGIT transcripts in high-frequency T cell clones (fig. S7A), whereas FACS analysis showed lower expression of TIGIT on OVA-specific CD8[+] T cells with anti-GITR or combination treatment compared with anti–PD-1 treatment alone (fig. S7B). This result is consistent with the observation that TIGIT expression is tightly regulated at the posttranscriptional level (35). Combination treatment significantly decreased the percentage of TIGIT[+] cells and the expression level on a per-cell basis on total tumor-infiltrating CD8[+] and CD4[+] T$_{Effs}$ and T$_{regs}$, the effect of which was mainly driven by anti-GITR Ab treatment (fig. S7, C and D). Combination and/or monotherapy treatment had no effect on the percentage of CD226[+] tumor-infiltrating CD4[+] T cells, splenic CD4[+] T cells, or splenic T$_{regs}$ (fig. S7E). TIGIT expression level (fig. S7F) on the dysfunctional T cell cluster 14 (identified in Fig. 3D) is uniformly high (~100%), whereas the memory-like T cell cluster 11 (identified in Fig. 4D) shows much lower levels of TIGIT expression.

In addition, because GITR is highly expressed on T$_{regs}$, we performed a comprehensive profiling [t-distributed stochastic neighbor embedding (t-SNE)–based visualization (viSNE)/SPADE] analysis on the remaining intratumoral T$_{regs}$ upon combination treatment and compared the results to a well-studied T$_{reg}$-depleting Ab (anti-CD25, PC61). We found that GITR and PD-1 combination therapy has a distinct effect on the phenotype of the remaining T$_{regs}$ compared with the anti-CD25 Ab treatment group. GITR and PD-1 combination treatment skewed the remaining T$_{reg}$ subsets toward a less suppressive phenotype, as defined by a lower expression of TIGIT, KLRG1, LAG3, and TIM3 (fig. S8). All these markers have been reported to identify a highly activated and phenotypically suppressive T$_{reg}$ population (36, 37). To assess whether the distinct effect of GITR and CD25 Abs on T$_{reg}$ depletion could differentially alter the phenotype of CD8[+] T cells, we analyzed the intratumoral OVA-specific CD8[+] T cells (fig. S9). Using an extended 18-parameter viSNE analysis, we found that GITR and PD-1, but not CD25 and PD-1, combination Ab treatment significantly decreased the dysfunctional phenotype (cluster 1: low Ki67, CD226, Sca1, CD95, CD122, high KLRG1, PD-1, LAG3, and CD244) and the naïve-like T cells (cluster 10: CD62L[+]CD44[−], low Ki67, CD226, PD-1, and other T cell activation markers), whereas it increased the cluster with memory-like activated T cell phenotype (cluster 6: high CD226, Ki67, Sca1, CD127, CD95, and CD122 and lower level of PD-1, TIM3, LAG3, KLRG1, and CD244). These data show that the shift in balance between CD226 and TIGIT is driven by the direct combination of PD-1 and GITR Abs and not solely through a reduction in intratumoral T$_{regs}$ (fig. S9, E and F).

Overall, the single-cell RNA-seq and FACS phenotyping data showed that anti–PD-1 favored the increased expression of CD226, whereas anti-GITR treatment down-regulated surface expression of TIGIT and reduced intratumoral T$_{reg}$ subsets with a highly suppressive phenotype, therefore synergistically restoring the homeostatic CD8[+] T cell function.

CD226 signaling pathway is essential in mediating the antitumor response induced by combination treatment

Using a CD226-blocking monoclonal Ab, we showed that costimulatory signaling through CD226 is required for the antitumor immunity mediated by combination treatment (Fig. 5F). To validate the T cell intrinsic role of CD226, we genetically inactivated it in C57BL/6 background mice (fig. S10, A and B). CD226 knockout (KO) mice showed no defect on T cell (CD4[+], CD8[+], and T$_{regs}$) development and homeostasis (fig. S10, C and E) or responsiveness to TCR activation when compared with wild-type (WT) littermates (fig. S10F). After tumor challenge in the CD226 KO mice, we found that combination treatment with anti–PD-1 and anti-GITR no longer conferred any antitumor effect or survival benefit, suggesting that CD226 is essential for the antitumor effect of the combination therapy (Fig. 5G). In addition, we validated the specificity of the CD226 pathway mediating this effect because inhibition of TNFR superfamily pathways (OX40/OX40L or 4-1BB/4-1BBL) or blockade of the B7 costimulatory molecule (CD28) with CTLA4-Ig had no impact on the antitumor effect observed with the combination treatment (fig. S11, A to C). In mice lacking CD226, combination therapy can no longer maintain the activated and highly proliferative memory T cell phenotype (PD-1[+/−], Eomes[+], Ki67[+], CD122[+], Sca1[+], and CD95[+]), as observed in WT control tumor-bearing mice (Fig. 5, H and I).

To test whether the mechanisms found in the MC38 model also extend to other tumor cell models, we evaluated the role of PD-1 and GITR combination therapy in the RENCA tumor model (murine kidney carcinoma tumor in Balb/c background). This tumor cell line was chosen because it also showed relatively good expression of CD155, the endogenous ligand for CD226 (fig. S12). Consistent with the MC38 studies, anti-GITR and anti–PD-1 combination therapy can synergize and promote long-term survival of RENCA tumor–bearing mice (Fig. 6A). FACS analysis of tumor-infiltrating T cells revealed a significant reduction in T$_{reg}$ driven by GITR Ab and expansion of tumor-specific CD8[+] T cells (gp70-tetramer[+]) driven primarily by PD-1 Ab (Fig. 6, B and C). In addition, we also observed a significant increase of CD226 expression on tumor-specific CD8[+] T cells (Fig. 6D, gp70 tetramer[+] cells), consistent with our observations in the MC38 tumor model. viSNE and SPADE analyses (Fig. 6, E to G) revealed an increase in memory-like effector cell cluster (Fig. 6G, C5; high CD95, CD122, Ki67, and Tbet and low CD44 and TIM3) accompanied by reduced dysfunctional T cell cluster (Fig. 6G, C1; high TIM3 and low Ki67, CD95, CD122, Tbet, and Sca1). Likewise, when CD226 signaling is blocked by an anti-CD226 Ab during combination treatment, the antitumor efficacy is completely abrogated (Fig. 6I), as previously demonstrated in the MC38 tumor model. Thus, our data provide additional evidence that the role of CD226 is not limited to one tumor model or mouse genetic background.

DISCUSSION

We systematically profiled tumor-infiltrating CD8[+] T cells using single-cell TCR sequencing, transcriptomics, and high-dimensional flow cytometry clustering analysis to unveil the molecular mechanisms driving the potent synergism of a costimulatory agonist,

anti-GITR Ab, and a coinhibitory antagonist, anti–PD-1 Ab. We showed that a combination of these Abs synergistically enhanced the effector function of expanded CD8$^+$ T cells by restoring the balance of two key homeostatic regulators, CD226 and TIGIT, resulting in robust survival benefit.

Recently, it has been shown that T cell dysfunction represents an important hurdle for the generation of durable responses upon PD-1 inhibition (3). PD-1 blockade induces an incomplete rescue of T$_{Ex}$, thus failing to restore T$_{Ex}$ into T$_{EM}$, a necessary requirement for effective immune memory (38, 39). We demonstrated that combination therapy not only decreased the intratumoral CD8$^+$ T cells with a nonprogrammable dysfunctional profile (Fig. 3, F and G) (31, 40) but also induced a highly proliferative TEM precursor population (Fig. 4) (32, 41), thus eliciting potent antitumor immunity in a CD226-dependent manner (Figs. 5, F and G, and 6I). The similarity in the efficacy between both MC38 and RENCA tumor models despite different mouse genetic backgrounds and tumor types (C57BL/6

and Balb/c; colon and kidney adenocarcinoma, respectively) suggests that this mechanism is not limited to a specific tumor model. However, further research is needed to determine whether the reduction of dysfunctional CD8$^+$ T cells is through the reinvigoration of these populations and/or epigenetic reprogramming of them into a functional state (24).

Furthermore, we validated our findings using gene pathway analysis to show that although anti-GITR and anti–PD-1 monotherapies regulated distinct molecular pathways, combination treatment synergistically integrated the pathways modulated by each single agent, resulting in a distinct transcriptional state of CD8$^+$ T cells (Table 1). In addition, applying to our data the uncoupled T cell dysfunction and activation gene expression module score, as described by Singer et al. (17), enabled us to quantitatively compare the complex T cell activation profile between different treatment groups. We found that the activation module had significantly higher scores in the combination versus monotherapy treatments groups, indicative of a more activated
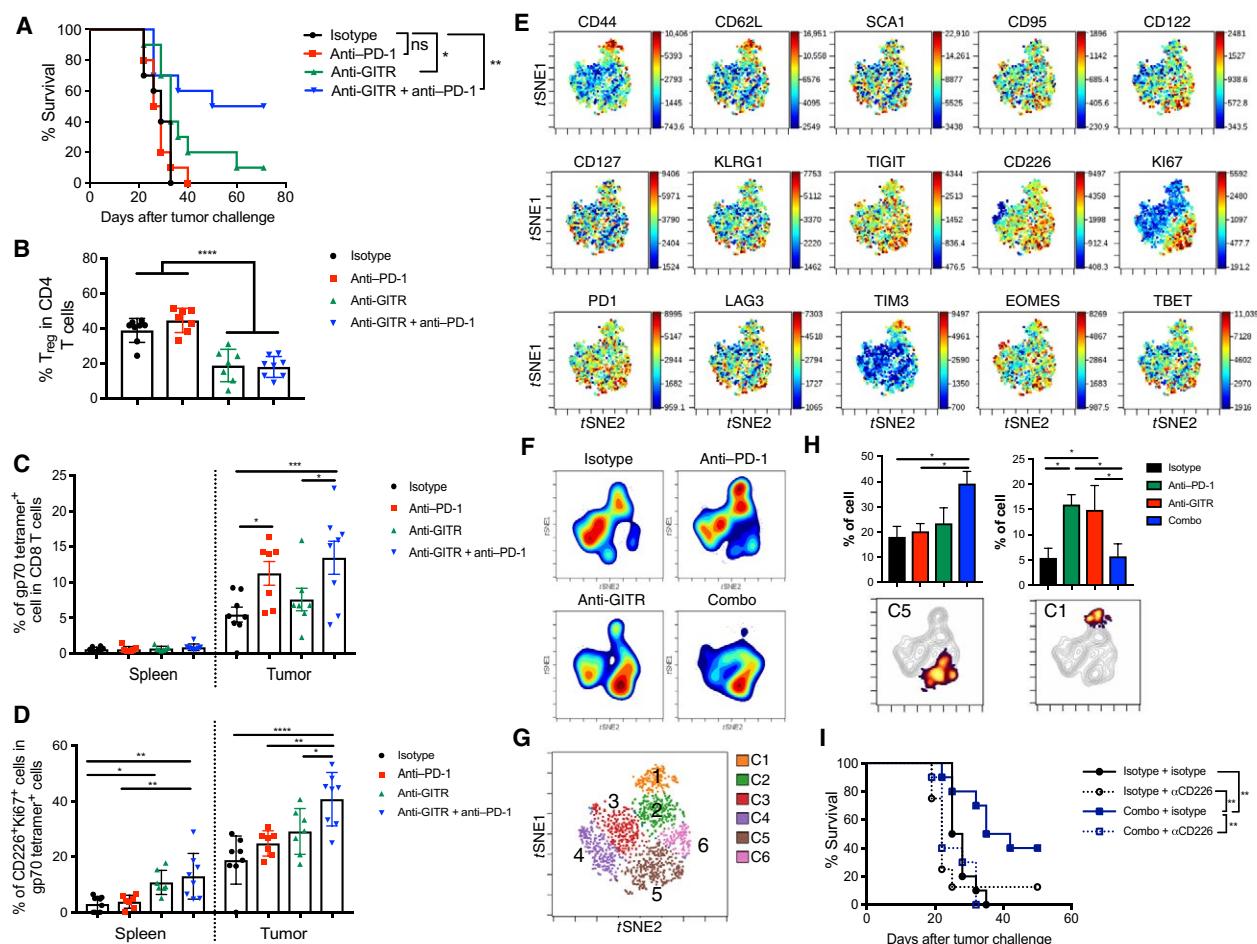

Fig. 6. CD226 signaling pathway is essential in mediating antitumor response induced by combination treatment in RENCA tumor model. (**A**) RENCA tumor–bearing mice were treated with either CD226-blocking Ab or isotype IgG before immunotherapy with anti-GITR and anti–PD-1 or isotype IgGs. The percentages of survival are shown here (n = 8 mice per group). ns, not significant. (**B**) Effect of Ab treatment on intratumoral T$_{reg}$ frequency (n = 7 to 8 mice per group). (**C**) Frequency of tumor and spleen gp70-specific CD8$^+$ T cells from tumor-bearing mice treated with anti-GITR and/or anti–PD-1 Ab (n = 8 mice per group). (**D**) Frequency of CD226$^+$Ki67$^+$ cell in gp70-specific CD8$^+$ T cell population (n = 7 to 8 mice per group). (**E**) viSNE plot of tumor-infiltrating T cells overlaid with the expression of selected markers. (**F**) Density viSNE plot of gp70-specific tumor-infiltrating CD8$^+$ T cells from each treatment group day 11 after tumor challenge. (**G**) viSNE plot of RENCA infiltrating gp70-specific CD8$^+$ T cells overlaid with color-coded T cell clusters identified by SPADE. (**H**) Frequency of selected T cell clusters (means ± SEM). Indicated cluster was highlighted on viSNE plot below the bar graph. (**I**) RENCA tumor–bearing mice were treated with CD226-blocking Ab or isotype IgG before immunotherapy with anti-GITR and anti–PD-1 or isotype IgGs. The percentages of survival are shown here (n = 10 mice per group). * P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, one-way ANOVA with Tukey's test.

T cell state (Fig. 2, B to D). Last, we confirmed *CD226* as the most up-regulated gene upon combination treatment.

We developed a new bioinformatic pipeline rpsTCR that allowed us to selectively profile clonally expanded intratumoral CD8[+] T cells, which displayed distinct gene signature upon treatment with anti–PD-1 and/or anti-GITR Ab. Expression analysis of different subsets of intratumoral CD8[+] T cells (total, clonally expanded, or nonexpanded) across treatment groups revealed that *CD226* mRNA level was significantly increased by combination treatment in clonally expanded T cells, although this difference was diluted and lost in bulk and nonexpanded CD8[+] T cells. This observation stresses the importance of performing genome profiling on putative tumor-reactive clones (high-frequency T cell clones) to unmask critical gene changes. These genetic findings were validated at the protein level by comparing cell surface CD226 expression on tetramer-positive, antigen-specific intratumoral CD8[+] T cells after different Ab treatments. One potential limitation of the MC38 tumor model is that we tracked the clonally expanded CD8[+] T cells using a surrogate OVA antigen due to lack of reagents to track endogenous tumor antigen–specific CD8[+] T cells. However, we were able to validate the findings using the RENCA tumor model and endogenous tumor antigen–specific T cells.

Recently, both cis- and trans-inhibitory mechanisms have been proposed for the TIGIT/CD226 signaling pathway (*31*). Therefore, the net output of this pathway may result from the balance between the expression level of CD226 on CD8[+] T cells and TIGIT on both CD8[+] T cells and bystander lymphocytes. We found CD226 as a previously unidentified target for PD-1–initiated SHP2 dephosphorylation (Fig. 7) in addition to CD28 (*32*), thus revealing additional complexity of the PD-1 pathway. In addition, we showed that PD-1 blockade rescued CD226 dephosphorylation by PD-1–SHP2, whereas GITR agonism reduced TIGIT expression, which competes for CD155/PRV binding with CD226 (*42*). Therefore, only the combination treatment restored T cell effector function by favorably tipping the balance between CD226 and TIGIT leading to durable antitumor responses (model shown in Fig. 7).

In summary, the systematic approaches used in this study can shed light on important functional T cell regulatory pathways implicated in the synergy between two immunotherapeutic Abs, representing distinct immunotherapeutic modalities (costimulatory agonism and
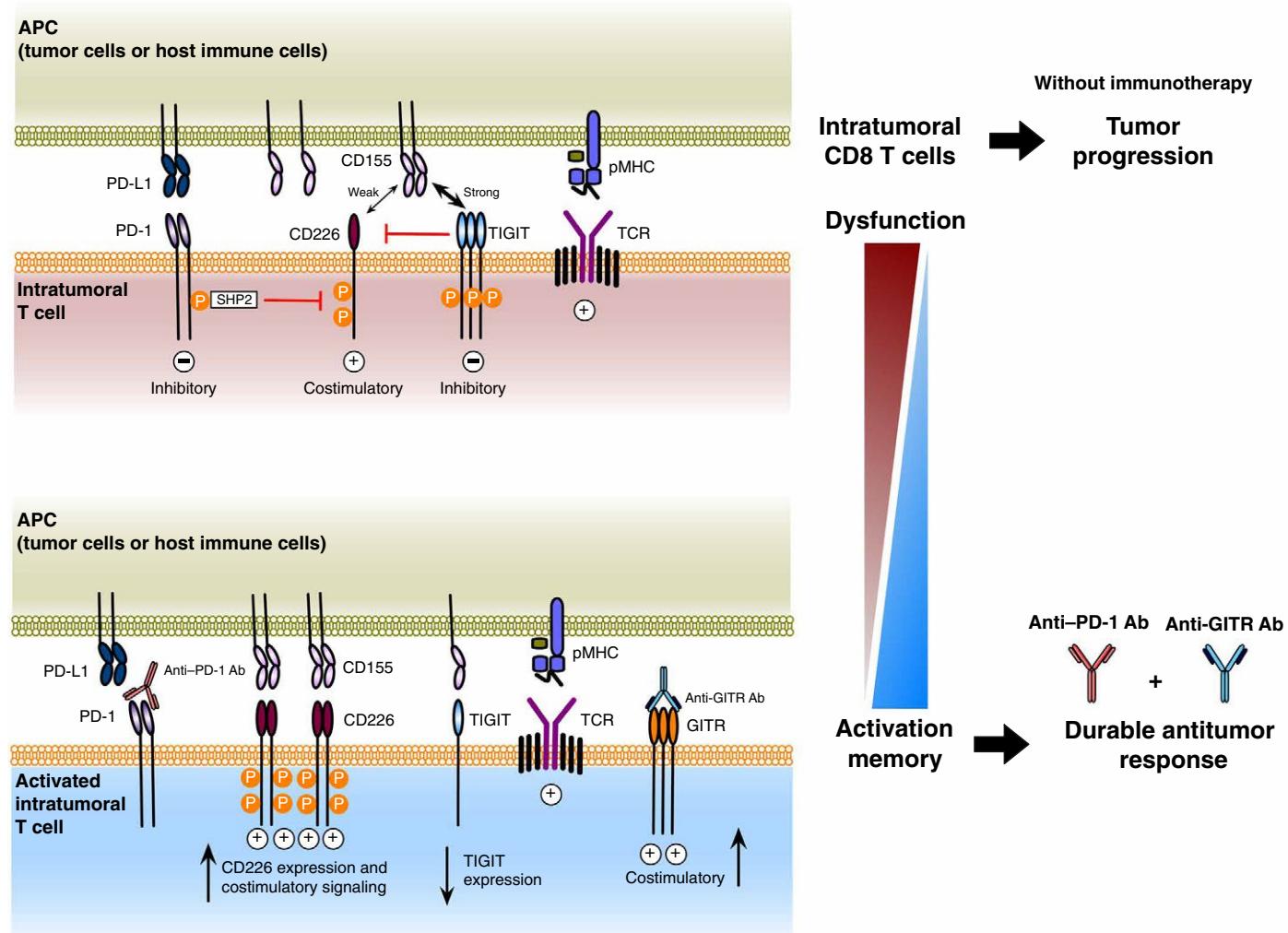

Fig. 7. Proposed model for anti–PD-1 and anti-GITR Ab combination therapy. pMHC, peptide–major histocompatibility complex; APC, antigen-presenting cells.

checkpoint inhibition). Similar methodology may be useful in future clinical trials of cancer immunotherapy to help unmask the molecular pathways driving durable antitumor responses.

MATERIALS AND METHODS

Study design

The aim of this study was to dissect the molecular and cellular mechanisms mediated by combination immunotherapy. Single-cell RNA-seq analysis study of tumor-infiltrating CD8[+] T cells was designed to identify the specific gene signature induced upon treatment. Single-cell suspensions were profiled by high-dimensional computational flow cytometry to identify phenotypic changes in cell populations. To validate our findings, we designed in vitro biochemistry experiments and additional in vivo studies using either genetically engineered mice or blocking Abs. Control and experimental treatments were administered to age- and sex-matched mice. Investigators were blinded for tumor measurements. The number of experimental replicates is indicated in the figure legends.

Mice and treatments

Six- to eight-week-old female C57BL/6 and Balb/c mice were obtained from the Jackson laboratory. CD226[−/−] mice on a C57BL/6 background were generated at Regeneron Pharmaceuticals Inc. using the VelociGene method (43, 44). Briefly, enhanced green fluorescent protein complementary DNA (cDNA) was inserted in-frame to the start codon, followed by a selection cassette that disrupts transcription of the gene body and results in a CD226 null allele. Heterozygous targeted mice were interbred to produce homozygous KO mice for study. All animals were maintained under pathogen-free conditions, and experiments were performed according to protocols approved by the Institute of Animal Care and Use Committee of Regeneron Pharmaceuticals Inc. For tumor studies, 3×10^5 MC38 or 1×10^6 RENCA cells were subcutaneously injected on the right flank of age-matched C57BL/6 or Balb/c, respectively. Six days after tumor implantation, mice (randomly distributed in different groups) were grouped on the basis of tumor size and treated by intraperitoneal injection with anti-GITR (5 mg/kg; DTA-1) and/or anti–PD-1 (RPM1-14) Ab or isotype control IgGs (rat IgG2b, LTF-2 and rat IgG2a, 2A3; Bio X Cell). Abs were administered again on day 13. Depletion/blocking Abs were given 1 to 2 days before immunotherapy. Perpendicular tumor diameters were measured blindly two to three times weekly using digital calipers (VWR, Radnor, PA). Volume was calculated using the formula $L \times W \times W \times 0.5$, where L is the longest dimension and W is the perpendicular dimension.

Flow cytometry

For flow cytometry analysis of in vivo experiments, blood, spleen, thymus, lymph node, and tumor were harvested on indicated days after treatment. Single-cell suspensions were prepared, and red blood cells were lysed using ACK lysis buffer (Thermo Fisher Scientific). Live/dead cell discrimination was performed using Live/Dead Fixable Blue Dead Cell Staining Kit (Thermo Fisher Scientific). Ag-specific CD8[+] T cells were detected with H-2K[b]/OVA SIINFEKL-Pentamer (ProImmune) for MC38-OVA model and H-2L[d]/MuLV gp70-SPSYVYHQF tetramer (MBL International) for RENCA model. For intracellular cytokine staining, cells were stimulated with or without SIINFEKL peptide for 36 hours and with a protein transport inhibitor (BD Biosciences) for the last 4 hours. After stimulation, cells were stained as described above for surface and intracellular proteins.

To quantify cell numbers in tissue, a fixed number of CountBright absolute counting beads (Thermo Fisher Scientific) were added to each sample before acquiring. Samples were acquired on LSRFortessa X-20, LSR II, or Symphony (BD Biosciences) and analyzed using FlowJo (TreeStar) and Cytobank. Further details are in Supplementary Materials and Methods.

Single-cell sorting RNA-seq

FACS-sorted tumor and spleen CD8[+] T cells were mixed with C1 cell suspension reagent (Fluidigm) before loading onto a 5- to 10-μm C1 integrated fluidic circuit (Fluidigm). Cell lysing, reverse transcription, and cDNA amplification were performed on the C1 Single-Cell Auto Prep Integrated Fluidic Circuit, as specified by the manufacturer (protocol 100-7168 E1). The SMARTer Ultra Low RNA Kit (Clontech) was used for cDNA synthesis from the single cells. Illumina next-generation sequencing libraries were constructed using the Nextera XT DNA Sample Prep Kit (Illumina). Cells were sequenced on Illumina NextSeq (Illumina) by multiplexed single-read run with 75 cycles. Raw sequence data (BCL files) were converted to FASTQ format via Illumina Casava 1.8.2. Reads were decoded on the basis of their barcodes. Read quality was evaluated using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Further details are in Supplementary Materials and Methods.

LUV reconstitution and phosphotyrosine Western blot

LUVs reconstituted different components (ICOS, CD28, and CD226) involved in T cell signaling on the liposomes, together with LCK, ZAP70, SLP76, and PI3K as previously described (45–47). Proteins of interest were premixed at desired ratios in 1× reaction buffer containing bovine serum albumin (0.5 mg/ml) and then mixed with LUVs (1 mM total lipids) as described previously (32). Fifty micrograms of protein for each sample was used for the Western blot. Further details are in Supplementary Materials and Methods.

Statistical analysis

Sample sizes were chosen empirically to ensure adequate statistical power and were in the line with field standards for the techniques used in the study. Statistical significance was determined with analysis of variance (ANOVA) or unpaired two-tailed Student's t test, assuming unequal variance at $P < 0.05$ level of significance (or indicated in figure legends).

SUPPLEMENTARY MATERIALS

immunology.sciencemag.org/cgi/content/full/3/29/eaat7061/DC1

Fig. S1. T cell depletion with Abs.

Fig. S2. Bioinformatic pipeline rpsTCR.

Fig. S3. Combination therapy expands intratumoral high-frequency tumor-reactive CD8[+] T cell clones.

Fig. S4. Combination treatment expands tumor antigen–specific CD8[+] T cells with effector function.

Fig. S5. Mean fluorescence intensity of markers for dysfunctional cell clusters identified in Fig. 2.

Fig. S6. Mean fluorescence intensity of markers for effector/memory cell clusters identified in Fig. 3.

Fig. S7. TIGIT expression at single-cell RNA level and FACS analysis of TIGIT/CD226 expression level on different T cell subsets.

Fig. S8. GITR and PD-1 combination treatment significantly reduced highly activated T_{reg} subsets.

Fig. S9. GITR and PD-1 combination treatment induced intratumoral CD8[+] T cell subsets distinct from CD25 and PD-1 combination therapy.

Fig. S10. CD226[−/−] mice show normal T cell development and homeostatic function.

REFERENCES AND NOTES

1. S. C. Wei, C. R. Duffy, J. P. Allison, Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discov.* **8**, 1069–1086 (2018).

2. Y. Iwai, J. Hamanishi, K. Chamoto, T. Honjo, Cancer immunotherapies targeting the PD-1 signaling pathway. *J. Biomed. Sci.* **24**, 26 (2017).

3. K. E. Pauken, M. A. Sammons, P. M. Odorizzi, S. Manne, J. Godec, O. Khan, A. M. Drake, Z. Chen, D. R. Sen, M. Kurachi, R. A. Barnitz, C. Bartman, B. Bengsch, A. C. Huang, J. M. Schenkel, G. Vahedi, W. N. Haining, S. L. Berger, E. J. Wherry, Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).

4. D. R. Sen, J. Kaminski, R. A. Barnitz, M. Kurachi, U. Gerdemann, K. B. Yates, H.-W. Tsao, J. Godec, M. W. LaFleur, F. D. Brown, P. Tonnerre, R. T. Chung, D. C. Tully, T. M. Allen, N. Frahm, G. M. Lauer, E. J. Wherry, N. Yosef, W. N. Haining, The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).

5. S. Yao, Y. Zhu, L. Chen, Advances in targeting cell surface signalling molecules for immune modulation. *Nat. Rev. Drug Discov.* **12**, 130–146 (2013).

6. D. A. Knee, B. Hewes, J. L. Brogdon, Rationale for anti-GITR cancer immunotherapy. *Eur. J. Cancer* **67**, 1–10 (2016).

7. L. L. Siu, N. Steeghs, T. Meniawy, M. Joerger, J. L. Spratlin, S. Rottey, A. Nagrial, A. Cooper, R. Meier, X. Guan, P. Phillips, G. Bajaj, J. Gokemeijer, A. J. Korman, K. L. Aung, M. S. Carlino, Preliminary results of a phase I/IIa study of BMS-986156 (glucocorticoid-induced tumor necrosis factor receptor–related gene [GITR] agonist), alone and in combination with nivolumab in pts with advanced solid tumors. *J. Clin. Oncol.* **35**, 104 (2017).

8. C. T.-S. Victor, A. J. Rech, A. Maity, R. Rengan, K. E. Pauken, E. Stelekati, J. L. Benci, B. Xu, H. Dada, P. M. Odorizzi, R. S. Herati, K. D. Mansfield, D. Patsch, R. K. Amaravadi, L. M. Schuchter, H. Ishwaran, R. Mick, D. A. Pryma, X. Xu, M. D. Feldman, T. C. Gangadhar, S. M. Hahn, E. J. Wherry, R. H. Vonderheide, A. J. Minn, Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. *Nature*, 1–18 (2015).

9. L. Lu, X. Xu, B. Zhang, R. Zhang, H. Ji, X. Wang, Combined PD-1 blockade and GITR triggering induce a potent antitumor immunity in murine cancer models and synergizes with chemotherapeutic drugs. *J. Transl. Med.* **12**, 36 (2014).

10. S. I. S. Mosely, J. E. Prime, R. C. A. Sainson, J.-O. Koopmann, D. Y. Q. Wang, D. M. Greenawalt, M. J. Ahdesmaki, R. Leyland, S. Mullins, L. Pacelli, D. Marcus, J. Anderton, A. Watkins, J. Coates Ulrichsen, P. Brohawn, B. W. Higgs, M. McCourt, H. Jones, J. A. Harper, M. Morrow, V. Valge-Archer, R. Stewart, S. J. Dovedi, R. W. Wilkinson, Rational selection of syngeneic preclinical tumor models for immunotherapeutic drug discovery. *Cancer Immunol. Res.* **5**, 29–41 (2017).

11. S. Chen, L.-F. Lee, T. S. Fisher, B. Jessen, M. Elliott, W. Evering, K. Logronio, G. H. Tu, K. Tsaparikos, X. Li, H. Wang, C. Ying, M. Xiong, T. VanArsdale, J. C. Lin, Combination of 4-1BB agonist and PD-1 antagonist promotes antitumor effector/memory CD8 T cells in a poorly immunogenic tumor model. *Cancer Immunol. Res.* **3**, 149–160 (2015).

12. A. E. Mahne, S. Mauze, B. Joyce-Shaikh, J. Xia, E. P. Bowman, A. M. Beebe, D. J. Cua, R. Jain, Dual roles for regulatory T-cell depletion and costimulatory signaling in agonistic GITR targeting for tumor immunotherapy. *Cancer Res.* **77**, 1108–1118 (2017).

13. Y. Bulliard, R. Jolicoeur, M. Windman, S. M. Rue, S. Ettenberg, D. A. Knee, N. S. Wilson, G. Dranoff, J. L. Brogdon, Activating Fc γ receptors contribute to the antitumor activities of immunoregulatory receptor-targeting antibodies. *J. Exp. Med.* **210**, 1685–1693 (2013).

14. E. Cha, M. Klinger, Y. Hou, C. Cummings, A. Ribas, M. Faham, L. Fong, Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci. Transl. Med.* **6**, 238ra70 (2014).

15. A. Han, J. Glanville, L. Hansmann, M. M. Davis, Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).

16. L. Robert, C. Harview, R. Emerson, X. Wang, S. Mok, B. Homet, B. Comin-Anduix, R. C. Koya, H. Robins, P. C. Tumeh, A. Ribas, Distinct immunological mechanisms of CTLA-4 and PD-1 blockade revealed by analyzing TCR usage in blood lymphocytes. *Oncoimmunology* **3**, e29244 (2014).

17. M. Singer, C. Wang, L. Cong, N. D. Marjanovic, M. S. Kowalczyk, H. Zhang, J. Nyman, K. Sakuishi, S. Kurtulus, D. Gennert, J. Xia, J. Y. H. Kwon, J. Nevin, R. H. Herbst, I. Yanai, O. Rozenblatt-Rosen, V. K. Kuchroo, A. Regev, A. C. Anderson, A distinct gene module for

18. C. C. Ku, M. Murakami, A. Sakamoto, J. Kappler, P. Marrack, Control of homeostasis of CD8+ memory T cells by opposing cytokines. *Science* **288**, 675–678 (2000).

19. S. Choi, C. Warzecha, E. Zvezdova, J. Lee, J. Argenty, R. Lesourne, L. Aravind, P. E. Love, THEMIS enhances TCR signaling and enables positive selection by selective inhibition of the phosphatase SHP-1. *Nat. Immunol.* **18**, 433–441 (2017).

20. S. Gilfillan, C. J. Chan, M. Cella, N. M. Haynes, A. S. Rapaport, K. S. Boles, D. M. Andrews, M. J. Smyth, M. Colonna, DNAM-1 promotes activation of cytotoxic lymphocytes by nonprofessional antigen-presenting cells and tumors. *J. Exp. Med.* **205**, 2965–2973 (2008).

21. D. Peter, S. L. C. Jin, M. Conti, A. Hatzelmann, C. Zitt, Differential expression and function of phosphodiesterase 4 (PDE4) subtypes in human primary CD4+ T cells: Predominant role of PDE4D. *J. Immunol.* **178**, 4820–4831 (2007).

22. M. Villalba, K. Bi, F. Rodriguez, Y. Tanaka, S. Schoenberger, A. Altman, Vav1/Rac-dependent actin cytoskeleton reorganization is required for lipid raft clustering in T cells. *J. Cell Biol.* **155**, 331–338 (2001).

23. C. Y. Yang, J. A. Best, J. Knell, E. Yang, A. D. Sheridan, A. K. Jesionek, H. S. Li, R. R. Rivera, K. C. Lind, L. M. D'Cruz, S. S. Watowich, C. Murre, A. W. Goldrath, The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8+ T cell subsets. *Nat. Immunol.* **12**, 1221–1229 (2011).

24. A. Schietinger, M. Philip, V. E. Krisnawan, E. Y. Chiu, J. J. Delrow, R. S. Basom, P. Lauer, D. G. Brockstedt, S. E. Knoblaugh, G. J. Hämmerling, T. D. Schell, N. Garbi, P. D. Greenberg, Tumor-specific T cell dysfunction is a dynamic antigen-driven differentiation program initiated early during tumorigenesis. *Immunity* **45**, 389–401 (2016).

25. E.-A. D. Amir, K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan, D. Pe'er, viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).

26. P. Qiu, E. F. Simonds, S. C. Bendall, K. D. Gibbs Jr., R. V. Bruggner, M. D. Linderman, K. Sachs, G. P. Nolan, S. K. Plevritis, Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* **29**, 886–891 (2011).

27. M. Yadav, S. Jhunjhunwala, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk, Z. Modrusan, I. Mellman, J. R. Lill, L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).

28. B. Homet Moreno, J. M. Zaretsky, A. Garcia-Diaz, J. Tsoi, G. Parisi, L. Robert, K. Meeth, A. Ndoye, M. W. Bosenberg, A. T. Weeraratna, T. G. Graeber, B. Comin-Anduix, S. Hu-Lieskovan, A. Ribas, Response to programmed cell death-1 blockade in a murine melanoma syngeneic model requires costimulation, CD4, and CD8 T cells. *Cancer Immunol. Res.* **4**, 845–857 (2016).

29. V. R. Buchholz, T. N. M. Schumacher, D. H. Busch, T cell fate at the single-cell level. *Annu. Rev. Immunol.* **34**, 65–92 (2016).

30. R. J. Johnston, L. Comps-Agrar, J. Hackney, X. Yu, M. Huseni, Y. Yang, S. Park, V. Javinal, H. Chiu, B. Irving, D. L. Eaton, J. L. Grogan, The immunoreceptor TIGIT regulates antitumor and antiviral CD8+ T cell effector function. *Cancer Cell* **26**, 923–937 (2014).

31. K. E. Pauken, E. J. Wherry, TIGIT and CD226: Tipping the balance between costimulatory and coinhibitory molecules to augment the cancer immunotherapy toolkit. *Cancer Cell* **26**, 785–787 (2014).

32. E. Hui, J. Cheung, J. Zhu, X. Su, M. J. Taylor, H. A. Wallweber, D. K. Sasmal, J. Huang, J. M. Kim, I. Mellman, R. D. Vale, T cell costimulatory receptor CD28 is a primary target for PD-1-mediated inhibition. *Science* **355**, 1428–1433 (2017).

33. M. F. Bachmann, K. McKall-Faienza, R. Schmits, D. Bouchard, J. Beach, D. E. Speiser, T. W. Mak, P. S. Ohashi, Distinct roles for LFA-1 and CD28 during activation of naive T cells: Adhesion versus costimulation. *Immunity* **7**, 549–557 (1997).

34. K. Shibuya, J. Shirakawa, T. Kameyama, S.-I. Honda, S. Tahara-Hanaoka, A. Miyamoto, M. Onodera, T. Sumida, H. Nakauchi, H. Miyoshi, A. Shibuya, CD226 (DNAM-1) is involved in lymphocyte function-associated antigen 1 costimulatory signal for naive T cell differentiation and proliferation. *J. Exp. Med.* **198**, 1829–1839 (2003).

35. N. Joller, J. P. Hafler, B. Brynedal, N. Kassam, S. Spoerl, S. D. Levin, A. H. Sharpe, V. K. Kuchroo, Cutting edge: TIGIT has T cell-intrinsic inhibitory functions. *J. Immunol.* **186**, 1338–1342 (2011).

36. N. Joller, E. Lozano, P. R. Burkett, B. Patel, S. Xiao, C. Zhu, J. Xia, T. G. Tan, E. Sefik, V. Yajnik, A. H. Sharpe, F. J. Quintana, D. Mathis, C. Benoist, D. A. Hafler, V. K. Kuchroo, Treg cells expressing the coinhibitory molecule TIGIT selectively inhibit proinflammatory Th1 and Th17 cell responses. *Immunity* **40**, 569–581 (2014).

37. A. C. Anderson, N. Joller, V. K. Kuchroo, Lag-3, Tim-3, and TIGIT: Co-inhibitory receptors with specialized functions in immune regulation. *Immunity* **44**, 989–1004 (2016).

38. M. M. Gubin, X. Zhang, H. Schuster, E. Caron, J. P. Ward, T. Noguchi, Y. Ivanova, J. Hundal, C. D. Arthur, W.-J. Krebber, G. E. Mulder, M. Toebes, M. D. Vesely, S. S. K. Lam, A. J. Korman, J. P. Allison, G. J. Freeman, A. H. Sharpe, E. L. Pearce, T. N. Schumacher, R. Aebersold,

dysfunction uncoupled from activation in tumor-infiltrating T cells. *Cell* **166**, 1500–1511.e9 (2016).

H.-G. Rammensee, C. J. M. Melief, E. R. Mardis, W. E. Gillanders, M. N. Artyomov, R. D. Schreiber, Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).

39. B. Bengsch, A. L. Johnson, M. Kurachi, P. M. Odorizzi, K. E. Pauken, J. Attanasio, E. Stelekati, L. M. McLane, M. A. Paley, G. M. Delgoffe, E. J. Wherry, Bioenergetic insufficiencies due to metabolic alterations regulated by the inhibitory receptor PD-1 are an early driver of CD8[+] T cell exhaustion. *Immunity* **45**, 358–373 (2016).

40. R. J. Johnston, X. Yu, J. L. Grogan, The checkpoint inhibitor TIGIT limits antitumor and antiviral CD8[+] T cell responses. *Oncoimmunology* **4**, e1036214 (2015).

41. A. O. Kamphorst, A. Wieland, T. Nasti, S. Yang, R. Zhang, D. L. Barber, B. T. Konieczny, C. Z. Daugherty, L. Koenig, K. Yu, G. L. Sica, A. H. Sharpe, G. J. Freeman, B. R. Blazar, L. A. Turka, T. K. Owonikoko, R. Pillai, S. S. Ramalingam, K. Araki, R. Ahmed, Rescue of exhausted CD8 T cells by PD-1–targeted therapies is CD28-dependent. *Science*, eaaf0683 (2017).

42. E. Lozano, M. Dominguez-Villar, V. Kuchroo, D. A. Hafler, The TIGIT/CD226 axis regulates human T cell function. *J. Immunol.* **188**, 3869–3875 (2012).

43. D. M. Valenzuela, A. J. Murphy, D. Frendewey, N. W. Gale, A. N. Economides, W. Auerbach, W. T. Poueymirou, N. C. Adams, J. Rojas, J. Yasenchak, R. Chernomorsky, M. Boucher, A. L. Elsasser, L. Esau, J. Zheng, J. A. Griffiths, X. Wang, H. Su, Y. Xue, M. G. Dominguez, I. Noguera, R. Torres, L. E. Macdonald, A. F. Stewart, T. M. DeChiara, G. D. Yancopoulos, High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat. Biotechnol.* **21**, 652–659 (2003).

44. W. T. Poueymirou, W. Auerbach, D. Frendewey, J. F. Hickey, J. M. Escaravage, L. Esau, A. T. Doré, S. Stevens, N. C. Adams, M. G. Dominguez, N. W. Gale, G. D. Yancopoulos, T. M. DeChiara, D. M. Valenzuela, F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat. Biotechnol.* **25**, 91–99 (2007).

45. F. Pagès, M. Ragueneau, R. Rottapel, A. Truneh, J. Nunes, J. Imbert, D. Olive, Binding of phosphatidyl-inositol-3-OH kinase to CD28 is required for T-cell signalling. *Nature* **369**, 327–329 (1994).

46. H. Wang, T. A. Kadlecek, B. B. Au-Yeung, H. E. S. Goodfellow, L.-Y. Hsu, T. S. Freedman, A. Weiss, ZAP-70: An essential kinase in T-cell signaling. *Cold Spring Harb. Perspect. Biol.* **2**, a002279 (2010).

47. W. Zhang, J. Sloan-Lancaster, J. Kitchen, R. P. Trible, L. E. Samelson, LAT: The ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell* **92**, 83–92 (1998).

48. R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani, G. P. Nolan, Automated identification of stratifying signatures in cellular subpopulations. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2770–E2777 (2014).

49. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).

50. X. Yang, D. Liu, F. Liu, J. Wu, J. Zou, X. Xiao, F. Zhao, B. Zhu, HTQC: A fast quality control toolkit for Illumina sequencing data. *BMC Bioinform.* **14**, 33 (2013).

51. X. Yang, D. Liu, N. Lv, F. Zhao, F. Liu, J. Zou, Y. Chen, X. Xiao, J. Wu, P. Liu, J. Gao, Y. Hu, Y. Shi, J. Liu, R. Zhang, C. Chen, J. Ma, G. F. Gao, B. Zhu, TCRklass: A new K-string-based algorithm for human and mouse TCR repertoire characterization. *J. Immunol.* **194**, 446–454 (2014).

52. A. Han, J. Glanville, L. Hansmann, M. M. Davis, Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).

Citation: B. Wang, W. Zhang, V. Jankovic, J. Golubov, P. Poon, E. M. Oswald, C. Gurer, J. Wei, I. Ramos, Q. Wu, J. Waite, M. Ni, C. Adler, Y. Wei, L. Macdonald, T. Rowlands, S. Brydges, J. Siao, W. Poueymirou, D. MacDonald, G. D. Yancopoulos, M. A. Sleeman, A. J. Murphy, D. Skokos, Combination cancer immunotherapy targeting PD-1 and GITR can rescue CD8[+] T cell dysfunction and maintain memory phenotype. *Sci. Immunol.* **3**, eaat7061 (2018).

TUMOR IMMUNOLOGY

A melanin-mediated cancer immunotherapy patch

Yanqi Ye,[1,2]* Chao Wang,[1,2]* Xudong Zhang,[1,2] Quanyin Hu,[1,2] Yuqi Zhang,[1,2] Qi Liu,[1,2] Di Wen,[1,2] Joshua Milligan,[1] Adriano Bellotti,[1,3] Leaf Huang,[1,2] Gianpietro Dotti,[4] Zhen Gu[1,2,3]†

Melanin is capable of transforming 99.9% of the absorbed sunlight energy into heat, reducing the risk of skin cancer. We here develop a melanin-mediated cancer immunotherapy strategy through a transdermal microneedle patch. B16F10 whole tumor lysate containing melanin is loaded into polymeric microneedles that allow sustained release of the lysate upon insertion into the skin. In combination with the near-infrared light irradiation, melanin in the patch mediates the generation of heat, which further promotes tumor-antigen uptake by dendritic cells, and leads to enhanced antitumor vaccination. We found that the spatiotemporal photoresponsive immunotherapy increases infiltration of polarized T cells and local cytokine release. These immunological effects increase the survival of mice after tumor challenge and elicited antitumor effects toward established primary tumor and distant tumor. Collectively, melanin generates local heat, boosts T cell activities by transdermal vaccines, and promotes antitumor immune responses.

INTRODUCTION

Emerging technologies associated with immunotherapy hold tremendous promise in cancer therapy (*1*). Micro- or nanoformulations or engineered immune cells can be used to deliver a variety of immunomodulators (*2*, *3*). Scaffolds, such as hydrogel, have also been developed to generate an immunogenic microenvironment that recruits and activates immune cells in situ (*4–7*). Moreover, T cell engineering that incorporates antibodies or therapeutics facilitates immune targeting and treatment (*8–11*). Also, dendritic cell (DC)–based vaccination can effectively capture antigens to improve the effectiveness of the immune response and is a powerful tool for cancer therapy (*12–16*). However, engineering of DCs often involves complex and expensive ex vivo manipulation (*4*, *17*). In addition, the limited lymph node–homing capability of ex vivo manipulated DCs is, at least in part, responsible for the limited anticancer efficacy (*18*). Vaccination with whole tumor antigens provides a broad source of tumor-associated antigens that elicit substantially enhanced immune responses compared with narrowly defined tumor antigens (*19*). Moreover, presenting a broad spectrum of immunogenic epitopes not only augments the immunity mediated by the DC antigen uptake and processing but also improves the direct activation of $CD4^+$ T helper and $CD8^+$ cytotoxic T lymphocytes (*7*, *20*).

Here, we describe a B16F10 melanoma vaccine patch that targets antigen-presenting cells (APCs) directly via transdermal delivery of tumor lysates combined with melanin (Fig. 1A). This vaccine involves the encapsulation of inactive whole tumor lysate that is gradually released by an intradermal microneedle (MN) patch inserted into the skin. MNs facilitate the uptake and presentation of antigens by DCs and, in turn, promote immune activation through the extensive network of lymphatic vessels in the dermis (*21–23*). At the same time, the presence of melanin, the existing natural biological

pigment in the whole tumor lysate, allows the local release of heat via remotely controllable near-infrared (NIR) light emission. Local heat causes the release of inflammatory cytokines that attract immune cells, generation of immunogenic substrates such as extracellular heat shock proteins (HSPs), reactive oxygen species (ROS), antigen adjuvants, and some other danger signals that activate the immune system (*24*, *25*). The mild increase in the local temperature of the interstitial tissues also contributes to the increased blood and lymphatic flow that facilitates the migration of APCs and T cells (*24*) and finally initiates B16F10-specific immune responses (*23*, *26*). The increased blood flow may also allow the recruitment of other cell subsets such as natural killer (NK) cells (*7*, *27*, *28*). We found that the administration of the vaccine MN patch generated robust innate and adaptive immune responses and induced tumor regression in the B16F10 melanoma model. Moreover, the NIR-enhanced transdermal vaccination delayed the growth of distant tumor and improved long-term survival, paving a strong rationale for pursuing this strategy in clinical studies.

RESULTS

Preparation and characterization of the cancer immunotherapy patch

We first investigated whether the tumor lysate can be loaded and released from transdermal MN patches in a sustained manner. We fabricated the MN patch within a micromold to form hyaluronic acid–based MNs that encapsulate the whole tumor lysate (with melanin) and adjuvants, such as granulocyte-macrophage colony-stimulating factor (GM-CSF) (*29*, *30*). Patches with and without tumor lysate were illustrated from the axial and transverse perspectives, respectively (Fig. 1B). An array of 15×15 MNs was assembled on a 9 mm by 9 mm patch with a center-to-center interval of 600 μm. The detailed dimensions of the MNs were visualized by the scanning electron microscopy (Fig. 1C). Each MN had a conical construction with a diameter of 300 μm at the base, a height of 800 μm, and a sharp tip tapering to a 5-μm radius of curvature. Upon loading with the tumor lysate, MNs appeared notably darker than the blank hyaluronic acid–based MNs because of the presence of melanin in the patch (*31*, *32*). The amount of melanin was around 50 μg per patch, which is within the safe dosage range of a single administration (*33*). A

[1]Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and North Carolina State University, Raleigh, NC 27695, USA. [2]Division of Pharmacoengineering and Molecular Pharmaceutics, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [3]Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA. [4]Department of Microbiology and Immunology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: zgu@email.unc.edu
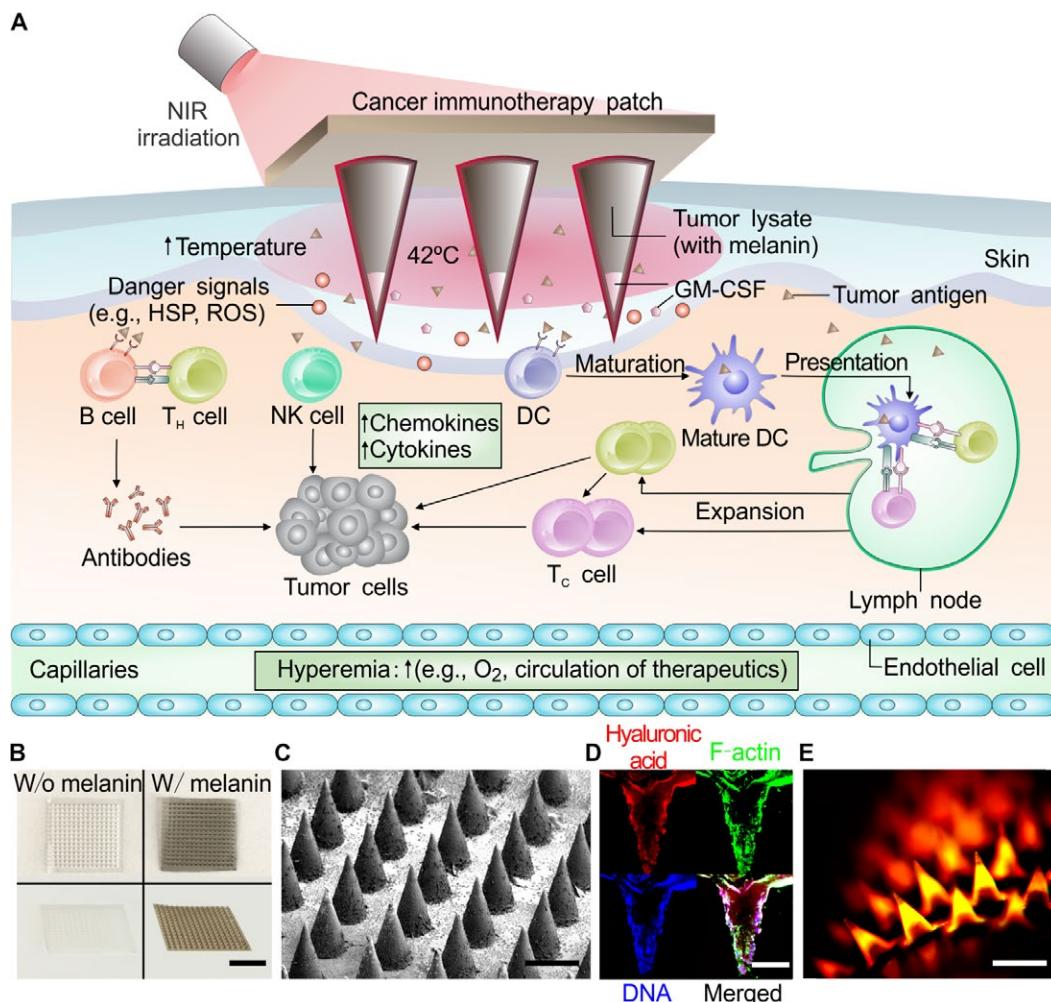
Fig. 1. Schematic of melanin-mediated cancer immunotherapy through a transdermal MN-based vaccine patch. (**A**) Schematic illustration of MN-based transdermal vaccination. T_H cell, T helper cell; T_c cell, cytotoxic T cell. (**B**) Photograph of representative MN patches without (W/o) melanin and with (W/) melanin (scale bar, 4 mm). (**C**) Scanning electron microscopy image of the MN patch (scale bar, 400 μm). (**D**) Fluorescence cross-sectional images of a representative MN. Actin filaments in cells were visualized by Alexa Fluor 488 phalloidin (green), cell DNA fragments were stained with Hoechst (blue), and hyaluronic acid polymer matrix was labeled with rhodamine B (red) (scale bar, 200 μm). (**E**) Fluorescence imaging of a representative MN patch that contained the Alexa Fluor 488 phalloidin–labeled tumor lysate and rhodamine B–labeled hyaluronic acid (scale bar, 400 μm).

representative MN was constructed with rhodamine B–labeled hyaluronic acid encapsulating tumor lysate with actin filaments and DNA stained with phalloidin and Hoechst, respectively (Fig. 1D). A fluorescent view of the array further showed the uniform loading of tumor lysate and alignment of the MNs (Fig. 1E). The MN underwent failure at 0.39 N in strain testing while being compressed to a quarter of its height (fig. S1), demonstrating sufficient strength for potential skin insertion without buckling (*34*).

Next, we evaluated the effects of the NIR laser irradiation on melanin-induced heat generation. The temperature variations of the MNs were recorded in real time using an infrared thermal camera (Fig. 2A). Compared with the transparent hyaluronic acid–based MNs, the surface temperature of the MNs loaded with tumor lysate increased dramatically within 1 min because of the melanin-induced light-to-heat transduction (Fig. 2B). NIR irradiation at 808 nm could be absorbed by the melanin contained in the tumor lysate as demonstrated by the melanin's absorption spectrum (fig. S2A). The absorption coefficient of the melanin content in the tumor lysate was also comparable with that of the synthetic melanin (fig. S2A and table S1). Despite the temperature increase, the morphology of the MNs remained unchanged, and the photothermal property of melanin was stable during the treatment (fig. S2B) (*35*). The heating behavior of the MN patch was also maintained when repeated NIR light exposures were performed (fig. S3). Furthermore, the steady state of the patch's surface temperature was lysate concentration– and NIR light intensity–dependent, whereas the thickness of the MN backing had a minimum effect (fig. S4 and Fig. 2C). The surface temperature of the patch could be controlled under hyperthermia (42°C) to minimize the potential thermal-induced denaturation of the tumor antigen and other biomolecules.

In vitro activation of DCs in response to vaccine MN
GM-CSF that serves as a potent cytokine for DC recruitment and activation was physically encapsulated into MN tips by cross-linking
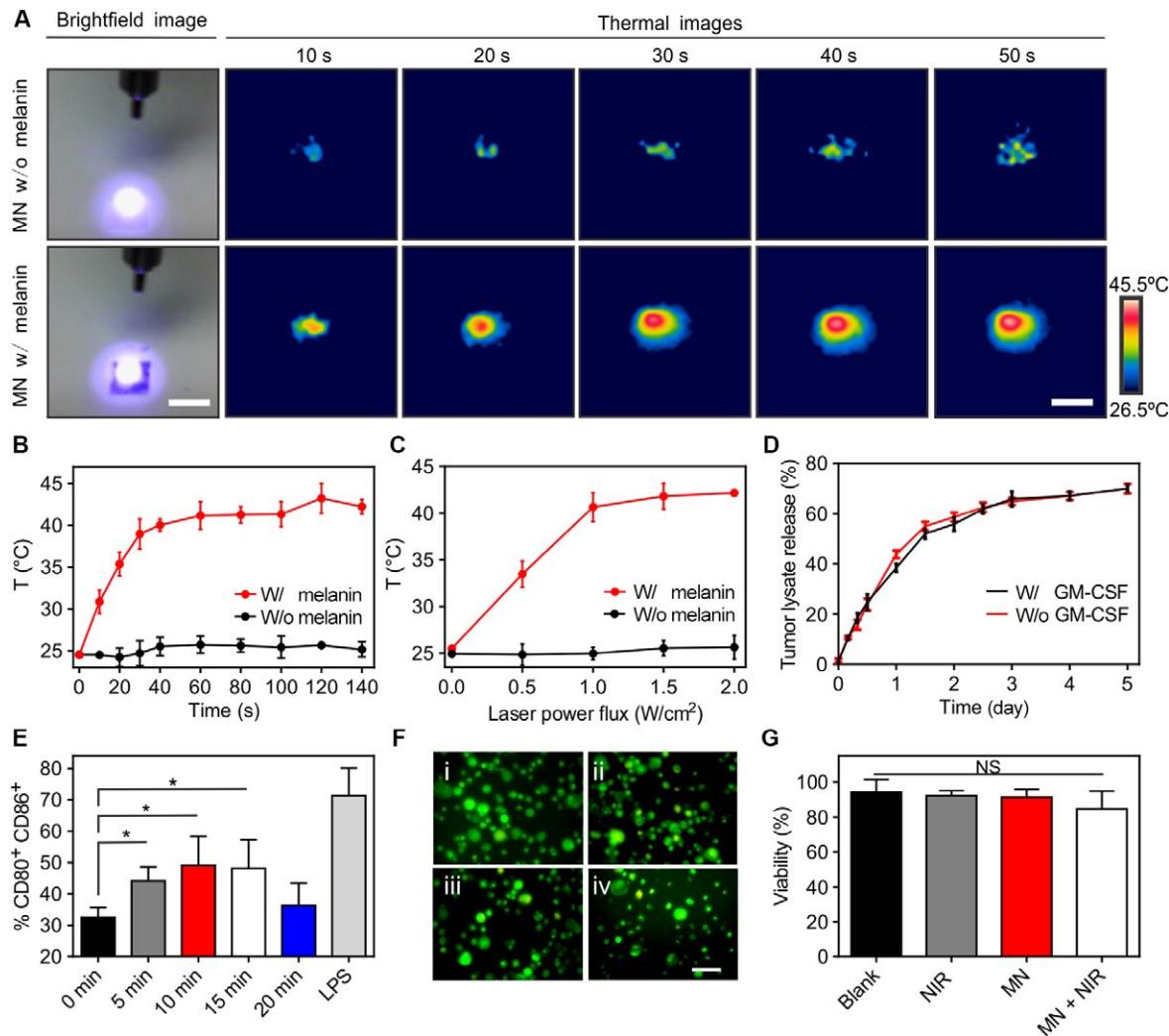
Fig. 2. Characterization of the light-responsive transdermal MNs. (**A**) Surface temperature changes of the MNs with or without tumor lysate in real time with continuous 808-nm NIR irradiation at 1.0 W/cm², characterized by an infrared thermal camera (scale bars, 1 mm). (**B**) Quantitative surface temperature changes of representative MNs with continuous NIR irradiation at 1.0 W/cm² ($n = 3$). (**C**) Quantitative temperature changes of representative MN patch with increasing laser power flux ($n = 3$). (**D**) In vitro collective release of tumor lysate proteins from the MN patch ($n = 3$). (**E**) In vitro activation of DCs in response to MNs loaded with tumor lysate and GM-CSF or LPS and exposed to NIR irradiation for different time ($n = 3$). (**F**) LIVE/DEAD assay of DCs after treatments with (i) blank MN, (ii) 10 min of NIR, (iii) MN, and (iv) MN and 10 min of NIR. Live cell (green), dead cell (red). (**G**) Quantification of DC viability after treatments. Data points represent mean ± SD ($n = 3$). Error bars indicate SD. Statistical significance was calculated by Student's t test [not significant (NS), $P > 0.05$; *$P < 0.05$].

methacrylated hyaluronic acid upon ultraviolet irradiation. We found that 60% of the bioactive GM-CSF was released from MN within 48 hours (fig. S5, A and B), whereas sustained release of the tumor lysate protein was observed over 5 days (Fig. 2D). Incorporation of the NIR treatment did not alter the release profiles of GM-CSF and tumor lysate (fig. S5, C and D). Scanning electron microscopy images of vaccine MNs showed a gradual dissociation of tips over time (fig. S6).

To evaluate whether GM-CSF in MNs provided signaling cues that could efficiently promote DC maturation, we exposed bone marrow–derived DCs to the dissolved MN suspension. The percentage of matured DCs (CD80⁺ and CD86⁺) substantially increased from 36.7 ± 2.3% to 48.9 ± 3.1% after treatment with tumor lysate– and GM-CSF–loaded MNs with 10 min of NIR laser irradiation (Fig. 2E) (*36*). The effect of varying NIR laser irradiation time on the activation

of DCs was also measured. Ten minutes of NIR irradiation allowed optimal DC activation as compared with samples treated with 5 or 15 min (Fig. 2E). Only 20 min of NIR exposure slightly impaired DC viability and functionality (fig. S7). DCs in all other experimental conditions of either MN suspension or 10 min of NIR irradiation exhibited high viability (Fig. 2, F and G, and fig. S8).

In vivo efficacy of MN-mediated immunization upon NIR

To characterize the in vivo efficacy of MN-based immunization, we transdermally treated C57BL/6J mice with vaccine MN patch loaded with B16F10 whole tumor lysate containing 1.5 mg of extracted protein. We measured the magnitude and duration of the immune response after NIR irradiation and subsequent tumor challenge (Fig. 3A). MN patches were applied on the mice skin of the caudal-dorsal area for about 10 min and further affixed using the Skin Affix surgical
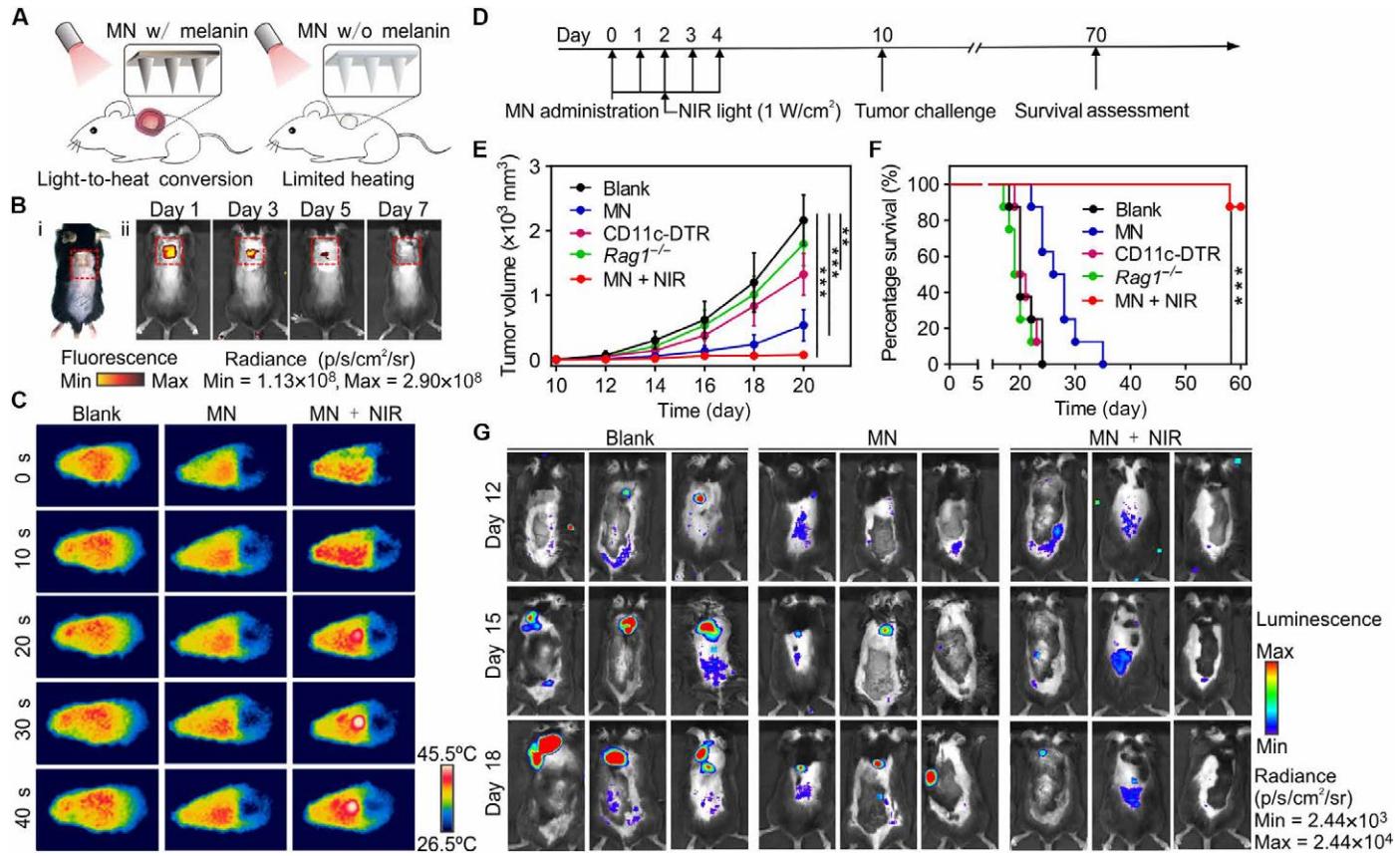
Fig. 3. Vaccine MN confer protective innate and adaptive immunity. (**A**) Schematic illustration of MN cancer immunotherapy. (**B**) Characterization of the MNs after insertion. (i) Photo of mouse dorsal skin (the area within the red line) that was treated transdermally with one MN patch and (ii) fluorescence signals of Cy5.5-labeled MN patch over time. (**C**) Surface temperature changes of individual animal after MN insertion into the skin measured by an infrared thermal camera. (**D**) Schematic representation of the B16F10 vaccine tumor model. (**E**) Average tumor volumes in treated mice after tumor challenge. (**F**) Kaplan-Meier survival curves for treated and control mice. Data points represent mean ± SD ($n = 8$). Error bars indicate SD. Statistical significance was calculated by Student's test and log-rank test (**$P < 0.01$; ***$P < 0.001$). (**G**) In vivo bioluminescence imaging of the B16F10 melanoma in different experimental groups and at different time points after tumor challenge. Three representative mice per treatment group are shown. The tumor growth in (E) to (G) was measured 10 days after tumor cell inoculation.

adhesive. Staining with trypan blue and hematoxylin and eosin (H&E) indicated successful penetration of MNs in the excised skin (fig. S9). The transdermal patch remained in the skin for at least 5 days (Fig. 3B). We then performed localized NIR irradiation on the MN region for 10 min daily for 5 days (MN + NIR). Control mice were treated with either vaccine MN patch without NIR irradiation (MN), MN patch loaded only with melanin (melanin), or MNs containing only hyaluronic acid with NIR irradiation (blank).

Temperature changes in the regional skin surface after vaccine MN insertion were recorded in real time using an infrared thermal camera. Light-to-heat transduction upon NIR irradiation caused a local heating effect observed in mice treated with vaccine MNs (Fig. 3C). The melanin in the vaccine patch mediated the transdermal heating at temperatures between 38° and 42°C within 30 s. Mild hyperthermia at the local treated site was similarly observed in mice treated with synthetic melanin–loaded MNs (fig. S10A). In contrast, mice treated with blank MNs and NIR and mice implanted with loaded MNs but without NIR showed limited variations in skin surface temperature within the normal range of 33° to 36°C (Fig. 3C).

In a prophylactic mouse model, mice were implanted with B16F10 melanoma cells 10 days after vaccination (Fig. 3D). All mice treated with the blank MNs had appreciable tumor growth within 15 days

after tumor cell inoculation and required euthanasia by day 25. MNs loaded with melanin and treated with NIR irradiation slightly improved the survival of the mice because some mice survived to day 25 (fig. S10, B and C). Similarly, MNs loaded with tumor lysate and melanin but without NIR irradiation caused tumor protection in 13% of the mice until day 30 (Fig. 3, E and F). In sharp contrast, mice receiving the combined vaccination (MNs loaded with tumor lysate and GM-CSF and NIR irradiation) showed long-term survival with complete tumor rejection in 87% of the treated mice (Fig. 3, E and F, and fig. S11). Bioluminescence imaging of the B16F10 melanoma–bearing mice confirmed significant inhibition of tumor growth (Fig. 3G and fig. S12). This was further evidenced by the measurement of tumor weight (fig. S13) and histologic analysis (fig. S14A).

We next assessed the requirement of immune cells for the antitumor effects observed by combined vaccination–based treatments. Depletion of CD11c+ DCs in diphtheria toxin receptor (DTR) mice was sufficient to abrogate the antitumor effect of the vaccine MN (Fig. 3, E and F). A study in $Rag1^{-/-}$ mice deficient in T cells and B cells showed a significant loss of tumor growth suppression during treatment with combined vaccination (fig. S11). Selective depletions of CD8+ and CD4+ T cells before combined MN vaccination were also studied. Eliminating CD8+ T lymphocytes showed no significant

tumor regression compared with blank control (fig. S15). When anti-CD4 antibody was given to the mice, there was a decrease ($P < 0.01$) in tumor size in contrast to the control, suggesting the benefit of CD4 T cells to a lesser extent than CD8 T cells for the antitumor response (fig. S15). The depletion of B cells and NK cells also had deleterious impact on the immune response while not diminishing the vaccination effect toward tumor challenge (fig. S15). Together, these results showed that the MN vaccination was associated with CD11c⁺ DCs and other immune cells such as T cells, B cells, and NK cells.

Loading of GM-CSF in the MNs played an important role in the local recruitment of DCs [CD11c⁺, paired immunoglobulin (Ig)–like receptors (PIRs) of activating-A/B⁺] (Fig. 4, A and C). Three days after combined vaccination, a 5.9-fold increase in accumulated DCs was observed in the skin section compared with mice treated with blank MN. NIR further augmented the effect of GM-CSF–loaded MNs on recruiting DCs (Fig. 4C). Increased localization of NK cells was also observed (Fig. 4, B and D). In addition, the elevated local microcirculatory blood perfusion observed after NIR and MN treatment could contribute in enhancing the migration of immune cells

(table S2 and fig. S16). Tumor infiltration by T cells upon treatment was analyzed by flow cytometry on day 15 after tumor inoculation. About 9.8-fold increase in CD8⁺ T cells was observed in mice receiving the combined vaccination compared with control mice, whereas MN-only group showed a 5.8-fold increase (Fig. 5A). Furthermore, staining with H-2Dᵇ gp100 tetramers identified B16F10-specific CD8⁺ T cells in tumors of treated mice (37). The percentage of tetramer-positive CD8⁺ T cells was found to be greater in the MN-treated mice compared with control mice lacking immunization (fig. S17). Mice with combined treatment or MN-only treatment exhibited 1.5- and 1.3-fold increases in activated DCs (CD80⁺ and CD86⁺), respectively, in regional skin as compared with control mice (Fig. 5B). Immunofluorescence staining and in situ cell apoptosis confirmed the results obtained with flow cytometry (Fig. 5C and fig. S14B). Local immune activation was associated with systemic immune responses. We measured an eightfold increase in IgG titers in the serum of immunized mice as compared with mice treated with blank control (Fig. 5D). NIR treatment promoted further increase in IgG titers by day 15 and prolonged immune responses compared with
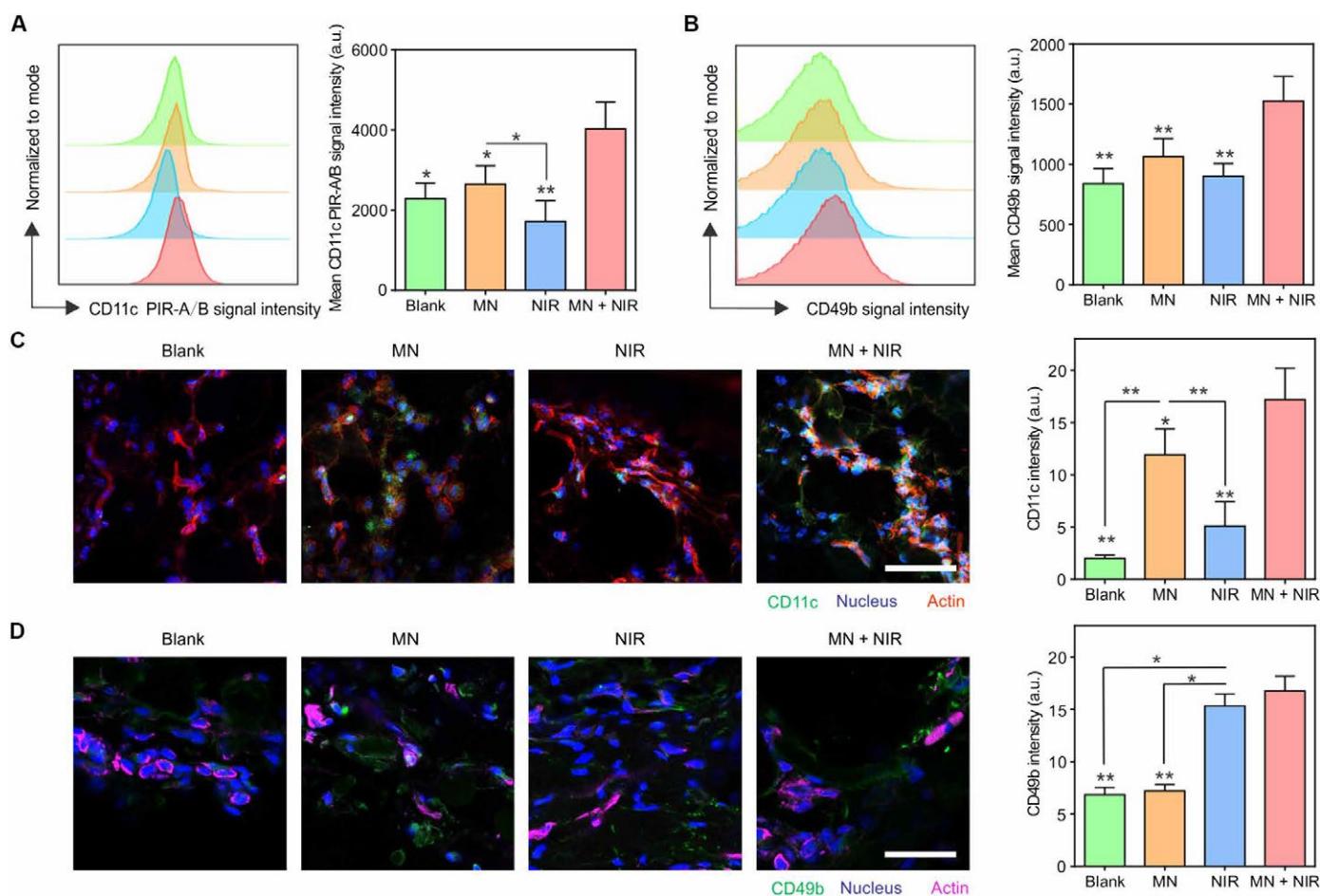

Fig. 4. Immune cell recruitment after the NIR-boosted and MN-mediated cancer immunotherapy. (**A**) Representative quantitative analysis of DCs (CD11c⁺ and PIR-A/B⁺) infiltrated in the skin 3 days after treatments as assessed by flow cytometry. The indicated samples were treated with blank MN (blank), vaccine MN (MN), MN loaded with tumor lysate without GM-CSF and treated with NIR (NIR), and vaccine MN and treated with NIR (MN + NIR). a.u., arbitrary units. (**B**) Representative quantitative analysis of NK cells (CD49b⁺) in the skin upon transdermal cancer immunotherapy as assessed by flow cytometry. (**C** and **D**) Immunofluorescence staining and quantitative analysis of (C) CD11c⁺ DCs (scale bar, 100 μm) and (D) CD49b⁺ NK cells (scale bar, 100 μm). Statistical significance was calculated by Student's *t* test (*$P < 0.05$; **$P < 0.01$). Asterisks indicate significant differences between the MN + NIR group and all other treatment groups. Data points represent mean ± SD ($n = 3$). Error bars indicate SD.
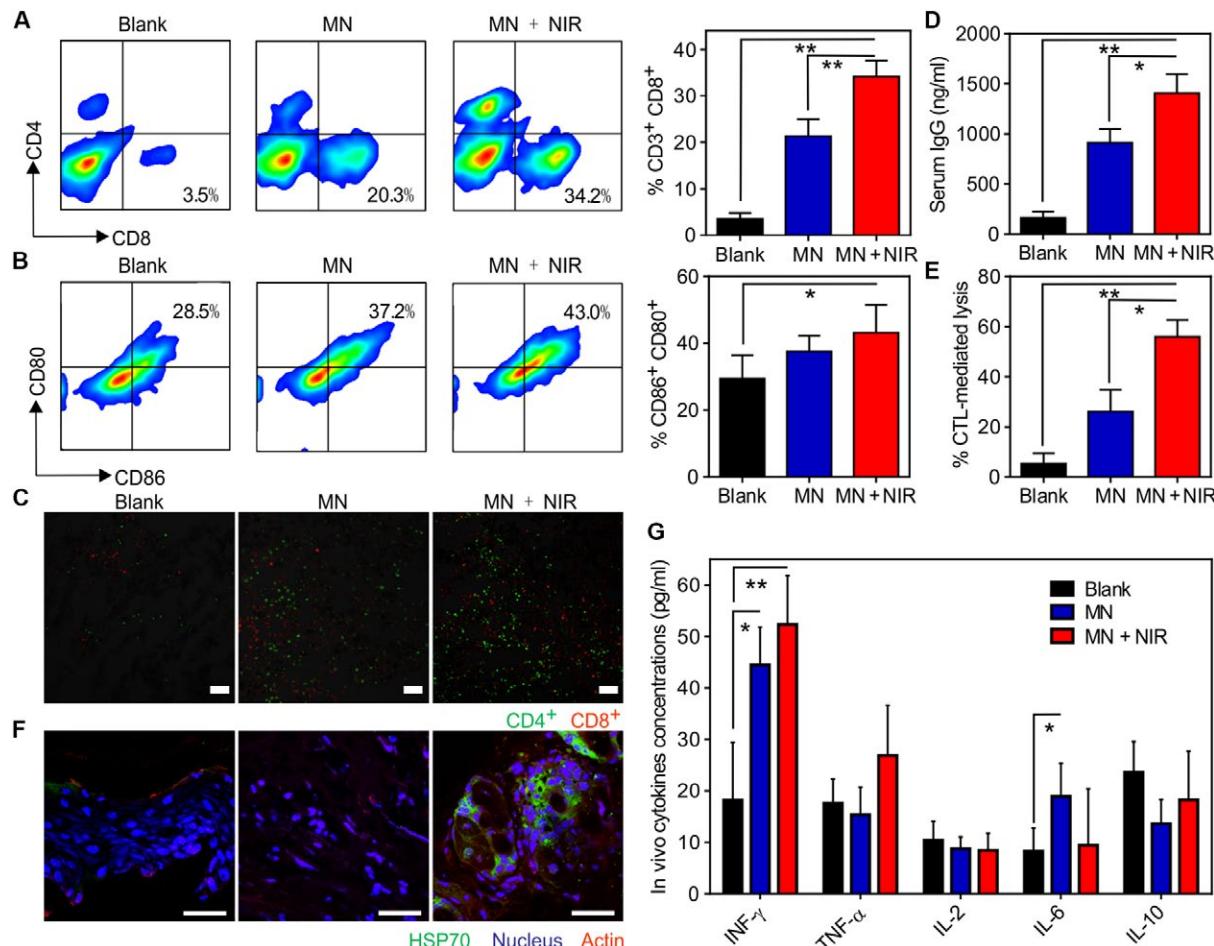
Fig. 5. Immunologic responses after the MN-mediated cancer immunotherapy. (**A**) Representative quantitative analysis of T cells (gated on CD3$^+$ T cells) in treated tumors analyzed by flow cytometry. (**B**) Representative quantitative analysis of activated DCs (CD86$^+$ and CD80$^+$) in the draining lymph nodes analyzed by flow cytometry. Data points represent mean ± SD (n = 8). (**C**) Immunofluorescence staining of the tumors showing CD4$^+$ T cell and CD8$^+$ T cell infiltration (scale bars, 100 μm). (**D**) Quantification of IgG1 subtypes in serum collected at day 10. Data points represent mean ± SD (n = 8). (**E**) Cytotoxic responses of splenocytes against B16F10 cells in vitro. Data points represent mean ± SD (n = 6). (**F**) Immunofluorescence staining of HSP70 (green) in the regional skin with actin filaments visualized by Alexa Fluor 660 phalloidin (red) and cell nuclei stained with Hoechst (blue) (scale bars, 100 μm). (**G**) In vivo local detection of cytokines from extracted patches at day 3. Statistical significance was calculated by Student's t test (*P < 0.05; **P < 0.01). Data points represent mean ± SD (n = 8). Error bars indicate SD. CTL, cytotoxic T lypmphocyte.

control groups (fig. S18). The in vitro analysis of splenocytes revealed 10-fold higher frequency of T cells responding to B16F10 tumor lysate in mice receiving the combined vaccination (Fig. 5E).

To examine that the light-to-heat transduction induced the production of danger signals and proinflammatory cytokines, we first measured local levels of ROS in the MN-treated surrounding tissue by flow cytometry. Samples from mice receiving the combined treatment showed about fourfold increase in ROS levels compared with the untreated group, whereas 1.5-fold increase was observed compared with the MN control without the NIR irradiation (fig. S19). In line with the elicitation of danger signals, combined vaccination caused the expression of HSP70 and HSP90 (Fig. 5F and fig. S20) (*38*). Danger signals from the local tissue and the antigenic molecules promote proinflammatory cytokine production (*39*). Consistent with this effect, we found local increases of interferon-γ (IFN-γ), tumor necrosis factor–α (TNF-α), and interleukin-6 (IL-6) in mice treated with the combined vaccination compared with MN controls (Fig. 5G and fig. S21).

Efficacy of MN patch in distant tumors

We further analyzed whether the local vaccination confers protection toward secondary tumors distant from the NIR- and MN-treated site (Fig. 6A). Local NIR irradiation and MN treatment were performed only on the left-side tumor in B16F10 mice bearing bilateral tumors. Right-side tumor was not injected and was shielded from light (Fig. 6B). We observed that the tumor sizes and bioluminescence signals decreased significantly on both sides of the mice that had the combined vaccination (Fig. 6, C to G). Meanwhile, substantial increase of activated DCs in the regional lymph node (Fig. 5B) and enhanced cytotoxic responses to B16F10 cells of the splenocytes in vitro (Fig. 5E) indicated that systemic antitumor effect could be achieved by the transdermal immunotherapy. This effect was parallel by fivefold increases in CD8$^+$ T cell infiltration compared with control, which was consistent with the role of immune cells for antitumor efficacy (Fig. 6D). Distant metastases were not observed in the lungs of the mice with combinational treatment (fig. S22).
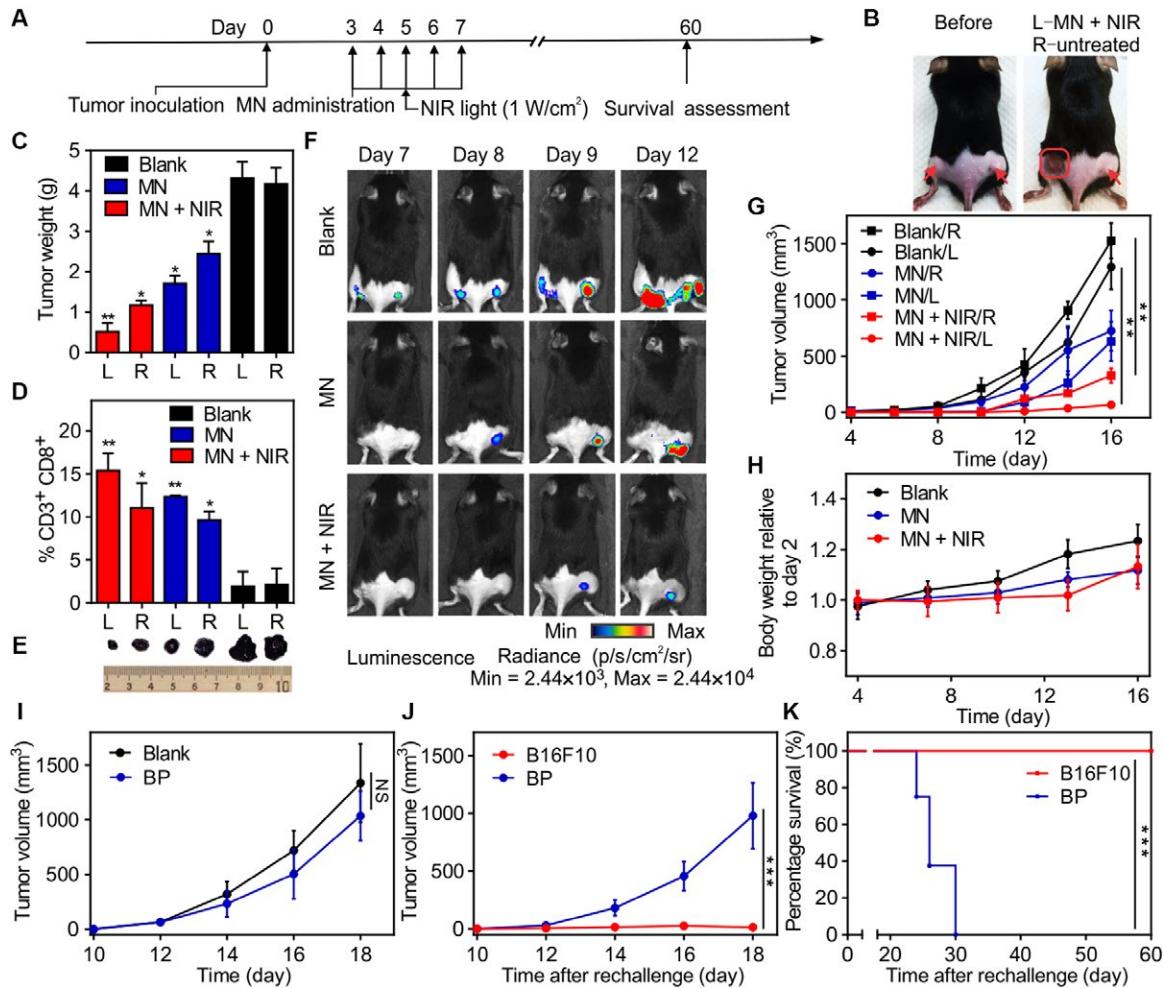
Fig. 6. Antitumor effect of local cancer immunotherapy treatment toward distant B16F10 tumors. (**A**) Schematic representation of the B16F10 tumor model. (**B**) Photograph of a mouse before and after MN administration. Red arrows indicate established tumors on both sides. L, left; R, right. The red line indicates the MN injection site. (**C**) Tumor weights in different experimental groups ($n = 3$). (**D**) Tumor-infiltrating CD8$^+$ T cells after treatments in different experimental groups ($n = 6$). Asterisks in (C) and (D) indicate statistically significant differences between blank MN and other groups. (**E**) Images of tumors extracted from treated mice indicated by the labels in (D). (**F**) In vivo bioluminescence imaging of treated B16F10 melanoma at different time points after treatment. One representative mouse per treatment group is shown. (**G**) Average tumor volumes in treated mice. (**H**) Body weights of treated mice ($n = 6$). (**I**) Average B16F10 tumor volumes in mice treated with MNs loaded with BP lysate with melanin and blank MN ($n = 8$). (**J**) Average tumor volumes in vaccinated mice rechallenged with either B16F10 cells or BP cells on day 80. (**K**) Kaplan-Meier survival curves for rechallenged mice. Data points represent mean ± SD ($n = 8$). Error bars indicate SD. Statistical significance was calculated by Student's t test and log-rank test (NS, $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

Body weight measurements indicated that the treated mice gained weight within the normal ranges (Fig. 6H). When the mice were vaccinated with tumor lysate of a different melanoma cell type [$BRAF^{V600E}PTEN^{-/-}$ Duke-clone 6 cell line (BPD6 or BP), syngeneic with C57BL/6J] (40, 41), minimal changes in the B16F10 tumor growth were observed (Fig. 6I), indicating the specificity of the immunological memory. Similarly, when vaccinated mice were rechallenged with B16F10 cells or BP cells, tumor protection was observed only in mice challenged with B16F10 (Fig. 6, J and K, and fig. S23).

Efficacy of MN patch in other tumor models
To demonstrate that the potency of proposed vaccination is not limited to the B16 melanoma model in which there is an up-regulation of the melanogenesis, we used a $BRAF^{V600E}$-mutated BP melanoma in C57BL/6J mice and a triple-negative breast cancer 4T1 carcinoma

tumor in BALB/cJ mice. Studies of MN loaded with synthetic melanin (with the same amount of pigment as quantified by spectrophotometry; table S1) and tumor lysate showed similar hyperthermic effect and enhanced immune responses with combined vaccination (fig. S24). Tumor regression and long-term survival were also achieved. Vaccination with the combined approach rendered 75 and 87% of mice resistant to BP and 4T1 engraftment, respectively (Fig. 7, A and B, and fig. S25). In the tumor-bearing mice, MN and NIR treatments induced complete remissions in 87 and 37% of mice engrafted with BP and 4T1 cells, respectively (Fig. 7, C and D, and fig. S25). Analysis of the local HSP70 expression showed a 2.5-fold increase in BP and a 4-fold increase in 4T1 models upon combined vaccination compared with the MN alone (fig. S26). Production of proinflammatory cytokines was also induced with the combined treatment, resulting in enhanced DC activation (figs. S27 and S28). Treatments were well
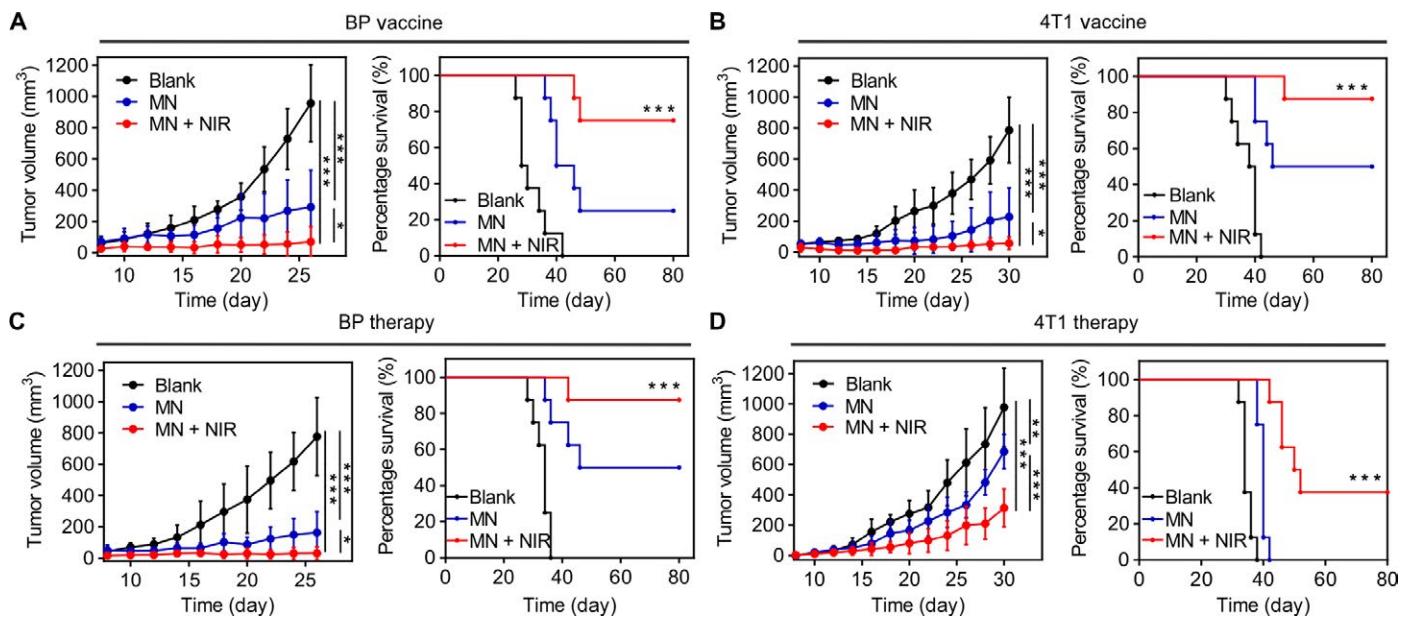
Fig. 7. Antitumor effect of local cancer immunotherapy treatment in various tumor models. (**A**) Average tumor growth and Kaplan-Meier survival rate of vaccinated C57BL/6J mice after BP tumor cell challenge. Mice were pretreated with blank MN (blank), MN loaded with BP tumor lysate and melanin (MN), or loaded MN combined with NIR irradiation (MN + NIR). (**B**) Average tumor growth and Kaplan-Meier survival rate of vaccinated BALB/cJ mice after 4T1 tumor cell challenge. Mice were pretreated with blank MN, MN loaded with 4T1 tumor lysate and melanin, or loaded MN combined with NIR irradiation. (**C**) Average tumor growth and Kaplan-Meier survival rate of C57BL/6J mice bearing established BP tumors. Mice were treated with blank MN, MN loaded with BP tumor lysate and melanin, or loaded MN combined with NIR irradiation. (**D**) Average tumor growth and Kaplan-Meier survival rate of BALB/cJ mice bearing established 4T1 tumors. Mice were treated with blank MN, MN loaded with 4T1 tumor lysate and melanin, or loaded MN combined with NIR irradiation. Data points represent mean ± SD ($n = 8$). Error bars indicate SD. Statistical significance was calculated by Student's t test and log-rank test (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

tolerated and did not cause weight loss or clinical signs of distant metastasis (figs. S29 and S30).

DISCUSSION

In the current study, we have integrated the B16F10 whole tumor lysate and GM-CSF into a transdermal MN patch to sustain the delivery of the cancer vaccine and target immune cell populations in the epidermis. By incorporating the melanin, we have demonstrated that the combined vaccine facilitated local immune activation upon NIR irradiation through the recruitment of DCs and other immune cells. Compared with other photosensitizing agents (*42*, *43*), melanin is a natural pigment with high biocompatibility and broad absorption spectrum. The efficient light-to-heat transduction mediated by melanin contributed to the rapid increase in skin temperature toward 42°C.

The combined treatment caused long-term survival with tumor rejection in 87% of vaccinated C57BL/6J mice. Complete tumor protection was further observed in mice rechallenged with B16F10 tumor cells. After NIR treatment, the elevated local microcirculatory blood perfusion was associated with the migration of local DCs and NK cells. Depletion of B cells, NK cells, and T cells attenuated the treatment efficacy in vaccinated mice, highlighting the relevance of the immune control in the tumor development. In accordance, the production of danger signals and proinflammatory cytokines triggered the immune activation. We have identified the elevated HSP70 and HSP90 expression and ROS levels in the surrounding tissues. This effect was also associated with local enrichment of IFN-γ, TNF-α, and IL-6. Characterization of other relevant danger signals and antigen adjuvants needs to be further examined.

In summary, the whole tumor lysate with melanin in a light-irradiated MN patch generates a hyperthermic-mimicking microenvironment that effectively recruits and activates immune cells at the vaccination site. As a result, the melanin-mediated patch prevents tumor engraftment in prophylactic models and causes sustained tumor regression in tumor-bearing mice. In addition, the natural melanin used in this treatment can also be extended to other biological pigments, such as carotenoid, xanthophylls, and bilirubin, for photomediated therapy. This method is also adaptable to deep-tissue photoacoustic imaging, biological labeling, and targeting a variety of diseases in a photo-/thermoresponsive manner (*44*–*46*). Nevertheless, the parameters associated with this immune activation need further optimization, for instance, the duration of NIR irradiation and the selection of light wavelength, as well as the treatment site. For the future study of the proposed approach, personal neoantigens and checkpoint inhibitors could be incorporated into the MNs to further enhance treatment specificity and efficacy (*47*). For the potential translation of the vaccine patch, it relies profoundly on a prolonged safety evaluation of the local and systemic toxicity, as well as other potential side effects.

MATERIALS AND METHODS
Study design
The objective of this study was to assess the effect of a melanin-mediated transdermal patch on cancer vaccine–based immunotherapy. The B16F10 whole tumor lysate containing melanin was integrated into an MN patch that allows controlled release of the cancer vaccine.

Upon NIR light irradiation on the patch, the local heating effect on boosting immune responses was evaluated. Mice subjects were purchased from the Jackson Laboratory, weighed, and randomly divided into different experimental groups. The numbers of sampling and experimental replicates were included in each figure legend.

Preparation and characterization of MNs

All MNs were prepared using silicone molds with arrays of conical holes machined by laser ablation (Blueacre Technology Ltd.). Each MN had a 300 μm by 300 μm round base tapering to a height of 800 μm with a tip radius of around 5 μm. MNs were arranged in a 15×15 array with a 600-μm center-to-center spacing. GM-CSF solution (0.1 mg/ml, 10-μl final volume) was directly deposited by pipetting onto each silicone micromold surface, followed by vacuum (600 mmHg) condition for 5 min to allow the solution to flow into the cavities. After that, 0.2 ml of 4.0 weight % (wt %) methacrylated hyaluronic acid solution mixed with N,N'-methylenebisacrylamide (2.0 wt %) and photoinitiator (Irgacure 2959; 0.05 wt %) were directly deposited by pipetting onto each silicone micromold surface, followed by vacuum (600 mmHg) condition for 5 min and centrifugation at 2000 rpm for 5 min (Hettich Universal 32R centrifuge). GM-CSF layer was cross-linked via ultraviolet irradiation (wavelength, 365 nm) for 10 s, and a piece of 4 cm by 9 cm silver adhesive tape was applied around the 2 cm by 2 cm micromold baseplate. Three milliliters of homogenized B16F10 tumor lysate (containing 1.5 mg of extracted tumor lysate proteins) in 4.0 wt % mixed methacrylated hyaluronic acid solution was added to the prepared micromold reservoir. We also used hyaluronic acid solution with different concentrations of tumor protein from 0.25 to 0.5 to 1.0 mg/ml. Protein concentration was quantified by measuring extracted protein content in the tumor lysate supernatant using T-PER (tissue protein extraction reagent) and Bradford assay before the MN fabrication. For the BP or 4T1 tumor lysate–loaded MNs, 50-μg melanin was dissolved in 4.0 wt % mixed methacrylated hyaluronic acid solution with tumor lysate (containing 1.5 mg of extract tumor lysate proteins, 3.0-ml final volume) to fabricate the MN matrix. For the blank MNs without tumor lysate, mixed methacrylated hyaluronic acid solution was used without the tumor lysate. For the melanin-loaded MNs, 50-μg melanin was dissolved in 4.0 wt % mixed methacrylated hyaluronic acid solution to fabricate the MN matrix. The stock melanin solution (100 mg/ml) was prepared by dissolving melanin in 1.0 M sodium hydroxide and heating to 99°C for 10 min. Final formulation was dried at 25°C in a vacuum desiccator overnight. After desiccation was completed, needle arrays were carefully separated from the silicone molds and cross-linked via ultraviolet irradiation. The needle base was tailored into a square shape. Fluorescent MNs were fabricated with phalloidin-labeled tumor lysate and rhodamine B– or Cy5.5-labeled hyaluronic acid. The preparation and storage of MNs were under sterilized condition. Morphology of MNs was characterized on a FEI Verios 460L field-emission scanning electron microscope operating at 20 kV after sputter coating with gold/palladium at the Analytical Instrumentation Facility. Cross sections of MNs were obtained by cutting 5-μm slides using Thermo Fisher Scientific HM525 NX Cryostat and stained with Alexa Fluor 488 phalloidin and Hoechst. Fluorescence images of MNs were taken by the Olympus IX70 multiparameter fluorescence microscope.

NIR-responsive property of MN

To evaluate the property of the MN patch in response to NIR light irradiation, we irradiated samples with 808-nm NIR laser light. The diode infrared laser module (Opto Engine LLC, MDL-N-808) was approved by a laser safety officer of the North Carolina State University (NC State) Environmental Health and Safety Center. The MN patches were placed on a piece of white paper with the needle tips facing down. The laser wire was fixed in place using a plate stand with a support rod and a clamp. The distance between the patch and spotlight was fixed at 10 cm, and the spot size was ~1 cm^2. The output energy of the diode infrared laser module was adjusted within 1.0 W and led to the intensity per unit area ($E = I/d^2$) of 1.0 W/cm^2 (at 808 nm). The NIR light was irradiated continuously or intermittently as needed. Surface temperature changes of the MN patch backing were recorded, and the highest temperature was controlled below 42°C in real time using a thermal imager (FLIR E4) and a noncontact infrared thermometer gun (Leegoal). For intermittent irradiation, the MNs were repeatedly exposed to the laser (1.0 W/cm^2), and the local temperature reached about 42°C for 1 min. Subsequently, the laser was turned off for 1 min. This cycle was repeated four times to assess the responsive behavior of MN patch under repeated NIR irradiation. For the continuous irradiation, samples with different parameters were tested. The effect of melanin content on the light-responsive behavior was tested on MNs loaded with tumor lysate of different protein concentration or blank MN with hyaluronic acid. Quantitative surface temperature changes of the representative MNs were recorded under continuous NIR irradiation at 1.0 W/cm^2. In a separate experiment, the laser power flux was controlled. During continuous NIR exposure for 2 min, the maximal temperature of the MN surface was recorded and plotted. The effect of MN backing thickness on the light-responsive behavior was tested on MNs with continuous NIR irradiation at 1.0 W/cm^2. The thickness of the MN patch was measured with an average value of 181 μm and an SD of 12.5 μm using a digimatic indicator (Mitutoyo Corp. ID-C112E Series 543). MN patch samples with different backing thicknesses (169, 175, 179, 181, and 202 μm) were used.

In vitro DC activation

Bone marrow was collected by flushing the femur and tibia with complete RPMI 1640 containing 10% fetal bovine growth serum. After lysis of red blood cells (lysate solution, Cwbiotech), 1×10^6 bone marrow cells were seeded in six-well culture dishes with 3.0 ml of the culture medium containing GM-CSF (20 ng/ml) and 50 μM β-mercaptoethanol (Bio-Rad, Hercules, CA, USA). On day 3, an additional 4.0 ml of the same medium with GM-CSF was added into the plates. On day 6, half of the culture supernatant was collected and centrifuged. Cells were resuspended in RPMI 1640 and added back into the original plates. On day 7, nonadherent cells were collected and used as bone marrow–derived DCs for further research use. Another murine DC line JAWS II cells were cultured in minimum essential medium Eagle alpha modification (Sigma-Aldrich) supplemented with GM-CSF (5 ng/ml). DCs were left unstimulated or stimulated for 12 hours with blank MN (100 μg/ml) or tumor lysate–loaded MN release medium or lipopolysaccharide (LPS; 100 ng/ml). After the stimulation, the cells were exposed to NIR irradiation (1.0 W/cm^2) at a distance of 10 cm for 0, 5, 10, 15, and 20 min, respectively. After cell incubation for 4 hours, the supernatants were collected and measured by IL-12 p70 mouse ELISA kit (Thermo Fisher Scientific, MC0121). The cells were washed and stained with LIVE/DEAD assay (Thermo Fisher Scientific, L3224) and CD80$^+$, and CD86$^+$ maturation marker-specific antibodies and subsequently analyzed by the confocal microscopy and flow cytometer. Both bone marrow–derived DCs and JAWS II cell line were used for characterization of DC activation. The LIVE/DEAD assay imaging data were obtained from the JAWS II cell line.

Mice and in vivo tumor models

Female C57BL/6J mice, BALB/cJ mice, CD11c-DTR transgenic mice [B6.FVB-Tg(Itgax-DTR/EGFP)57Lan/J; stock no. 004509], and $Rag1^{-/-}$ knockout mice (B6.129S7-$Rag1^{tm1Mom}$/J; stock no. 002216) were purchased from the Jackson Laboratory. Mice were weighed and randomly divided into different groups. On day 0, healthy mice were treated with MNs loaded with tumor lysates and GM-CSF, blank MNs loaded with synthetic melanin without tumor lysates, or blank MNs containing hyaluronic acid only (blank). MN patches were applied into the skin of the caudal-dorsal area for about 10 min and further fixed using Skin Affix surgical adhesive (Medline Industries Inc.). After injection of the MNs, NIR irradiation was performed on the localized MN region for 10 min each day for five successive days after immunization (MN + NIR). The diode infrared laser module at 808 nm (Opto Engine LLC, MDL-N-808) was approved by a laser safety officer of the NC State University Environmental Health and Safety Center. Mice in the control groups were treated with either vaccine MNs without NIR irradiation (MN), blank MNs loaded with synthetic melanin without tumor lysate with NIR irradiation (melanin), or blank MNs containing hyaluronic acid only with NIR irradiation (blank). Surface temperature changes of the regional skin were recorded and controlled below 42°C in real time using a thermal imager (FLIR E4) and a noncontact infrared thermometer gun (Leegoal). On day 10, 1×10^6 B16F10 tumor cell lines in 25 μl of phosphate-buffered saline (PBS) were subcutaneously transplanted into the flank of the C57BL/6J mice, CD11c-DTR transgenic mice, and $Rag1^{-/-}$ knockout mice. Tumor-free mice were rechallenged with 1×10^6 B16F10 tumor cells in 25 μl of PBS 80 days after the first tumor inoculation. For the depletion antibody study, we depleted specific T cell, B cell, and NK cell populations in mouse models. Mice were intraperitoneally given 200 μg of antibody purified from mouse thymus or spleen dissolved in 200 μl of PBS. Antibodies against CD4 (BioLegend, LEAF 100435), CD8 (BioLegend, LEAF 100735), CD19 (BioLegend, LEAF 152402), and NK-1.1 (BioLegend, LEAF 108712) were administered twice weekly for 3 weeks, starting 1 week before the tumor inoculation. Depletions were confirmed by flow cytometry of splenic suspension. For another melanoma model, 1×10^6 BP tumor cells in 25 μl of PBS were subcutaneously transplanted into the flank of the C57BL/6J mice. For the carcinoma tumor model, 1×10^6 4T1 tumor cells in 25 μl of PBS were subcutaneously transplanted into the flank of the BALB/cJ mice. For experimental metastasis model, 1×10^5 tumor cells were intravenously infused into mice via the tail vein. In another set of experiments, tumor cells were subcutaneously transplanted into the flank of the mice on day 0. The tumor-bearing mice were weighed and randomly divided into four groups when the tumor volume reached around 50 mm³ on day 3. After that, the mice were peritumorally administered with sterile MN loaded with tumor lysate and GM-CSF (MN) or blank MN containing hyaluronic acid only (blank). After that, NIR light was irradiated on the MN patch for 10 min during the following 5 days from days 3 to 7. Tumor growth was measured by a digital caliper or monitored by bioluminescence signals of luciferase-tagged cells. The tumor volume (cubic millimeter) was calculated as $1/2 \times$ long diameter $\times$ (short diameter)^2.

To assess potential toxicity, we monitored mice daily for weight loss. H&E staining was performed on the organs collected from the mice following the standard procedure (Histology Laboratory at NC State College of Veterinary Medicine). Lungs were excised, and macroscopically visible metastases were counted. The resected primary tumors were stored at −20°C for immunofluorescence staining or fixed in 4% paraformaldehyde for subsequent analysis.

Identification of DC subsets, T cells, and cytokines

Antibodies used were purchased from Thermo Fisher Scientific and BioLegend Inc. Staining antibodies, including CD3 (Thermo Fisher Scientific, A18644), CD4 (Thermo Fisher Scientific, A18667), CD8 (Thermo Fisher Scientific, A18609), CD11c (BioLegend, 117309), CD49b (BioLegend, 108909), CD80 (BioLegend, 104707), CD86 (BioLegend, 105007), PIR-A/B (gp91, BioLegend, 144103), tetramer (MBL International, H-2D^b gp100, EGSRNQDWL-P), and fluorogenic CellROX Deep Red reagent (BioLegend, C10422) were used for fluorescence-activated cell sorting (FACS) and were analyzed following the manufacturers' instructions. Organs (skin, tumor, lymph node, and spleen) were harvested and minced into 2- to 4-mm pieces using scissors or scalpel blade. A single-cell suspension was prepared in the cell staining buffer (BioLegend, 420201). CellROX reagents at predetermined optimum concentrations were added to the cells and incubated on ice for 20 to 30 min in the dark for ROS detection. The relevant tetramer phycoerythrin (PE) was stained at room temperature for 30 min in the dark before additional staining with appropriately conjugated fluorescent antibodies. Stained cells were analyzed on a FACSCalibur instrument (BD) and using FlowJo software. To determine the concentration of different cytokines at the vaccine MN site, patch and adjacent tissues were excised and digested with T-PER (Pierce). Cytokine concentrations in the extracted patch were analyzed with a BioLegend's LEGENDplex bead-based immunoassays, according to the manufacturer's instructions.

Statistical analysis

Statistical analysis was evaluated using GraphPad Prism (6.0). Statistical analysis was performed with paired Student's *t* test and analysis of variance (ANOVA). *P* values for Kaplan-Meier curves were calculated with log-rank test. *P* values of 0.05 or less were considered significant.

Study approval

All mouse studies were performed following animal protocols approved by the Institutional Animal Care and Use Committee at the NC State University and the University of North Carolina at Chapel Hill (UNC-CH).

REFERENCES AND NOTES

1. J. Couzin-Frankel, Cancer immunotherapy. *Science* 342, 1432–1433 (2013).
2. D. Ricklin, G. Hajishengallis, K. Yang, J. D. Lambris, Complement: A key system for immune surveillance and homeostasis. *Nat. Immunol.* 11, 785–797 (2010).
3. J. B. Swann, M. J. Smyth, Immune surveillance of tumors. *J. Clin. Invest.* 117, 1137–1146 (2007).
4. L. Gu, D. J. Mooney, Biomaterials and emerging anticancer therapeutics: Engineering the microenvironment. *Nat. Rev. Cancer* 16, 56–66 (2015).
5. L. E. Klevorn, R. M. Teague, Adapting cancer immunotherapy models for the real world. *Trends Immunol.* 37, 354–363 (2016).
6. D. R. Littman, Releasing the brakes on cancer immunotherapy. *Cell* 162, 1186–1190 (2015).
7. C. G. Drake, E. J. Lipson, J. R. Brahmer, Breathing new life into immunotherapy: Review of melanoma, lung and kidney cancer. *Nat. Rev. Clin. Oncol.* 11, 24–37 (2013).
8. M. E. Dudley, S. A. Rosenberg, Adoptive-cell-transfer therapy for the treatment of patients with cancer. *Nat. Rev. Cancer* 3, 666–675 (2003).
9. S. A. Rosenberg, N. P. Restifo, J. C. Yang, R. A. Morgan, M. E. Dudley, Adoptive cell transfer: A clinical path to effective cancer immunotherapy. *Nat. Rev. Cancer* 8, 299–308 (2008).
10. M. T. Stephan, J. J. Moon, S. H. Um, A. Bershteyn, D. J. Irvine, Therapeutic cell engineering with surface-conjugated synthetic nanoparticles. *Nat. Med.* 16, 1035–1041 (2010).
11. B. Huang, W. D. Abraham, Y. Zheng, S. C. Bustamante López, S. S. Luo, D. J. Irvine, Active targeting of chemotherapy to disseminated tumors using nanoparticle-carrying T cells. *Sci. Transl. Med.* 7, 291ra94 (2015).
12. D. J. Irvine, M. A. Swartz, G. L. Szeto, Engineering synthetic vaccines using cues from natural immunity. *Nat. Mater.* 12, 978–990 (2013).
13. M. F. Bachmann, G. T. Jennings, Vaccine delivery: A matter of size, geometry, kinetics and molecular patterns. *Nat. Rev. Immunol.* 10, 787–796 (2010).
14. J. Banchereau, A. K. Palucka, Dendritic cells as therapeutic vaccines against cancer. *Nat. Rev. Immunol.* 5, 296–306 (2005).
15. M. V. Dhodapkar, M. Sznol, B. Zhao, D. Wang, R. D. Carvajal, M. L. Keohan, E. Chuang, R. E. Sanborn, J. Lutzky, J. Powderly, H. Kluger, S. Tejwani, J. Green, V. Ramakrishna, A. Crocker, L. Vitale, M. Yellin, T. Davis, T. Keler, Induction of antigen-specific immunity with a vaccine targeting NY-ESO-1 to the dendritic cell receptor DEC-205. *Sci. Transl. Med.* 6, 232ra51 (2014).
16. A. D. Garg, L. Vandenberk, C. Koks, T. Verschuere, L. Boon, S. W. Van Gool, P. Agostinis, Dendritic cell vaccines based on immunogenic cell death elicit danger signals and T cell–driven rejection of high-grade glioma. *Sci. Transl. Med.* 8, 328ra27 (2016).
17. T. R. Fadel, F. A. Sharp, N. Vudattu, R. Ragheb, J. Garyu, D. Kim, E. P. Hong, N. Li, G. L. Haller, L. D. Pfefferle, S. Justesen, K. C. Herold, T. M. Fahmy, Corrigendum: A carbon nanotube-polymer composite for T-cell therapy. *Nat. Nanotechnol.* 9, 723 (2014).
18. J. V. Kim, J.-B. Latouche, I. Rivière, M. Sadelain, The ABCs of artificial antigen presentation. *Nat. Biotechnol.* 22, 403–410 (2004).
19. C. L.-L. Chiang, G. Coukos, L. Kandalaft, Whole tumor antigen vaccines: Where are we? *Vaccines* 3, 344–372 (2015).
20. W. Kratky, C. Reis e Sousa, A. Oxenius, R. Spörri, Direct activation of antigen-presenting cells is required for CD8$^+$ T-cell priming and tumor vaccination. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17414–17419 (2011).
21. S. B. Stephan, A. M. Taber, I. Jileaeva, E. P. Pegues, C. L. Sentman, M. T. Stephan, Biopolymer implants enhance the efficacy of adoptive T-cell therapy. *Nat. Biotechnol.* 33, 97–101 (2015).
22. J. Kim, W. A. Li, Y. Choi, S. A. Lewin, C. S. Verbeke, G. Dranoff, D. J. Mooney, Injectable, spontaneously assembling, inorganic scaffolds modulate immune cells in vivo and increase vaccine efficacy. *Nat. Biotechnol.* 33, 64–72 (2015).
23. O. A. Ali, D. Emerich, G. Dranoff, D. J. Mooney, In situ regulation of DC subsets and T cells mediates tumor regression in mice. *Sci. Transl. Med.* 1, 8ra19 (2009).
24. S. S. Evans, E. A. Repasky, D. T. Fisher, Fever and the thermal regulation of immunity: The immune system feels the heat. *Nat. Rev. Immunol.* 15, 335–349 (2015).
25. K. Sato, N. Sato, B. Xu, Y. Nakamura, T. Nagaya, P. L. Choyke, Y. Hasegawa, H. Kobayashi, Spatially selective depletion of tumor-associated regulatory T cells with near-infrared photoimmunotherapy. *Sci. Transl. Med.* 8, 352ra110 (2016).
26. J. Fu, D. B. Kanne, M. Leong, L. H. Glickman, S. M. McWhirter, E. Lemmens, K. Mechette, J. J. Leong, P. Lauer, W. Liu, K. E. Sivick, Q. Zeng, K. C. Soares, L. Zheng, D. A. Portnoy, J. J. Woodward, D. M. Pardoll, T. W. Dubensky Jr., Y. Kim, STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade. *Sci. Transl. Med.* 7, 283ra52 (2015).
27. T. S. Kupper, R. C. Fuhlbrigge, Immune surveillance in the skin: Mechanisms and clinical consequences. *Nat. Rev. Immunol.* 4, 211–222 (2004).
28. K. Nagao, F. Ginhoux, W. W. Leitner, S.-I. Motegi, C. L. Bennett, B. E. Clausen, M. Merad, M. C. Udey, Murine epidermal Langerhans cells and langerin-expressing dermal dendritic cells are unrelated and exhibit distinct functions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3312–3317 (2009).
29. C. Wang, Y. Ye, G. M. Hochu, H. Sadeghifar, Z. Gu, Enhanced cancer immunotherapy by microneedle patch-assisted delivery of anti-PD1 antibody. *Nano Lett.* 16, 2334–2340 (2016).
30. J. Yu, Y. Zhang, Y. Ye, R. DiSanto, W. Sun, D. Ranson, F. S. Ligler, J. B. Buse, Z. Gu, Microneedle-array patches loaded with hypoxia-sensitive vesicles provide fast glucose-responsive insulin delivery. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8260–8265 (2015).
31. A. Filson, J. Hope, Isolation of melanin granules. *Nature* 179, 211 (1957).
32. A. J. Lea, Solubility of melanins. *Nature* 170, 709 (1952).
33. R. Zhang, Q. Fan, M. Yang, K. Cheng, X. Lu, L. Zhang, W. Huang, Z. Cheng, Engineering melanin nanoparticles as an efficient drug-delivery system for imaging-guided chemotherapy. *Adv. Mater.* 27, 5063–5069 (2015).
34. S. P. Sullivan, D. G. Koutsonanos, M. del Pilar Martin, J. W. Lee, V. Zarnitsyn, S.-O. Choi, N. Murthy, R. W. Compans, I. Skountzou, M. R. Prausnitz, Dissolving polymer microneedle patches for influenza vaccination. *Nat. Med.* 16, 915–920 (2010).
35. M. L. Moraes, P. J. Gomes, P. A. Ribeiro, P. Vieira, A. A. Freitas, R. Kohler, O. N. Oliveira Jr., M. Raposo, Polymeric scaffolds for enhanced stability of melanin incorporated in liposomes. *J. Colloid Interface Sci.* 350, 268–274 (2010).
36. B. M. Carreno, V. Magrini, M. Becker-Hapak, S. Kaabinejadian, J. Hundal, A. A. Petti, A. Ly, W. R. Lie, W. H. Hildebrand, E. R. Mardis, G. P. Linette, A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 348, 803–808 (2015).
37. J. C. Castle, S. Kreiter, J. Diekmann, M. Löwer, N. van de Roemer, J. de Graaf, A. Selmi, M. Diken, S. Boegel, C. Paret, M. Koslowski, A. N. Kuhn, C. M. Britten, C. Huber, Ö. Türeci, U. Sahin, Exploiting the mutanome for tumor vaccination. *Cancer Res.* 72, 1081–1091 (2012).
38. A. Murshid, J. Gong, S. K. Calderwood, The role of heat shock proteins in antigen cross presentation. *Front. Immunol.* 3, 63 (2012).
39. M. M. Hoffmann, C. Molina-Mendiola, A. D. Nelson, C. A. Parks, E. E. Reyes, M. J. Hansen, G. Rajagopalan, L. R. Pease, A. G. Schrum, D. Gil, Co-potentiation of antigen recognition: A mechanism to boost weak T cell responses and provide immunotherapy in vivo. *Sci. Adv.* 1, e1500415 (2015).
40. D. Dankort, D. P. Curley, R. A. Cartlidge, B. Nelson, A. N. Karnezis, W. E. Damsky Jr., M. J. You, R. A. DePinho, M. McMahon, M. Bosenberg, Braf^{V600E} cooperates with Pten loss to induce metastatic melanoma. *Nat. Genet.* 41, 544–552 (2009).
41. A. Holtzhausen, F. Zhao, K. S. Evans, M. Tsutsui, C. Orabona, D. S. Tyler, B. A. Hanks, Melanoma-derived Wnt5a promotes local dendritic-cell expression of IDO and immunotolerance: Opportunities for pharmacologic enhancement of immunotherapy. *Cancer Immunol. Res.* 3, 1082–1095 (2015).
42. Q. Chen, L. Xu, C. Liang, C. Wang, R. Peng, Z. Liu, Photothermal therapy with immune-adjuvant nanoparticles together with checkpoint blockade for effective cancer immunotherapy. *Nat. Commun.* 7, 13193 (2016).
43. J. Conde, N. Oliva, Y. Zhang, N. Artzi, Local triple-combination therapy results in tumour regression and prevents recurrence in a colon cancer model. *Nat. Mater.* 15, 1128–1138 (2016).
44. W. Sun, Q. Hu, W. Ji, G. Wright, Z. Gu, Leveraging physiology for precision drug delivery. *Physiol. Rev.* 97, 189–225 (2017).
45. Y. Zhang, J. Yu, A. Kahkoska, Z. Gu, Photoacoustic drug delivery. *Sensors* 17, 1400 (2017).
46. L. Cheng, C. Wang, L. Feng, K. Yang, Z. Liu, Functional nanomaterials for phototherapies of cancer. *Chem. Rev.* 114, 10869–10939 (2014).
47. C. Wang, Y. Ye, Q. Hu, A. Bellotti, Z. Gu, Tailoring biomaterials for cancer immunotherapy: Emerging trends and future outlook. *Adv. Mater.* 29, 1606036 (2017).

CANCER THERAPY

Developmental phosphoproteomics identifies the kinase CK2 as a driver of Hedgehog signaling and a therapeutic target in medulloblastoma

Teresa Purzner[1,2]*, James Purzner[1,2], Taylor Buckstaff[3], Giorgio Cozza[4], Sharareh Gholamin[5], Jessica M. Rusert[6], Tom A. Hartl[1], John Sanders[7], Nicholas Conley[8], Xuecai Ge[1,9], Marc Langan[†], Vijay Ramaswamy[10,11], Lauren Ellis[1], Ulrike Litzenburger[12], Sara Bolin[13], Johanna Theruvath[14], Ryan Nitta[13], Lin Qi[15], Xiao-Nan Li[15], Gordon Li[13], Michael D. Taylor[11,16], Robert J. Wechsler-Reya[6,17], Lorenzo A. Pinna[18,19], Yoon-Jae Cho[20,21,22‡], Margaret T. Fuller[1‡], Joshua E. Elias[7], Matthew P. Scott[1]*

A major limitation of targeted cancer therapy is the rapid emergence of drug resistance, which often arises through mutations at or downstream of the drug target or through intrinsic resistance of subpopulations of tumor cells. Medulloblastoma (MB), the most common pediatric brain tumor, is no exception, and MBs that are driven by sonic hedgehog (SHH) signaling are particularly aggressive and drug-resistant. To find new drug targets and therapeutics for MB that may be less susceptible to common resistance mechanisms, we used a developmental phosphoproteomics approach in murine granule neuron precursors (GNPs), the developmental cell of origin of MB. The protein kinase CK2 emerged as a driver of hundreds of phosphorylation events during the proliferative, MB-like stage of GNP growth, including the phosphorylation of three of the eight proteins commonly amplified in MB. CK2 was critical to the stabilization and activity of the transcription factor GLI2, a late downstream effector in SHH signaling. CK2 inhibitors decreased the viability of primary SHH-type MB patient cells in culture and blocked the growth of murine MB tumors that were resistant to currently available Hh inhibitors, thereby extending the survival of tumor-bearing mice. Because of structural interactions, one CK2 inhibitor (CX-4945) inhibited both wild-type and mutant CK2, indicating that this drug may avoid at least one common mode of acquired resistance. These findings suggest that CK2 inhibitors may be effective for treating patients with MB and show how phosphoproteomics may be used to gain insight into developmental biology and pathology.

INTRODUCTION

A key challenge in developing cancer therapeutics is the identification of a target protein that is essential to the growth, survival, or metastasis of a tumor. One path to such proteins is to test developmental regulators that operate in normal cells from which the tumor is derived. A prime example is medulloblastoma (MB), the most common malignant pediatric brain tumor. Developmental, genetic, and transcriptional analyses have established a clear parallel between sonic hedgehog (SHH)–subtype MB and granule neuron precursors (GNPs) (1–3). During normal cerebellar development, GNPs proliferate extensively in response to hedgehog (Hh) signaling (2) before differentiating into granule neurons, the most abundant type of neuron in the brain (Fig. 1A) (4). In mice, this period begins at postnatal day 1 (P1), peaks at P7, and is largely complete by P14, at which point remaining GNPs have stopped dividing and have begun differentiating (5) into granule neurons (2, 6). Sustained Hh target gene activity in GNPs, due to reduced function of pathway components that negatively regulate Hh signal transduction, such as Patched (PTCH1) (1), or to heightened function of activating Hh signal transducers or effectors, such as GLI2 (3), results in continued proliferation of GNPs beyond P14 and eventual rise of SHH-type MB.

To date, drugs for SHH-type MB inhibit Smoothened (SMO), a transmembrane protein that acts early in the Hh signal transduction pathway. Patients treated with SMO inhibitors initially have dramatic tumor regression but eventually develop resistance due to mutations in *SMO* or in genes encoding downstream components of the Hh pathway (7). Worse, at the time of diagnosis, 49% of infants and 59% of children have mutations downstream of SMO; hence, these tumors

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Division of Neurosurgery, University of Toronto, Toronto, Ontario M5S1A1, Canada. [3]Department of Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [4]Department of Molecular Medicine, University of Padua, Padova, PD 35122, Italy. [5]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Tumor Initiation and Maintenance Program, National Cancer Institute–Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA. [7]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [8]Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305, USA. [9]Department of Molecular and Cell Biology, University of California, Merced, Merced, CA 95340, USA. [10]Division of Haematology/Oncology, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [11]Developmental and Stem Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [12]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA. [13]Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA 94305, USA. [14]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA. [15]Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA. [16]Division of Neurosurgery, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [17]Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA 92123, USA. [18]Department of Biomedical Sciences, University of Padua, Padova, PD 35122, Italy. [19]National Research Council Neuroscience Institute, Padova, PD 35122, Italy. [20]Papé Family Pediatric Research Institute, Department of Pediatrics, Oregon Health and Science University, Portland, OR 97239, USA. [21]Knight Cancer Institute, Oregon Health & Science University, Portland, OR 97239, USA. [22]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA.
*Corresponding author. Email: mscott@stanford.edu (M.P.S.); tpurzner@stanford.edu (T.P.)
†Retired.
‡These authors contributed equally to this work.
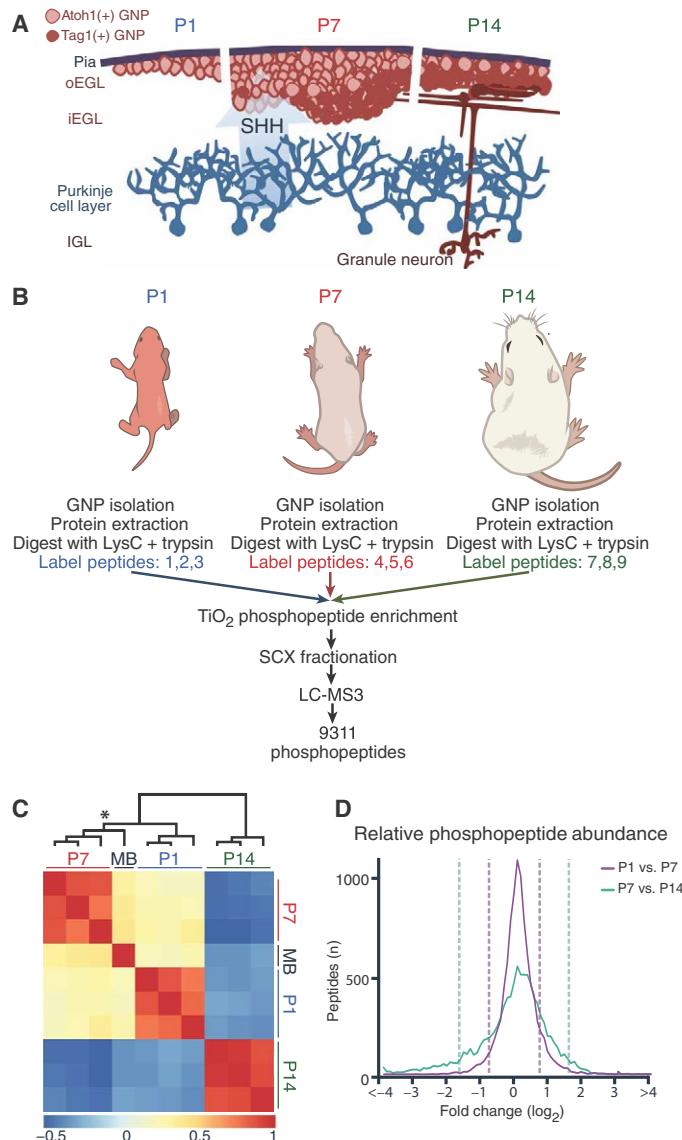
Fig. 1. Quantitative mapping of the phosphoproteome during GNP development. (**A**) Schematic of early postnatal proliferation and differentiation of GNPs. Pink, Atoh1-positive proliferative GNPs; red, postmitotic GNPs. oEGL/iEGL, outer/inner external granule layer; IGL, internal granule layer. (**B**) Experimental scheme for the phosphoproteomic assays. (**C**) Heat map representing relative phosphopeptide abundance and undirected clustering among three biological replicates and sample types (P1, P7, and P14 GNPs and $Ptch^{+/-}$ MB). Color axis = R^2. *Cluster branch of P7 GNPs versus MB sample. (**D**) Distribution of relative phosphopeptide abundance at developmental transitions. Dashed lines = 1.5 SDs. Mass spectrometry (MS) was performed in $n = 3$ biological replicates per experimental time point (14 to 40 mice per replicate at each time point).

are resistant to SMO inhibitors from the start (*8*). Despite the great potential of Hh pathway inhibitors, children with MBs continue to receive multiple nontargeted therapies and, consequently, sustain long-term neurological and cognitive problems. There is a pressing need to identify novel drug targets that affect Hh signal transduction downstream of SMO, preferably at late steps in the pathway. Ideally, inhibition of this target using a specific drug would be refractory to single mutations of Hh pathway components.

To identify important candidate drug targets for MBs, and to explore regulation of Hh transduction, we performed a proteome-wide analysis of in vivo phosphorylation events occurring in murine GNPs during the initiation, peak, and completion of Hh-driven proliferation. Unlike genome-wide transcription assays, phosphoproteomics provided measurement of the modified state of proteins at each developmental stage and thereby inferred which kinases have stage-specific changes in activity. Kinases that are specifically active during the MB-like P7 stage of GNP development may also be required for MB growth. Kinases are critical regulators of Hh signaling (*9*, *10*) and are attractive drug targets, susceptible to inhibition by competitive inhibitors, small molecules that interfere with the binding of adenosine 5′-triphosphate (ATP) (*11–13*). If a kinase has multiple substrates that are important for Hh transduction and/or survival of MB cells, then the drug will have multiple restraining influences on tumor growth, and mutation of any single substrate is unlikely to cause drug resistance.

RESULTS

Proliferating and postmitotic GNPs have distinct phosphoproteomes

Using a protocol summarized schematically in Fig. 1B, we identified 9311 unique phosphorylation events in GNPs purified from P1, P7, and P14 mouse cerebellum. The false discovery rate (FDR) was <1% (table S1). All samples were prepared in biological triplicate, and the results were strongly reproducible ($R^2 = 0.76$; Fig. 1C and fig. S1). A single MB sample generated from $Ptch1^{+/-}$ mice (*1*, *14*) was included for comparison (Fig. 1C and table S1). In keeping with previous literature (*5*), P1 and P7 Atoh1-positive cells were highly proliferative, whereas P14 GNPs were almost universally postmitotic (fig. S2). Many changes in phosphorylation distinguished proliferating (P1 and P7) from early postmitotic (P14) GNPs (Fig. 1, C and D). Only 4.9% of the phosphorylated peptides changed in relative phosphorylation level more than twofold from P1 to P7 (3.6% increased, 1.3% decreased); in contrast, 21.4% of phosphorylated peptides changed more than twofold between P7 and P14 (18.2% decreased, 3.2% increased; Fig. 1D). Strikingly, the pattern of protein abundance and phosphorylation in P7 GNP samples was more similar to that of MB cells than to that of either P1 or P14 GNPs, demonstrating strong similarity between peak proliferating GNPs and MB even at the minute-to-minute kinetics of phosphorylation (Fig. 1C, asterisk). Peptides more highly phosphorylated in P1, P7, and MB relative to P14 included known pro-proliferative phosphorylation sites of the cell cycle regulator retinoblastoma 1 (RB1) (fig. S2) and the proteins encoded by three of eight genes that commonly undergo copy number aberrations in human MBs (*Gab1*, *Irs2*, and *Yap-1*) (*15*).

CK2 substrates are more highly phosphorylated in GNPs isolated at P7

Motif analysis of phosphorylation events in GNPs that changed in frequency (>1.5 SD) over time suggested candidate kinases whose activity might be different during proliferation versus differentiation. Sixteen phosphorylation consensus motifs were prominent among 1522 protein sequences that had phosphorylation changes (Fig. 2A). Four observed motifs [for cyclin-dependent kinases (CDKs), protein kinase A (PKA), and PKC] were consistent with past studies (*4*, *10*, *16*). Of the remaining motifs, 7 of 12 matched those targeted by the protein kinase CK2 (casein kinase 2)—the target motifs of which have acidic residues C-terminal to a phosphorylated serine or threonine
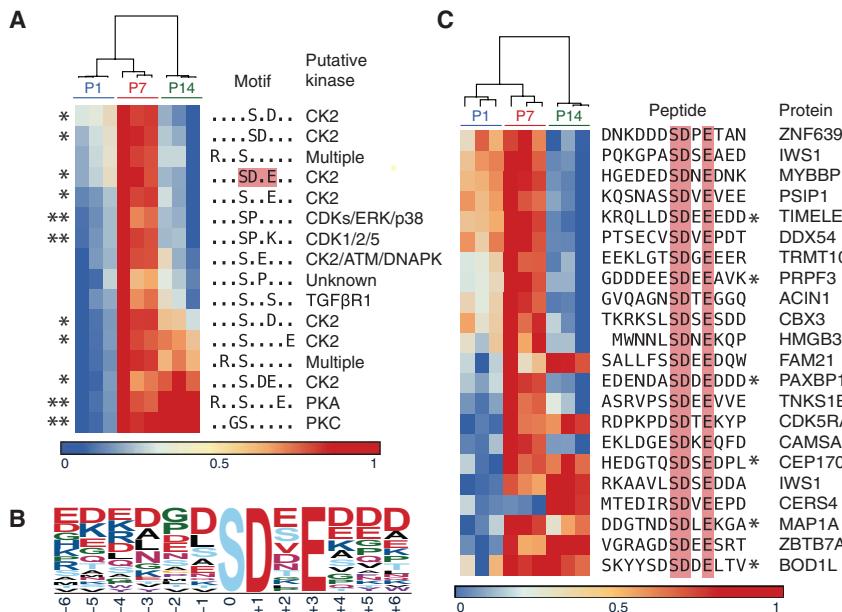
Fig. 2. Phosphoproteomic data indicate increased phosphorylation at CK2 motifs at P7. (A and B) The 16 motifs enriched in the fractions of altered phosphopeptide abundance (>1.5 SDs) in the phosphoproteomic analysis of GNPs purified from mice harvested on P1, P7, and P14 (A). The predicted kinase of each motif is noted down the right. *CK2 target motifs. **Known regulators of GNP development. Color axis: Fraction of maximum enrichment. Red box: CK2 motif characterized by motif logo (B). **(C)** Enriched phosphopeptides, and the corresponding protein, contributing to the SD.E motif–enriched phosphopeptides described in (A). Proteins include known (*) and currently uncharacterized CK2 substrates. Data are n = 3 biological replicates per experimental time point (14 to 40 mice per replicate at each time point).

(Fig. 2, A to C, and table S2) (17). Notably, of 278 phosphorylated peptides with possible CK2 target motifs, 83.8% were more highly phosphorylated in P7 compared to P1 or P14. CK2 itself was not more highly expressed at P7 compared to P1 or P14 in terms of either protein or transcript abundance (18), which is likely why CK2 has not been previously identified in large-scale transcriptional assays of cerebellar development and MB (18, 19). In keeping with this, many CK2 substrates did not change in phosphorylation abundance (table S1). Our findings therefore likely reflect regulated and localized change specifically in CK2 activity, which could have resulted from changes in kinase localization and/or trafficking of its subunits, co-regulator binding, phosphatase abundance, or accessibility of substrate motifs or docking sites (20). This highlights the importance of investigating kinase activity at its terminal readout: proteome-wide assessment of phosphorylation.

CK2 inhibition results in decreased GNP proliferation
Many peptides with CK2 consensus sites were more highly phosphorylated at the peak proliferative period (P7), suggesting that CK2 activity is important for GNP proliferation. We tested the role of CK2 in cerebellar development by injecting mice with the CK2 inhibitor 4,5,6,7-tetrabromobenzotriazole (TBB) (Fig. 3A) (21). Although several CK2 inhibitors exist, TBB was used because it is fairly selective (22), is cell-permeable (23), lacks short-term cytotoxicity in vivo (24), and is effective at preventing seizure activity in mice (24). Its potency suggests blood-brain barrier (BBB) permeability. Mice were treated twice a day, from P3 to P7, with TBB [30 mg/kg, intraperitoneally (ip)] (25) or vehicle control [dimethyl sulfoxide (DMSO)]. Four hours after the final injection, mice were sacrificed and their

cerebella were fixed. TBB-treated mice had no change in cerebellar folia length, which is affected by non–GNP-derived neurons (Fig. 3, B and C) (26). In contrast, folia width, which depends on the number of proliferating GNPs and their progeny, granule neurons, was significantly reduced by TBB treatment ($P < 0.01$; Fig. 3, B and C). The number of Atoh1-positive ($P < 0.01$) and proliferating GNPs at P7 ($P < 0.001$) was significantly reduced in treated mice (Fig. 3, B and D). No significant change in the rate of apoptosis or the ratio of proliferation to differentiation was detected (Fig. 3, B and D). Similarly, GNPs cultured in the presence of SHH and CX-4945, a CK2 inhibitor that is highly specific and structurally distinct from TBB, had a dose-dependent decrease in proliferation and a correlating decrease in total Atoh1-positive cells at 24 hours (fig. S3), as assessed by single-cell immunofluorescence imaging. The decrease in folia width was likely to be due to reduced GNP proliferation in response to inhibition of CK2 during the critical postnatal period.

CK2 facilitates Hh signal transduction
Because GNP proliferation in neonatal mice is driven by Hh signaling, we investigated whether CK2 affects the response to SHH. Hh signal transduction is initiated when SHH ligand binds and inhibits the cell surface receptor PTCH1 (Fig. 4A) (1, 27). SHH inhibition of PTCH1 unleashes the G protein (heterotrimeric GTP-binding protein)–coupled receptor SMO (28, 29), which then inhibits a protein called suppressor of fused (SUFU), in turn an inhibitor of the transcription factor GLI2 (30). Inhibition of SUFU thus allows transcription of GLI2 target genes, such as *GLI1*. Note that CK2 is an acronym derived from the misnomer "casein kinase-2" and is unrelated in both structure and function to the known GLI2 regulator CK1.

Genetic and small-molecule inhibitor studies in NIH3T3 cells (spontaneously immortalized murine embryonic fibroblasts) confirmed a role for CK2 in Hh signal transduction. NIH 3T3 cells are routinely used to assay the Hh pathway because they contain all canonical components of the Hh transduction pathway and have a robust transcriptional response to SHH ligand in culture. In contrast, GNPs in culture exit the cell cycle and rapidly differentiate into non–SHH-responsive granule neurons. A change in *Gli1* transcript expression in GNPs can result from perturbation of the Hh pathway or from non–Hh-dependent perturbations, such as cell cycle inhibitors or changes in ion homeostasis, so NIH3T3 cells are used instead.

CK2 is a heterotetrameric holoenzyme made up of four subunits: two catalytic α subunits (either two CSNK2A1 subunits or one CSNK2A1 subunit and one CSNK2A2 subunit) and two regulatory β subunits (CSNK2B). Genetic loss in mice of either CSNK2A1 or CSNK2B results in early embryonic death (31, 32). CSNK2A2 can generally be replaced by CSNK2A1, and knockout of *Csnk2a2* in mice results in infertility in males but otherwise has no apparent phenotypic consequence (31).

Small interfering RNA (siRNA)–mediated knockdown of the essential catalytic α subunit, CSNK2A1, in NIH3T3 cells strongly reduced the induction of *Gli1* expression by SHH (Fig. 4B). As expected,
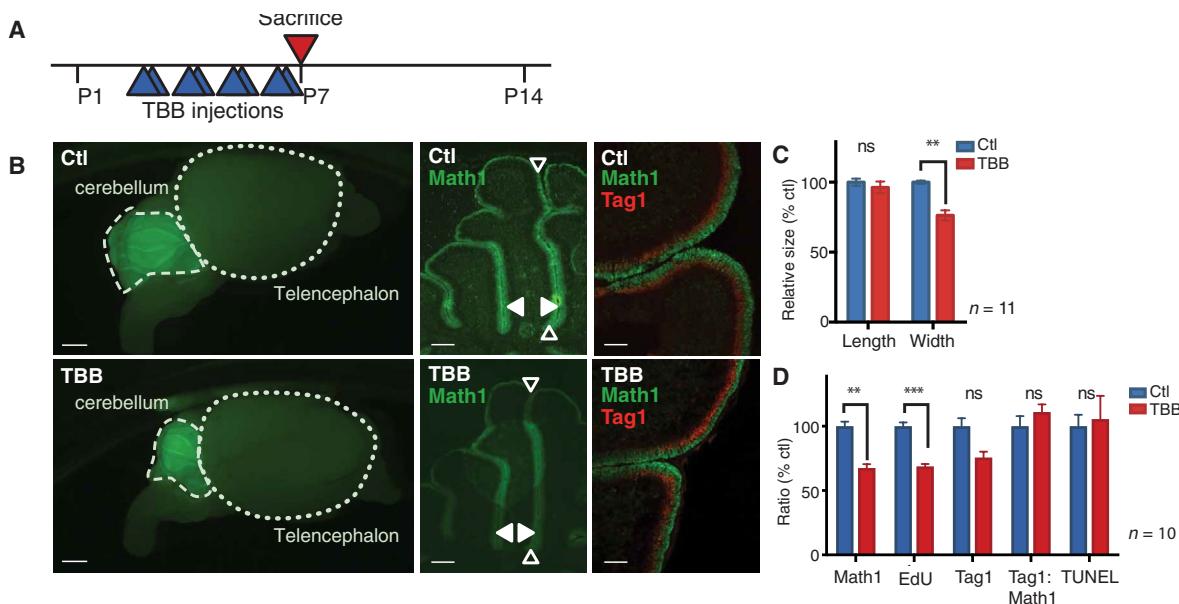
Fig. 3. CK2 is required for cerebellar development. (A) Experimental scheme for CK2 inhibitor (TBB) studies in early postnatal mice. **(B** and **C)** Immunofluorescent imaging of whole brains and sagittal cerebellar sections from control (DMSO) and CK2 inhibitor–treated *Math1/nGFP* transgenic mice. Green, Math1 (proliferating GNP marker); red, Tag1 (postmitotic GNP marker). Empty arrow: Folia length. Solid arrow: Folia width. **(D)** Ratio of Math1-expressing GNPs, early differentiating GNPs (Tag1), proliferation [5-ethynyl-2′-deoxyuridine (EdU)], and apoptosis [terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling (TUNEL)] in TBB-treated mice relative to control mice. Data are means ± SD from $n = 10$ to 11 mice, as indicated. ns = $P > 0.05$, **$P < 0.01$, ***$P < 0.001$, two-tailed t test. Scale bars, 1 mm (left), 0.1 mm (middle), and 0.05 mm (right).

the effect of knockdown of only the nonessential α prime subunit, CSNK2A2, on *Gli1* transcription was not significant. Knockdown of CSNK2B resulted in a significant but modest decrease in *Gli1* induction, consistent with previous evidence (*32*) that even low CSNK2B protein abundance, as in *Csnk2b* knockdown cells (fig. S4), is sufficient to make substantial amounts of CK2 holoenzyme. Each of the mRNA knockdowns was subunit-specific and did not reduce expression of the non-Hh target *Smo* (fig. S4).

Two structurally unrelated CK2 inhibitors, CX-4945 and TBB (*21, 33*), each significantly reduced the expression of the SHH-induced GLI2 targets (*Gli1*, *Gli2*, and *Ptch1*) in NIH3T3 cells as early as 4 hours after drug administration (Fig. 4C). CK2 inhibition did not decrease the transcript abundance of the non-Hh target gene *Smo* (fig. S5). The rapid influence of CK2 inhibitors on Hh pathway target gene expression suggests that the slower (72-hour) effects of CSNK2A1 knockdown in NIH3T3 cells were due to direct action on Hh signal transduction rather than indirect effects from prolonged CK2 inhibition.

CK2 facilitates Hh signal transduction at late steps in the pathway

When NIH3T3 cells were stimulated with an SMO agonist "SAG" (100 nM) (*34*), the CK2 inhibitor TBB blocked the increase in *Gli1* transcript that normally occurs in response to SAG, working as well as the SMO antagonist GDC-0449 (100 nM; also known as vismodegib) (Fig. 4D) (*35*). Therefore, CK2 regulates Hh signal transduction at or downstream of SMO. Treatment of *Sufu*$^{-/-}$ mouse embryonic fibroblasts (MEFs) (*36*) with TBB also strongly blocked transcription of *Gli1* (Fig. 4E), indicating that CK2 is needed at late steps of the Hh signal transduction pathway, at or downstream of SUFU. Treatment with the SMO antagonist GDC-0449 (100 nM) did not block transcription of *Gli1* in *Sufu*$^{-/-}$ cells, as expected, because SMO is upstream of SUFU in the pathway.

Our experiments demonstrating action of CK2 at a late step of the Hh pathway are consistent with previous work showing that the CK2 inhibitor TBB affects GLI2 protein stability, likely through regulation of GLI2 ubiquitination (*37, 38*). In keeping with this, TBB reduced the amount of endogenous full-length GLI2 expressed in NIH3T3 cells (Fig. 4F), as previously shown (*37*). TBB did not affect the abundance of an N-terminal–deleted form of Gli2 (HA-Gli2ΔN, hereafter simply Gli2ΔN; Fig. 4F), stably transfected into NIH3T3 cells (*39*). Gli2ΔN lacks sites required for ubiquitin-dependent degradation (*40*), remains constitutively nuclear, and normally causes constitutive induction of the GLI2 target gene *Gli1*. However, treatment with TBB still substantially lowered *Gli1* mRNA expression in Gli2ΔN-transfected cells (Fig. 4G). Together, the results suggest that CK2 regulates two steps in the Hh pathway: to stabilize GLI2 itself and to enhance the ability of GLI2 to activate its transcriptional targets. Our phosphoproteomic data point to several potential candidate substrates that might mediate effects of CK2 on GLI2-mediated transcription. For example, phosphorylated CK2 sites were found within enhancer-looping proteins [such as RING finger protein 1 (RING1), nipped-B–like protein (NIPBL), and mediator complex subunit 1 (MED1)] and epigenetic readers and writers [such as Polycomb repressive complex 1 (PRC1), PRC2, and histone deactylase (HDAC) components]. In two cases, the CK2 sites are well-established activating phosphorylation sites (in HDAC1 Ser$^{393/421}$ and HDAC2 Ser$^{294/422}$). CK2 may regulate accessibility of GLI2 target gene promoters or their interactions with enhancers.

CK2 inhibitors restrict the growth of mouse and human SHH MB cells in culture

SHH MBs occur because of sustained Hh target gene activity and require persistent action of the Hh signal transduction pathway for survival. Therefore, we reasoned that CK2 inhibitors may be effective
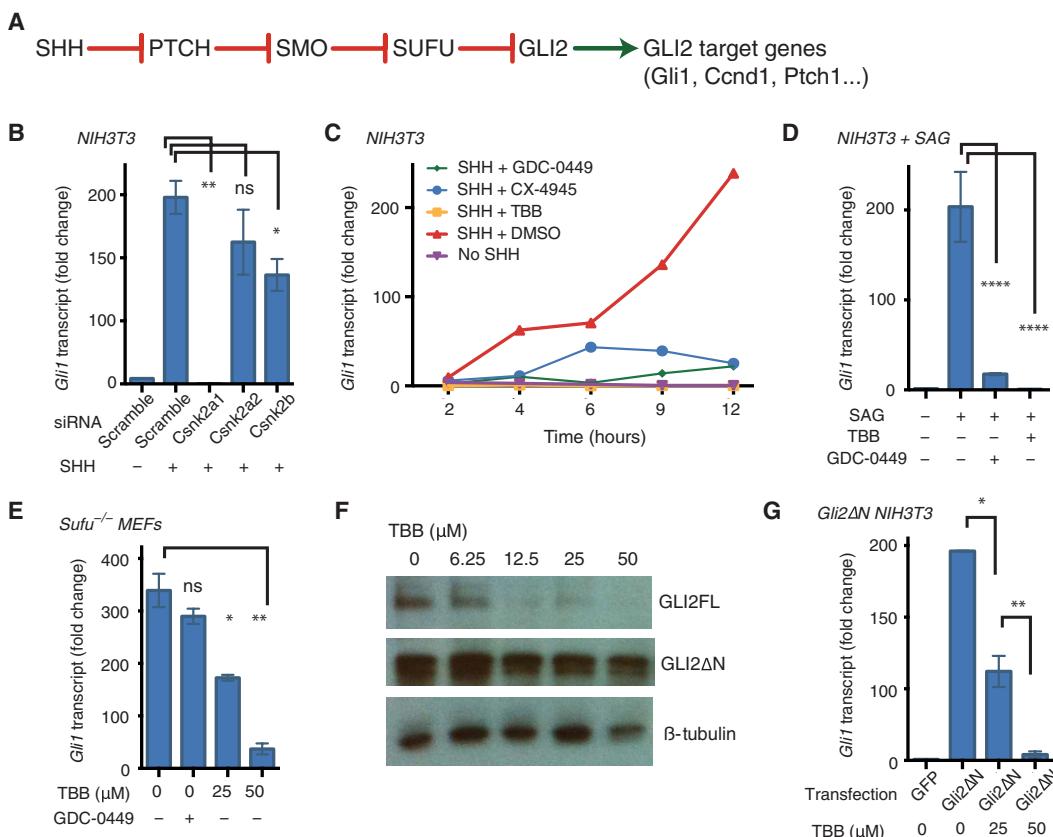
Fig. 4. CK2 is required for Hh signal transduction. (**A**) Canonical Hh signaling pathway, for reference in subsequent panels. (**B** and **C**) Fold change in *Gli1* mRNA expression in NIH3T3 cells exposed to murine SHH (3 μg/ml) for 6 hours after knockdown of CSNK2A1, CSNK2A2, or CSNK2B (B) or when treated with the CK2 inhibitor TBB (50 μM) or CX-4945 (10 μM) or SMO inhibitor GDC-0449 (100 nM) (C) relative to controls. (**D** and **E**) *Gli1* mRNA expression after exposure to TBB (as indicated) or GDC-0449 (100 nM) in NIH3T3 cells pretreated with SMO agonist SAG (100 nM) (D) and in *Sufu*$^{-/-}$ MEFs (E). (**F**) Immunoblotting for protein abundance of endogenous GLI2 (GLI2FL, full length) and exogenous constitutively active GLI2 (GLI2ΔN) in NIH3T3 cells treated with various concentrations of TBB for 6 hours. Blots are from a representative of three experiments. (**G**) *Gli1* mRNA expression after exposure to TBB in NIH3T3 cells expressing GLI2ΔN. Data in (A) to (D) and (G) were obtained by quantitative reverse transcription polymerase chain reaction (qRT-PCR) and are shown as means ± SD from *n* = 3 experiments. ns = *P* > 0.05, *P* < 0.05, **P* < 0.01, ****P* < 0.0001, two-tailed *t* test.

at restricting the growth of this MB subtype. Unfortunately, most SHH MB cell lines suppress Hh pathway activity when propagated in vitro and are therefore not useful models of MB. Several previously established SHH MB cell lines that we tested had suppressed SHH pathway activity as indicated by low expression of the Hh target *Gli1* (fig. S6). After careful screening, we chose to use the cell lines MB21, MB53, and MB55 (*41*) in further tests. These murine MB lines retain key characteristics of primary MB tumors that arise in *Ptch*$^{+/-}$ mice, including sustained, high amounts of *Gli1* transcription that is inhibited by SMO inhibitors both in our hands (fig. S6) and in published accounts (*41, 42*). Hh signaling activity and expression of Hh pathway components in these cell lines are in keeping with primary mouse SHH MB tumors, and when cells are implanted into the flank or cerebellum of mice, they grow into tumors with SHH-subtype histology (*41*). Gene expression patterns also strongly correlate with both primary mouse SHH MB cells and human SHH MBs, but not human non-SHH subgroup MBs (WNT, group 3, or group 4) (*41*).

Treatment of MB21, MB53, and MB55 *Ptch*$^{+/-}$ mouse MB cell lines in culture with the CK2 inhibitor CX-4945 reduced *Gli1* expression after 6 hours (fig. S6), suggesting inhibition of Hh pathway activity. The cells underwent dose-dependent cell death after 3 days of treatment [median inhibitory concentration (IC$_{50}$) = 2.5 to 5.3 μM,

relative, normalized to control (DMSO)–treated cells] (Fig. 5A). Three human SHH MB patient-derived xenograft (PDX) lines passaged only in mouse cerebellum were assayed in vitro and similarly demonstrated dose-dependent cell death after 3 days of treatment (Fig. 5B). A human SHH MB tumor cell line with the same activating mutation as the mouse MB cells (RCMB32, *Ptch*$^{+/-}$) had even greater sensitivity to CX-4945 (IC$_{50}$ = 0.76 μM) than MB21, MB53, or MB55. Two other human SHH MB cell lines have *TP53* mutations and SHH pathway–activating mutations downstream of SMO, so they are resistant to SMO inhibitors [ICb-984, IC$_{50}$ = 3.3 μM (*GLI2;MYCN*-amplified; *TP53*$^{-/-}$; chromothripsis); BT084, IC$_{50}$ = 1.8 μM (*GLI2;MYCN;TERT*-amplified; *TP53*$^{-/-}$; chromothripsis)] (*43, 44*). These cell lines represent some of the most aggressive forms of human SHH MB. Primary human SHH MB cells (ST01), obtained directly from the operating room and confirmed by pathology to be of the SHH subgroup, also had a dose-dependent decrease in cell viability when treated with CX-4945 for 3 days (IC$_{50}$ = 1.4 μM; Fig. 5, B and C). As expected, treatment with the SMO inhibitor vismodegib also reduced cell viability and at lower concentrations than CX-4945 (100 nM versus 10 μM), reflecting the favorable ligand abundance, binding accessibility, and/or drug-substrate interactions compared to CX-4945. Notably, cells in G$_0$, which are commonly resistant to chemotherapy and are a cause
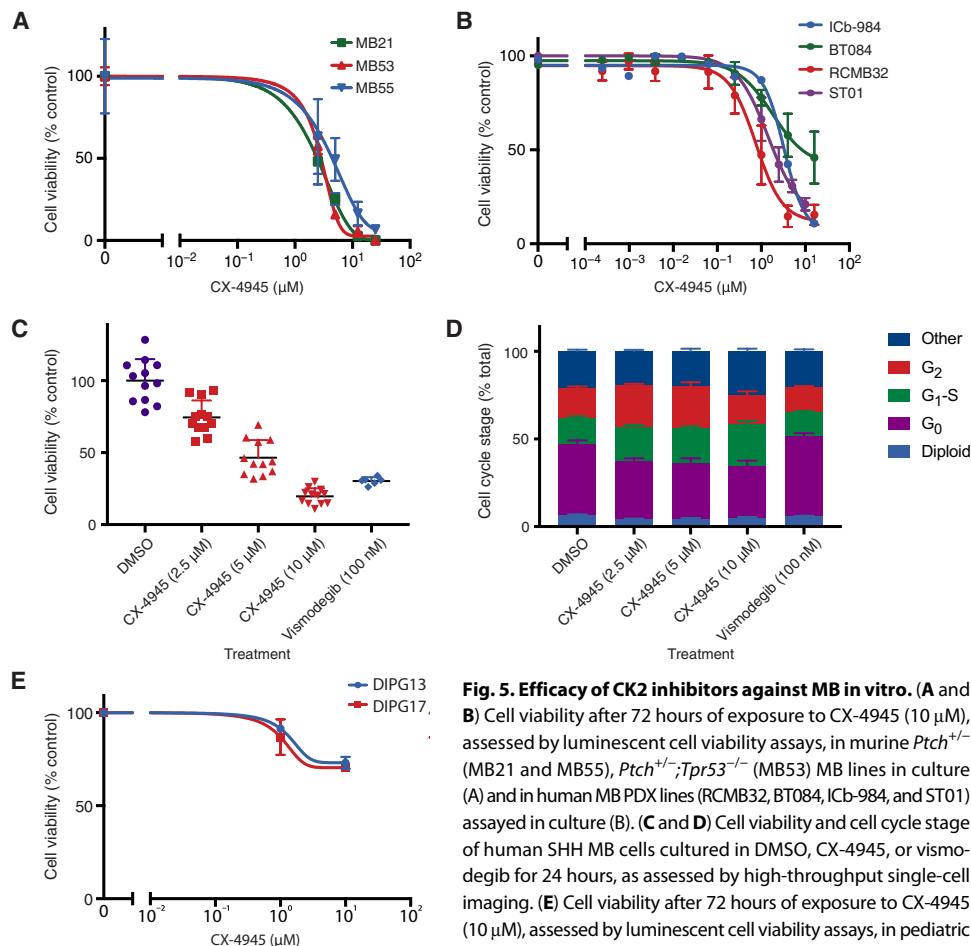
Fig. 5. Efficacy of CK2 inhibitors against MB in vitro. (**A** and **B**) Cell viability after 72 hours of exposure to CX-4945 (10 μM), assessed by luminescent cell viability assays, in murine *Ptch*$^{+/-}$ (MB21 and MB55), *Ptch*$^{+/-}$;*Tpr53*$^{-/-}$ (MB53) MB lines in culture (A) and in human MB PDX lines (RCMB32, BT084, ICb-984, and ST01) assayed in culture (B). (**C** and **D**) Cell viability and cell cycle stage of human SHH MB cells cultured in DMSO, CX-4945, or vismodegib for 24 hours, as assessed by high-throughput single-cell imaging. (**E**) Cell viability after 72 hours of exposure to CX-4945 (10 μM), assessed by luminescent cell viability assays, in pediatric glioma (DIPG) cells. Data are means ± SD from *n* = 3 experiments.

mors, near-complete cessation of tumor growth (Fig. 6, B to D).

CK2 inhibition prolonged survival in mice with cerebellar MB. Eighty thousand *Ptch*$^{+/-}$;*Tpr53*$^{-/-}$;SmoD477G mouse MB cells per mouse were injected into cerebella of NSG mice (Fig. 6A). Seventy-two hours after tumor cells were injected, a collaborator who was blinded to treatment randomized the mice into two groups and began administering either the CK2 inhibitor CX-4945 (37.5 mg/kg) or vehicle control (DMSO). CX-4945 was used because it is the only CK2 inhibitor optimized for clinical use, is provided orally, and has been shown to be safe in humans with solid tumors (*50*). In agreement with previous trials using these cells (*51*), 100% of control mice died because of tumors by day 17; in contrast, 43% of mice treated with CX-4945 survived past 100 days, despite termination of treatment at day 30 (Fig. 6E). CX-4945 was well tolerated, because body weights were similar between cohorts throughout the duration of treatment (fig. S7).

CK2 expression correlates with 5-year survival in patients with SHH MB

Human clinical data were consistent with the idea of using CK2 inhibition to treat MB. Kaplan-Meier survival analysis has revealed that patients with SHH-MB who

of SHH MB recurrence (*45*), were the most sensitive cell population to treatment with CX-4945 (11.1% relative decrease in G$_0$ cells) but the least sensitive population to treatment with vismodegib (5.3% relative increase in G$_0$ cells; Fig. 5D). Control cell lines derived from diffuse intrinsic pontine glioma (DIPG), another aggressive pediatric brain tumor that is not Hh pathway–dependent, did not die in response to CX-4945 (Fig. 5E).

CK2 inhibitors are effective against SHH MB models in vivo

Ptch$^{+/-}$;*Tpr53*$^{-/-}$;*SmoD477G* mouse MB cells, which have a mutation in SMO that causes resistance to the U.S. Food and Drug Administration (FDA)–approved Hh (Smo) inhibitor vismodegib (GDC-0449), were injected into the flanks of mice, and the resulting tumors were allowed to grow to 250 to 300 mm³ (Fig. 6A). A parallel experiment was done using implants of *Ptch*$^{+/-}$;*Tpr53*$^{-/-}$ MB cells, which lack the activating SMO mutation and are therefore susceptible to vismodegib. Preclinical work for the two FDA-approved Hh inhibitors (GDC-0449 and LDE225) relied on mouse flank allografts using the same cell lines as those used here (*40, 46, 47*). These mouse cells reportedly closely resemble human MB cells with respect to gene expression (*41*), response to treatment (*40, 47–49*), and even the spontaneous occurrence of treatment-induced, drug-resistant mutations (*35, 46*). Treatment of mice harboring flank allografts with the CK2 inhibitor TBB [30 mg/kg, ip twice daily (BID)] significantly inhibited MB growth as early as 2 days after initiating treatment and, in vismodegib-resistant tu-

had low expression of mRNA encoding the main CK2 catalytic subunit α (*CSNK2A1*) in their tumors had a mean 82% 5-year survival, whereas patients with high expression of *CSNK2A1* had a mean 61% 5-year survival (Fig. 6F). No correlation between *CSNK2A1* expression and patient survival was observed in other MB subgroups (group 3, group 4, and WNT; fig. S8), which have been previously shown to not require Hh pathway activity for tumor growth (*3*).

Single mutation in CK2 causes resistance to TBB but not CX-4945

To test the potential for rapid emergence of resistance to CK2 inhibitors, four *Ptch*$^{+/-}$ mouse MB cell lines (MB21, MB53, MB55, and MB56) were serially passaged for 1 month in the presence of TBB (50 μM), CX-4945 (10 μM), or a DMSO control. During this period, one line (MB55) evolved resistance to TBB, although it maintained sensitivity to CX-4945 (Fig. 7A). Sequencing of *Csnk2a1* from the TBB-resistant MB55 cell line revealed five missense mutations (fig. S9A), leading to amino acid substitutions [D175N (fig. S9B), R47M, H183R, G199P, and S217N] that were not present in any of the other cell lines, including the parental TBB-sensitive MB55 line.

Molecular dynamic (MD) simulations and molecular docking experiments were performed and revealed that, of these five mutations, D175N was the most likely candidate for the reduced efficacy of TBB. In the crystal structure of CSNK2A1 (*52*), alone and with the inhibitor CX-4945 (*33*) or TBB (*53*), Asp^{175} is located in the well-known and highly
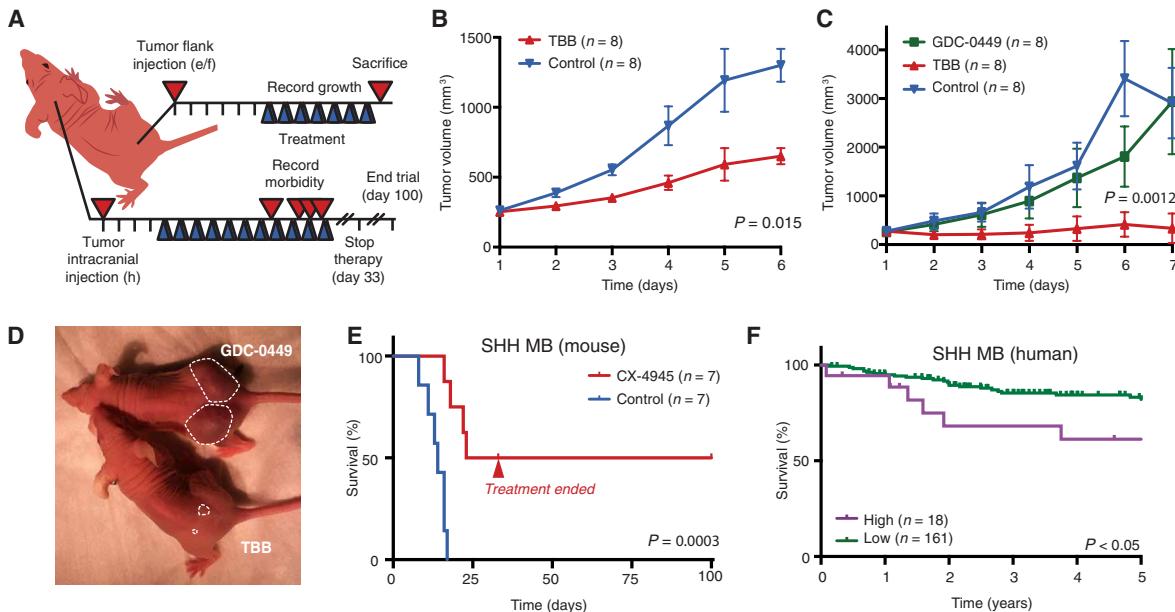
Fig. 6. Efficacy of CK2 inhibitors against MB in vivo. (A) Experimental scheme for flank and cerebellar tumor trials. Blue triangle, treatment. (**B** and **C**) Relative tumor growth of $Ptch^{+/-};Tpr53^{-/-}$ (B) and $Ptch^{+/-};Tpr53^{-/-};SmoD477G$ (C) MB flank allografts. Data are means ± SD; P values by a two-tailed t test. (**D**) Representative picture of mice treated with TBB versus GDC-0449. Dashed line demarcates the tumor. (**E**) Kaplan-Meier survival analysis of mice with $Ptch^{+/-};Tpr53^{-/-};SmoD477G$ MB cerebellar allografts treated with CX-4945 or vehicle control. $n = 7$ mice in each treatment group. (**F**) Kaplan-Meier analysis of CK2 gene expression and overall survival in human patients ($n = 179$) with SHH-subtype MB. P values in (E) and (F) were determined by a log-rank Mantel-Cox test.

conserved DFG motif (DWG in CK2) located in the catalytic binding cleft of CK2α. Together with Glu[81], Asp[175] is important for the correct orientation of the catalytic residue Lys[68] through salt bridges. Our MD simulations revealed that an asparagine in the 175 position was unable to form a salt bridge with catalytic residue Lys[68]. Consequently, Lys[68] strengthened the remaining salt bridge, and Asn[175] moved toward Glu[81] to protect itself from the hydrophobicity of the surrounding residues (especially Ile[174]). The resulting displacement of Lys[68] and Asn[175] increased the ATP-binding cavity volume by about 137 Å. In wild-type CK2, the binding of TBB requires interaction with both the hinge region and Lys[68] (53). Our substrate-specific docking experiments revealed that, when TBB was inside the enlarged cavity of the D175H mutant protein, TBB was too small to efficiently interact with both the hinge region and Lys[68] (Fig. 7B). In contrast, CX4945 spanned the entire catalytic site and maintained the same crystallographic binding pattern as in the wild-type protein (Fig. 7C), anchoring both the hinge region at 2.94 Å and to Lys[68] at 2.80 Å through hydrophilic interactions. These data indicate that the efficacy of CX-4945 in blocking growth of the MB cell lines—with wild-type and mutant CK2 alike— may be through direct binding–mediated inhibition of CK2.

DISCUSSION

By screening the phosphoproteome during developmental periods when GNPs switch from proliferation to differentiation, we identified CK2 as having increased activity on multiple substrates, specifically in the proliferative, MB-like, P7 stage. The work presented here has already directly led to a phase 1/2 study investigating the use of the CK2 inhibitor CX-4945 in patients with SHH MB. The rapid translation from a phosphoproteomic screen to a clinical trial reflects favorable aspects of the approach to cancer drug target identi-

fication. Traditional approaches to discovering novel drug targets often rely on identification of mutations that appear within tumors. However, tumor cells carrying any mutation that is advantageous to cell growth and survival will come to predominate, even if that advantage is small. A great challenge is to determine the biological function of each mutated protein in the context of any one tumor and then distinguishing proteins with activity only slightly advantageous to tumor growth from those that are essential to survival of the tumor. Drugs that target the former have only incremental benefit to patient survival, whereas the latter would be potentially clinically transformative. In contrast, a better understanding of proteins that regulate proliferation versus differentiation in the developmental lineage from which the tumor arose, as provided by our developmental phosphoproteomic approach, can identify important tumor drivers and their likely biological function in proliferation. The critical role of CK2 in GNP development may be the reason why CK2 inhibition resulted in robust tumor stasis and long-term regression of SHH MB but had less robust, often transient, benefit in many other cancer types (54, 55).

Another reason for the rapid path to clinic was the identification of a drug target that is readily conducive to small-molecule inhibition. Kinases, as would be identified through our phosphoproteomic approach, are often readily inhibited by small molecules that compete with ATP binding in the activation cleft. CK2 had many commercially available small-molecule inhibitors and a human-ready compound. This allowed us to bypass lengthy and costly drug development work as would be required for many other protein classes that have undesirable features of a drug target (presence of metal ions, lack of hydrogen bond donors and acceptors, need for adaptive changes in conformation, lipophilicity, and featureless binding sites) (56).

The many potential roles of CK2 in MB, and its influence on two late steps in SHH signal transduction, make it unlikely that a mutation
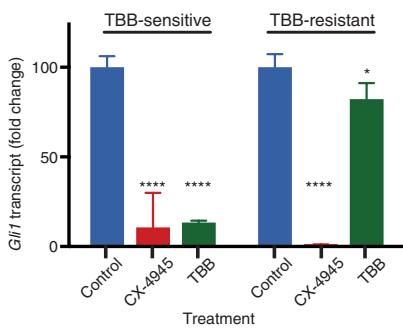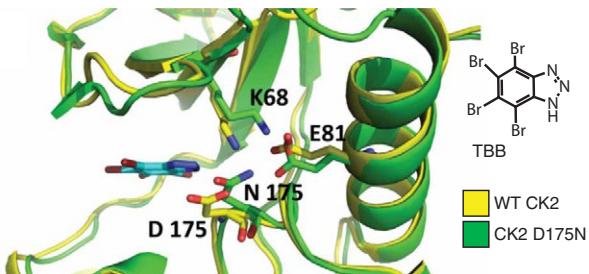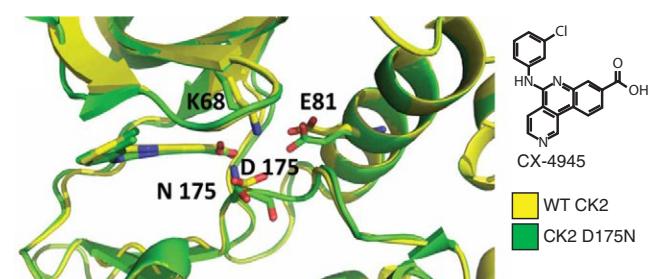
A

B

C

Fig. 7. TBB resistance results from mutation in *Csnk2a1*. (A) *Gli1* transcript abundance in parental, TBB-sensitive MB55 cells and serial passage/drug-induced, TBB-resistant MB55 cells treated with CX-4945 or TBB relative to that in control (DMSO-treated) cells. Data are means ± SD from *n* = 3 experiments. ns = *P* > 0.05, **P* < 0.05, ****P < 0.0001, two-tailed *t* test. **(B** and **C)** MD simulation and docking analysis of TBB (B) and CX-4945 (C) with wild-type (WT) CK2 and mutant (D175N) CK2 in MB55 cells.

in any one substrate will result in drug resistance, which is another common limitation of many targeted therapies. As we demonstrated, mutations within CK2 itself can still cause resistance. Yet, unlike off-target resistance resulting from mutation of any one of many proteins, predictable resistance due to a mutation in CK2 itself can be tackled through combinations of CK2 inhibitors and/or rational drug design. The quantitative developmental phosphoproteomics approach that we used enabled identification of a previously unknown regulatory mechanism of the Hh signaling pathway and a potential new treatment for this devastating pediatric cancer. Similar approaches may work for other cancers where the developmental cell of origin can be found.

MATERIALS AND METHODS
Animal experiments: Statistics, blinding, and randomization
Given that CK2 promotes Hh signal transduction, we estimated sample size on the basis of previous work performed with Hh inhibitors.

Given the dramatic SHH MB tumor regression previously demonstrated with the Hh inhibitor vismodegib (*57*), we conservatively predicted that CK2 inhibitors would have an effect size of at least 2.2. Therefore, if α error = 0.05 and power = 0.85, then five animals are required per condition. To allow for potential exclusion of one to two animals per condition because of poor tumor engraftment, we used seven to eight animals per condition. Ultimately, we had 100% engraftment [and which met the inclusion criteria (tumor = 250 to 300 mm^3)]; thus, all animals were included in the analyses of the tumor growth experiments. Mice were randomized into control or treatment groups daily. Daily randomization ensured that quicker growing tumors were equally represented in both groups. For cerebellar tumor studies, mice were randomized by a blinded collaborator 72 hours after implantation. A blinded collaborator treated and recorded weight, morbidity, and survival. Prism statistical analysis software was used (GraphPad Software) to calculate significance as determined by two-tailed *t* tests for each paired data point (Fig. 6, B and C) or log-rank Mantel-Cox test (Fig. 6E) for survival studies. All studies were performed in compliance with institutional regulations.

A previous study investigating the effect of Hh inhibition in the developing cerebellum showed that mice with inactive Hh signaling have about 50% reduction in the EGL size, with an effect size of about 5.7 (control: about 80, SD = 7; treatment: 40, SD = 7) (*6*). Therefore, we conservatively predicted that CK2 inhibitors would be at least 25% as potent as Hh inhibitors, giving an effect size of 1.4. Therefore, with α error = 0.05 and power = 0.85, we expected that we would need and therefore used 10 animals per condition for the developmental experiments. Mice were randomized to treatment or control group at P3, with each litter being equally divided between the two groups. Images were randomized, and all measurements (length, width, Atoh-1, Tag1, and TUNEL staining) were performed by a blinded collaborator. Cerebellar folia length was compared between matched sections of control and treated mice by measuring the maximal length from the mid-base to the tip of the culmen (lobule IV/V). Culmen width was measured 600 μm from the base of the folia between lobules V and VI. Prism statistical analysis software was used (GraphPad Software) to calculate significance as determined by two-tailed *t* tests. For the Hh epistasis assays specifically, all measurements were taken in biological triplicate, and Prism statistical analysis software was used (GraphPad Software) to calculate significance as determined by two-tailed *t* tests.

Human Kaplan-Meier curves
Kaplan-Meier survival curves on the human MB data sets were generated using GraphPad Prism, and significance was measured by log-rank Mantel-Cox test.

GNP isolation
Wild-type CD1-Elite mice (strain 482, Charles River) were sacrificed at P1, P7, and P14. Thirty-five to 40 mice were used per P1 and P14 biological replicate, and 14 to 17 mice were used per P7 biological replicate. Cerebella were dissected, and GNPs were isolated and purified using a Percoll gradient, as previously described (*58*, *59*). This approach results in a population that is 95 to 99% GNPs (*58*). Briefly, cerebella were minced and then placed in a digestion buffer consisting of 1× Hanks' balanced salt solution (HBSS) (14185, Thermo Fisher Scientific), papain (10 U/ml; LSOO3126, Worthington Biochemical), and deoxyribonuclease (DNase) (250 U/ml; D4627, Sigma). Papain solution was aspirated and replaced with 1× HBSS containing ovomucoid

(8 mg/ml; LK003182, Worthington Biochemical), bovine serum albumin (BSA) (8 mg/ml; Sigma), and DNase (250 U/ml) and was then triturated using a Pasteur pipette to obtain a single-cell suspension. Cells were passed through a 70-μm nylon cell strainer (21008-952, VWR International) and purified using a step gradient of 35 and 65% Percoll (P4937, Sigma). Pellets were flash-frozen in liquid nitrogen and kept at −80°C until all samples were collected for MS.

Sample preparation for quantitative MS analysis

GNP samples were prepared as previously described (60), with the following modifications. All solutions are reported as final concentrations. Lysis buffer [8 M urea, 1% SDS, 50 mM tris (pH 8.5), protease and phosphatase inhibitors from Roche] was added to the cell pellets to achieve a cell lysate with a protein concentration between 2 and 8 mg/ml. A micro-BCA (bicinchoninic acid) assay (Pierce) was used to determine the final protein concentration in the cell lysate. Proteins were reduced and alkylated as previously described. Proteins were precipitated using methanol/chloroform. In brief, four volumes of methanol were added to the cell lysate, followed by one volume of chloroform, and finally three volumes of water. The mixture was vortexed and centrifuged to separate the chloroform phase from the aqueous phase. The precipitated protein was washed with one volume of ice-cold methanol. The washed precipitated protein was allowed to air dry. Precipitated protein was resuspended in 4 M urea, 50 mM tris (pH 8.5). Proteins were first digested with LysC (1:50; enzyme/protein) for 12 hours at 25°C. The LysC digestion was diluted down to 1 M urea, 50 mM tris (pH 8.5) and then digested with trypsin (1:100; enzyme/protein) for another 8 hours at 25°C. Peptides were desalted using a C_{18} solid-phase extraction cartridges, as previously described. Dried peptides were resuspended in 200 mM EPPS [4-(2-hydroxyethyl)piperazine-1-propanesulfonic acid] (pH 8.0). Peptide quantification was performed using the micro-BCA assay (Pierce). The same amount of peptide from each condition was labeled with tandem mass tag (TMT) reagent (1:4; peptide/TMT label) (Pierce). The 6- and 10-plex labeling reactions were performed for 2 hours at 25°C. Modification of tyrosine residue with TMT was reversed by the addition of 5% hydroxyl amine for 15 min at 25°C. The reaction was quenched with 0.5% trifluoroacetic acid, and samples were combined at a 1:1 ratio. Combined samples were desalted and offline-fractionated into 24 fractions, as previously described.

Liquid chromatography–MS3 spectrometry

Twelve fractions (every other one) of the 24-peptide fraction from the basic reverse-phase step were analyzed with a liquid chromatography (LC)–MS3 data collection strategy on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) equipped with a Proxeon EASY-nLC 1000 for online sample handling and peptide separations. About 5 μg of peptide resuspended in 5% formic acid + 5% acetonitrile (ACN) was loaded onto a 100–μm–inner diameter fused-silica microcapillary with a needle tip pulled to an internal diameter less than 5 μm. The column was packed in-house to a length of 35 cm with a C_{18} reverse-phase resin (GP118 resin 1.8 μm, 120 Å; Sepax Technologies). The peptides were separated using a 120-min linear gradient from 3 to 25% buffer B (100% ACN + 0.125% formic acid) equilibrated with buffer A (3% ACN + 0.125% formic acid) at a flow rate of 600 nl/min across the column. The scan sequence for the Orbitrap Fusion began with an MS1 spectrum [Orbitrap analysis, resolution 120,000 (400 to 1400 m/z scan range); AGC target, 2×10^5; maximum injection time, 100 ms (dynamic exclusion of 75 s)].

"Top speed" (1 s) was selected for MS2 analysis, which consisted of collision-induced dissociation [quadrupole isolation set at 0.5 Da and ion trap analysis; AGC, 4×10^3; normalized collision energy (NCE), 35; maximum injection time, 150 ms]. The top 10 precursors from each MS2 scan were selected for MS3 analysis (synchronous precursor selection), in which precursors were fragmented by high-energy collisional dissociation before Orbitrap analysis [NCE, 55; maximum AGC, 5×10^4; maximum injection time, 150 ms; isolation window, 2.5 Da; resolution, 60,000 (10-plex experiments) or 15,000 (6-plex experiments)].

LC-MS3 data analysis

A suite of in-house software tools was used for .RAW file processing and controlling peptide and protein-level FDRs, assembling proteins from peptides, and protein quantification from peptides, as previously described. MS/MS spectra were searched against a UniProt human database (accessed February 2014), with both the forward and reverse sequences. Database search criteria are as follows: tryptic with two missed cleavages, a precursor mass tolerance of 50 parts per million (ppm), fragment ion mass tolerance of 1.0 Da, static alkylation of cysteine (57.02146 Da), static TMT labeling of lysine residues and N termini of peptides (229.162932 Da), and variable oxidation of methionine (15.99491 Da). TMT reporter ion intensities were measured using a 0.03-Da window (6-plex) or 0.003-Da window (10-plex) around the theoretical mass/charge ratio (m/z) for each reporter ion in the MS3 scan. Peptide spectral matches with poor-quality MS3 spectra were excluded from quantitation (<100 summed signal-to-noise ratio across 6 channels and <0.5 precursor isolation specificity for 6-plexes or <200 summed signal-to-noise ratio across 10 channels and <0.5 precursor isolation specificity for 10-plexes).

LC-MS3 sample comparison

Only peptides identified with full confidence were included in the analysis (binomial probability threshold of $P < 10^{-6}$, occurrence threshold = 20). Each biological replicate was median-adjusted to 10 to account for differences in sample loading, resulting in the following changes: proteomic data set: P1A, −0.0362; P1B, −0.0879; P1C, +0.20476; P7A, +0.03374; P7B, −0.2761; P7C, −0.2865; P14A, +0.78605; P14B, +0.65244; P14C, +0.71771; MB, +0.04692; phosphoproteomic data set: P1A, −0.040163003; P1B, −0.124877212; P1C, −0.009654128; P7A, −0.170361415; P7B, −0.341821765; P7C, −0.117111143; P14A, +1.114888992; P14B, +1.060676918; P14C, +0.955615374; MB, +0.43299154. Log2-transformed P1:P7 and P7:P14 ratios were determined, and phosphopeptide changes were normalized to protein changes. Values outside of 1.5 SD were considered "significant changers." Motif analysis of significantly changing phosphopeptides was performed using Motif-X, as previously described (61, 62). Briefly, Motif-X is an iterative statistical approach to identifying protein phosphorylation motifs in large-scale phosphoproteomic data sets built on a greedy recursive search of the sequence space to identify highly correlated residue/position pairs with the lowest P values. Here, a binomial probability threshold of $P < 10^{-6}$ and occurrence threshold of 20 were used.

GNP proliferation

To determine relative proliferation of Atoh1(+) GNPs at P1, P7, and P14, cerebella were dissected from *Atoh1/nGFP* transgenic mice (63) 1 hour after EdU injection [50 mg/kg using a stock (5 mg/ml) diluted in phosphate-buffered saline (PBS)]. Brains were dissected and fixed in 4% paraformaldehyde (PFA) overnight and then transferred into

20% sucrose for 24 hours. Fixed whole cerebella were mounted in optimal cutting temperature (OCT) embedding medium, sectioned at 20 μm, and then air-dried for 20 min before storing at −20°C for up to 2 months. Sections were blocked in 0.2% Triton X-100 and 2% goat serum diluted in sterile PBS for 1 hour at room temperature. EdU staining was performed as per the manufacturer's instructions (Click-iT Plus EdU Alexa Fluor 647 Imaging Kit, catalog no. C10640, Life Technologies) and counterstained with Hoechst 33258 (final concentration, 1 μl/ml in PBS) at room temperature for 10 min. Images were taken on a Leica TCS SP8 confocal microscope. Total EdU signal within Atoh1(+) regions was quantified using ImageJ (six cerebella per time point, four to seven folia per cerebellar section) after background subtraction via rolling ball radius (50 pixels). Prism statistical analysis software was used (GraphPad Software) to calculate significance as determined by two-tailed t tests.

GNP culture
GNPs from CD1-Elite mice (strain 482, Charles River) were cultured in Neurobasal-A with B27, sodium pyruvate, penicillin/streptomycin, Glutamax, glucose (0.6 mg/ml), and KCl (25 mM). Recombinant SHH (SHH 461-54/CF, R&D Systems) was added to media at 3 μg/ml to promote cell division.

Single-cell immunofluorescence imaging (GNP and MB)
Before cell culture, 96-well glass imaging plates (Cellvis P96-1.5H-N) were incubated with poly-D-lysine (A-003-E, Millipore) at 100 μg/ml for 2 hours followed by laminin (CC095, Millipore) at 10 μg/ml overnight at 37°C. GNPs or MB cells were plated at 1×10^5 cells per well and grown at 37°C. After treatment, the cells were fixed with 4% PFA for 10 min at room temperature. The cells were then blocked using 5% donkey serum, 1% BSA, and 0.2% Triton X-100 for 1 hour at room temperature. Primary antibodies (Abs) were incubated overnight at 4°C. The primary Abs used are as follows: anti-NeuN mouse monoclonal at 1:200 (MAB377, Millipore) and phospho-Rb (Ser$^{807/811}$) rabbit monoclonal at 1:1000 (clone D20B12, 8516, Cell Signaling Technology). Cells were costained with rhodamine phalloidin (R415, Molecular Probes). Donkey anti–immunoglobulin G (IgG) secondary Abs against mouse and rabbit conjugated to Alexa Fluor 488 and Alexa Fluor 647 at 1:500 (Jackson ImmunoResearch) were used. All cell imaging was performed using the ImageXpress Micro XLS Widefield High Content Screening System (Molecular Devices) using 20× (0.45 or 0.75 numerical aperture) Nikon objectives. The intensity of fluorescence in each cell was automatically calculated using custom MATLAB scripts: Nuclei were segmented using 4′,6-diamidino-2-phenylindole, as previously described (*64*); downstream analysis was performed in R; cell cycle classification was performed using Gaussian finite mixture modeling with the mclust package (*65*); and cell cycle clustering analysis in GNPs used the $\log_{10}$-transformed median intensity of each cell, measuring Math1-GFP, phosphorylated Rb, and NeuN, as well as the area of the cell. All studies were done with six individual cultures.

CK2 inhibition in the developing mouse
Math1/nGFP transgenic mice (*63*) were treated with DMSO or TBB (30 mg/kg, ip BID) (218697, EMD Millipore) from P4 to P7 and then sacrificed 6 hours after their final treatment. Brains were dissected and immediately homogenized in lysis buffer (sc-24948, Santa Cruz Biotechnology) supplemented with protease and phosphatase inhibitors (no. 1861218, Thermo Fisher Scientific) or fixed in 4% PFA overnight and then transferred into 20% sucrose for 24 hours. Lysates

were separated on a 12.5% SDS–polyacrylamide gel electrophoresis (PAGE) gel and immunoblotted with pAKT:S129 (no. 13461, Cell Signaling Technology) and AKT (no. 4685, Cell Signaling Technology). Fixed whole cerebella were mounted in OCT, sectioned at 20 μm, and then air-dried for 20 min before storing at −20°C for up to 2 months. Sections were blocked in 0.2% Triton X-100 and 2% goat serum diluted in sterile PBS for 1 hour at room temperature. The primary Abs used were anti-NeuN Ab (1:200, COT: 2459079, EMD Millipore), anti-Tag1 Ab (1:100, Ab derived in the Scott laboratory), TUNEL staining as per the manufacturer's instructions (In Situ Cell Death Detection Kit, catalog no. 1 684 809, Roche), and EdU staining as per the manufacturer's instructions (Click-iT Plus EdU Alexa Fluor 647 Imaging Kit, catalog no. C10640, Life Technologies). All sections were incubated in primary Ab overnight at 4°C and counterstained with Hoechst 33258 (final concentration, 1 μl/ml in PBS) at room temperature for 10 min. Images were taken on a Leica TCS SP8 confocal microscope. Images were randomized, and all measurements (length, width, Atoh-1, Tag1, and TUNEL staining) were performed by a blinded collaborator. For each group, two images were taken and measured for each of 10 cerebella. Cerebellar folia length was compared between matched sections of control and treated mice by measuring the maximal length from the mid-base to the tip of the culmen (lobule IV/V). Culmen width was measured (600 μm) from the base of the folia between lobules V and VI. One section was measured from each of 11 cerebella analyzed. Prism statistical analysis software (GraphPad Software) was used to calculate significance as determined by two-tailed t tests.

siRNA studies
Pooled custom Dharmacon ON-TARGETplus specificity enhanced siRNA was created against *Csnk2a1*, *Csnk2a2*, and *Csnk2b*. siRNA sequences were provided by D. Seldin (Boston University School of Medicine) and are listed in table S3. Scramble siRNA was purchased from Dharmacon (D-001810-01-05). NIH3T3 cells or *Sufu*$^{−/−}$ MEFs were plated in six-well plates at 400,000 cells per well in 10% fetal bovine serum (FBS) media constituted of Dulbecco's modified Eagle's medium (DMEM) (11960-069, Invitrogen) supplemented with 10% FBS (SH30070.02, Hyclone Laboratories), penicillin/streptomycin (15140-122, Life Technologies), sodium pyruvate (11360-070, Invitrogen), and Glutamax (35050-061, Invitrogen). Cells were then starved in 0.5% FBS for 24 hours before treatment with 25% SHH-N conditioned media in 0.5% FBS. SHH-N conditioned medium was made using a human embryonic kidney (HEK) 293 cell line that stably secretes Shh-N, as previously described (*34*).

Hh pathway epistasis studies
Sufu$^{−/−}$ MEFs [conditional deletion of exons 4 to 8 (*36*)] were provided by P.-T. Chang [University of California, San Francisco (UCSF)]. As in siRNA experiments, NIH3T3 cells or *Sufu*$^{−/−}$ MEFs were plated in six-well plates at 400,000 cells per well in 10% FBS media (DMEM supplemented with 10% FBS, penicillin/streptomycin, sodium pyruvate, and Glutamax). Cells were then starved in 0.5% FBS for 24 hours before treatment with 25% SHH-N conditioned media in 0.5% FBS ± drug treatment or SAG (100 nM; S7779, Selleckchem). Gli2ΔN cells were treated 48 hours after transfection. Myc-GLI2-DN (17649, pCS2-MT-GLI2-deltaN) plasmid was purchased from Addgene. Plasmid transfection was performed using TurboFect transfection reagent (no. R0531, Thermo Scientific) according to the manufacturer's instructions, starting by plating NIH3T3 cells in a six-well

dish at 400,000 cells per well. Unless otherwise stated, all treatments were applied for 6 hours. Cell lysis and RNA extraction were performed using QIAshredder homogenizers (79654, Qiagen) and RNeasy Mini kits (74104, Qiagen) according to the manufacturer's instructions. RNA was quantified on a NanoDrop spectrophotometer (ND-2000C, NanoDrop Technologies). Transcript quantification was performed using the TaqMan RNA-to-Ct 1-Step Kit (4392656, Life Technologies) and the Applied Biosystems 7500 RT-PCR system, as per the manufacturer's instructions. All primers were commercially available TaqMan Gene Expression Assay probes purchased from Life Technologies (csnk2a1, Mm00786779_s1; csnk2a2, Mm01243455_m1; csnk2b, Mm00487216_m1; Gli1, Mm00494654_m1; Gli2, Mm.273292; Ccnd1, Mm.273049; and Ptch, Mm00436047_m1). All studies were done in biological triplicate (cultures).

Western blotting
Gli2ΔN cells were lysed with radioimmunoprecipitation assay buffer (sc-24948, Santa Cruz Biotechnology) supplemented with Halt Protease and Phosphatase Inhibitor (no. 1861218, Thermo Fisher Scientific) for 30 min on ice, and lysates were cleared by centrifugation at 13,000 rpm for 15 min at 4°C. Supernatants were incubated with 6× Laemmli sample buffer (10570021-1, BioWorld) at 95°C for 5 min. The samples were then separated with a 7% SDS-PAGE gel and immunoblotted with anti-Gli2 (AF3635-SP, R&D Systems) and anti–β-tubulin (ab6046, Abcam).

Cell viability
$Ptch^{+/-}$ MB cell lines MB55 and MB21 were gifts from R. Segal [Dana-Farber Cancer Institute (DFCI), Harvard Medical School], and $Ptch^{+/-};p53^{-/-}$ MB53 cells were gifts from C. Rudin [Memorial Sloan Kettering Cancer Center]. All cell lines were validated for high Gli activity via qRT-PCR and cultured at low density as neurospheres in DMEM/F12 supplemented with B27 and 1% penicillin/streptomycin. Cell viability was assessed using CellTiter-Glo (G7573, Promega) according to the manufacturer's instructions. Cells were plated at 10,000 cells per well in 96-well plates and treated with drugs as indicated, and data were collected on a TECAN Infinite 200 plate reader. All measurements were done in biological triplicate (cultures). Prism statistical analysis software (GraphPad Software) was used to determine relative IC_{50}s, normalized to control (DMSO)–treated cells [nonlinear regression, log(inhibitor) versus response (variable slope) curve].

Human MB cell lines
The RCMB32 and ICb-984 cell lines were generated in the laboratories of R.J.W.-R. and X.-N.L., respectively. The BT084 cell line was provided by T. Milde (Heidelberg University Hospital). The primary tumor sample (ST01) (Fig. 5, B to D) obtained from the operating room was minced and then placed in a digestion buffer consisting of 1× HBSS (14185, Thermo Fisher Scientific), papain (10 U/ml; LSOO3126, Worthington Biochemical), and DNase (250 U/ml; D4627, Sigma). Papain solution was aspirated and replaced with 1× HBSS containing ovomucoid (8 mg/ml; LK003182, Worthington Biochemical), BSA (8 mg/ml; Sigma), and DNase (250 U/ml) and then triturated using a Pasteur pipette to obtain a single-cell suspension. Cells were pelleted and resuspended in 0.02% HBSS-BSA. Cells were passed through a 70-μm nylon cell strainer (21008-952, VWR International). Of 20 tumors implanted, 5 grew to a sufficient size to resect. Cell viability and cell cycle stage were assayed by high-throughput single-cell immunofluorescence imaging, as described above; 12 individual

cultures were assessed at each dose. $Gli1$ transcript expression was assessed by qRT-PCR, as described above, and three cultures were assessed at each dose. Final pathological report is provided in table S4.

Acquired resistance
MB53, MB55, and MB21 cells were cultured in Neurobasal (10888-022, Life Technologies) + B27 supplement (17504044, Invitrogen) + DMSO (0.1%), CX-4945 (10 μM), or TBB (50 μM). Surviving cells were passaged for 1 month, after which only control (DMSO)–treated and MB55 TBB–treated (50 μM) cells remained. Complementary DNA (cDNA) from the three control lines and MB55 TBB–resistant lines was synthesized using the ABI High-Capacity cDNA Reverse Transcription Kit (4368814, Thermo Fisher Scientific) and amplified using the Q5 High-Fidelity DNA Polymerase (M0491, New England Biolabs) as per the manufacturer's protocol. Primer sequences are provided in table S3.

In silico analysis of mutant versus wild-type binding of TBB and CX-4945
All the crystal structures of CK2 were retrieved from the Protein Data Bank (PDB). To perform docking and MD simulations, the crystal structure of human CK2 in complex with CX4945 was used (PDB code: 3PE1). However, to compare the binding motif of TBB and to evaluate its position in the structure 3PE1, the crystallographic complex TBB/CK2 was also considered (PDB code: 1J91). The crystal structures were processed to remove unwanted ligands and water molecules. Hydrogen atoms were added to the protein structure using standard geometries (66). To minimize contacts between hydrogens, the structures were subjected to AMBER99 force-field minimization until the root mean square of conjugate gradient was <0.1 kcal mol^{-1} Å^{-1} (1 Å = 0.1 nm), keeping the heavy atoms fixed at their crystallographic positions (66). A similar protocol was used also to minimize the structures subjected to in silico mutagenesis. To strictly calibrate the high-throughput docking protocol, a small database of known CK2 inhibitors was built and a set of docking runs was performed. After the calibration phase, TBB and CX-4945 were docked directly into the ATP-binding site of selected CK2 crystal structures, by using AutoDock software (66). MD simulations of the considered structures (parameterized with AMBER99) were performed with NAMD 2.8 (67) to verify their stability over time; in particular, a 100-ns NPT (isothermal-isobaric) ensemble (1 atm, 300 K) MD simulation was performed after an equilibration phase of 1 ns (positional restraints were applied on carbon atoms to equilibrate the solvent around the protein) (68, 69).

Flank allograft studies
In vivo assays of CK2 inhibitor efficacy were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee at Stanford University. SmoWT-MB and SmoD477G-MB cells isolated from either parental SmoWT or SmoD477G mouse $Ptch^{+/-};p53^{-/-}$ MB hind-flank allografts were provided by C. Rudin (Memorial Sloan Kettering Cancer Center). $Ptch^{+/-};Tpr53^{-/-}$ or $Ptch^{+/-};Tpr53^{-/-};SmoD477G$ MB cells were mixed 1:1 with Matrigel, and 10^7 cells were injected into both flanks of a 6- to 7-week-old Nu/Nu mouse (strain 088, Charles River). Once tumors reached 250 to 300 mm^3, mice were randomized to treatment versus control groups. Treatment groups received TBB (30 mg/kg, ip BID; reconstituted in DMSO at 50 μg/μl) or GDC-0449 (37.5 mg/kg, ip BID; reconstituted in DMSO at 62.5 μg/μl) (S1082, Selleckchem). Control mice were

treated with an equivalent volume of DMSO. Corn oil was used in all groups to a total volume of 180 μl per injection, and all doses were given with a 28-gauge insulin syringe. Tumors were measured, and mice were weighed daily. Once tumors reached 1.5 cm^2 or met institutional euthanasia criteria, mice were sacrificed and tumors were harvested. Growth curves were generated using GraphPad Prism, and significance was measured by two-way analysis of variance (ANOVA).

Cerebellar allografts

Cells from *Ptch*$^{+/-}$;*Tpr53*$^{-/-}$;*SmoD477G* mice were suspended in sterile PBS to an adjusted final concentration of 80,000 cells per 2-μl injection. NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice (6 to 7 weeks old; The Jackson Laboratory) were anesthetized and fixed to a stereotactic frame, and the head was sterilized. A small sagittal incision was made in the mouse's scalp to expose the area surrounding lambda. The skull was sterilized, and a 0.7-mm drill bit was used to create a midline hole located 2 mm down from lambda. A Hamilton syringe with affixed 26-gauge needle was lowered to a depth of ~3 mm and then withdrawn to 2.75 mm before injecting 2-μl (80,000) cells at 0.5 μl/min. The drill hole was closed with a thin slip of bone wax, and the wound was closed. Mice were randomized, and a blinded collaborator began treatment at 72 hours with either CX-4945 (37.5 mg/kg, orally BID; reconstituted in DMSO) or DMSO (equivalent volume to treatment group, dosed ip BID). Corn oil was used as vehicle for both CX-4945 and DMSO. Weight, morbidity, and survival were measured by a blinded collaborator.

Mouse strains

All animal experiments were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee at Stanford University. Male and female wild-type CD1-Elite mice (strain 482, Charles River) were sacrificed at P1, P7, and P14 for initial phosphoproteomic study. Female and male Atoh1/nGFP transgenic mice (*63*) were used for cerebellar development studies. Female Nu/Nu mice (6 to 7 weeks old; strain 088, Charles River) were used for flank studies, and female NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG) mice (6 to 7 weeks old; strain 005557, The Jackson Laboratory) were used for cerebellar tumor studies.

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/11/547/eaau5147/DC1

Fig. S1. Relative phosphopeptide abundance between P14 biological replicates.

Fig. S2. Relative proliferation in the P1, P7, and P14 cerebellum.

Fig. S3. GNP proliferation after CX-4945 treatment in culture.

Fig. S4. Knockdown specificity of pooled siRNA constructs targeting *Csnk2a1, Csnk2a2, Csnk2b,* and *Smo.*

Fig. S5. *Gli1* and Hh-associated transcript expression in NIH3T3 cells after CK2 inhibition.

Fig. S6. *Gli1* transcript expression in cultured SHH MB cell lines.

Fig. S7. Weight of control or CX-4945–treated mice with cerebellar SHH MB.

Fig. S8. *CSNK2A1* expression and 5-year survival of patients with SHH MB.

Fig. S9. Sequence conservation of *Csnk2a1* mutations in TBB-resistant SHH MB cells.

Table S1. Phosphoproteome of GNPs isolated at P1, P7, and P14.

Table S2. Phosphopeptides of enriched kinase motifs.

Table S3. Primer and siRNA sequences.

Table S4. Pathological report for the primary human MB sample (ST01).

REFERENCES AND NOTES

1. R. H. Zurawel, C. Allen, R. Wechsler-Reya, M. P. Scott, C. Raffel, Evidence that haploinsufficiency of *Ptch* leads to medulloblastoma in mice. *Genes Chromosomes Cancer* **28**, 77–81 (2000).

2. R. J. Wechsler-Reya, M. P. Scott, Control of neuronal precursor proliferation in the cerebellum by Sonic hedgehog. *Neuron* **22**, 103–114 (1999).

3. M. D. Taylor, P. A. Northcott, A. Korshunov, M. Remke, Y.-J. Cho, S. C. Clifford, C. G. Eberhart, D. W. Parsons, S. Rutkowski, A. Gajjar, D. W. Ellison, P. Lichter, R. J. Gilbertson, S. L. Pomeroy, M. Kool, S. M. Pfister, Molecular subgroups of medulloblastoma: The current consensus. *Acta Neuropathol.* **123**, 465–472 (2012).

4. S. K. Gupta, K. F. Meiri, K. Mahfooz, U. Bharti, S. Mani, Coordination between extrinsic extracellular matrix cues and intrinsic responses to orient the centrosome in polarizing cerebellar granule neurons. *J. Neurosci.* **30**, 2755–2766 (2010).

5. J. S. Espinosa, L. Luo, Timing neurogenesis and differentiation: Insights from quantitative clonal analyses of cerebellar granule cells. *J. Neurosci.* **28**, 2301–2312 (2008).

6. J. D. Corrales, G. L. Rocco, S. Blaess, Q. Guo, A. L. Joyner, Spatial pattern of sonic hedgehog signaling through *Gli* genes during cerebellum development. *Development* **131**, 5581–5590 (2004).

7. G. W. Robinson, B. A. Orr, G. Wu, S. Gururangan, T. Lin, I. Qaddoumi, R. J. Packer, S. Goldman, M. D. Prados, A. Desjardins, M. Chintagumpala, N. Takebe, S. C. Kaste, M. Rusch, S. J. Allen, A. Onar-Thomas, C. F. Stewart, M. Fouladi, J. M. Boyett, R. J. Gilbertson, T. Curran, D. W. Ellison, A. Gajjar, Vismodegib exerts targeted efficacy against recurrent sonic hedgehog–subgroup medulloblastoma: Results from phase II Pediatric Brain Tumor Consortium studies PBTC-025B and PBTC-032. *J. Clin. Oncol.* **33**, 2646–2654 (2015).

8. M. Kool, D. T. W. Jones, N. Jäger, P. A. Northcott, T. J. Pugh, V. Hovestadt, R. M. Piro, L. A. Esparza, S. L. Markant, M. Remke, T. Milde, F. Bourdeaut, M. Ryzhova, D. Sturm, E. Pfaff, S. Stark, S. Hutter, H. Seker-Cin, P. Johann, S. Bender, C. Schmidt, T. Rausch, D. Shih, J. Reimand, L. Sieber, A. Wittmann, L. Linke, H. Witt, U. D. Weber, M. Zapatka, R. König, R. Beroukhim, G. Bergthold, P. van Sluis, R. Volckmann, J. Koster, R. Versteeg, S. Schmidt, S. Wolf, C. Lawerenz, C. C. Bartholomae, C. von Kalle, A. Unterberg, C. Herold-Mende, S. Hofer, A. E. Kulozik, A. von Deimling, W. Scheurlen, J. Felsberg, G. Reifenberger, M. Hasselblatt, J. R. Crawford, G. A. Grant, N. Jabado, A. Perry, C. Cowdrey, S. Croul, G. Zadeh, J. O. Korbel, F. Doz, O. Delattre, G. D. Bader, M. G. McCabe, V. P. Collins, M. W. Kieran, Y. J. Cho, S. L. Pomeroy, O. Witt, B. Brors, M. D. Taylor, U. Schüller, A. Korshunov, R. Eils, R. J. Wechsler-Reya, P. Lichter, S. M. Pfister; ICGC PedBrain Tumor Project, Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405 (2014).

9. M. A. Price, D. Kalderon, Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by glycogen synthase kinase 3 and casein kinase 1. *Cell* **108**, 823–835 (2002).

10. M. Hammerschmidt, M. J. Bitgood, A. P. McMahon, Protein kinase A is a common negative regulator of Hedgehog signaling in the vertebrate embryo. *Genes Dev.* **10**, 647–658 (1996).

11. J. Zhang, P. L. Yang, N. S. Gray, Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009).

12. Z. A. Knight, H. Lin, K. M. Shokat, Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10**, 130–137 (2010).

13. P. Wu, T. E. Nielsen, M. H. Clausen, FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36**, 422–429 (2015).

14. L. V Goodrich, L. Milenković, K. M. Higgins, M. P. Scott, Altered neural cell fates and medulloblastoma in mouse *patched* mutants. *Science* **277**, 1109–1113 (1997).

15. P. A Northcott, D. T. W. Jones, M. Kool, G. W. Robinson, R. J. Gilbertson, Y.-J. Cho, S. L. Pomeroy, A. Korshunov, P. Lichter, M. D. Taylor, S. M. Pfister, Medulloblastomics: The end of the beginning. *Nat. Rev. Cancer* **12**, 818–834 (2012).

16. J. Bloom, F. R. Cross, Multiple levels of cyclin specificity in cell-cycle control. *Nat. Rev. Mol. Cell Biol.* **8**, 149–160 (2007).

17. F. Meggio, L. A Pinna, One-thousand-and-one substrates of protein kinase CK2? *FASEB J.* **17**, 349–368 (2003).

18. X. Zhu, D. Girardo, E. E. Govek, K. John, M. Mellén, P. Tamayo, J. P. Mesirov, M. E. Hatten, Role of Tet1/3 genes and chromatin remodeling genes in cerebellar circuit formation. *Neuron* **89**, 100–112 (2016).

19. P. A. Northcott, D. J. Shih, J. Peacock, L. Garzia, A. S. Morrissy, T. Zichner, A. M. Stütz, A. Korshunov, J. Reimand, S. E. Schumacher, R. Beroukhim, D. W. Ellison, C. R. Marshall, A. C. Lionel, S. Mack, A. Dubuc, Y. Yao, V. Ramaswamy, B. Luu, A. Rolider, F. M. Cavalli, X. Wang, M. Remke, X. Wu, R. Y. Chiu, A. Chu, E. Chuah, R. D. Corbett, G. R. Hoad, S. D. Jackman, Y. Li, A. Lo, K. L. Mungall, K. M. Nip, J. Q. Qian, A. G. Raymond, N. T. Thiessen, R. J. Varhol, I. Birol, R. A. Moore, A. J. Mungall, R. Holt, D. Kawauchi, M. F. Roussel, M. Kool, D. T. Jones, H. Witt, A. Fernandez-L, A. M. Kenney, R. J. Wechsler-Reya, P. Dirks, T. Aviv, W. A. Grajkowska, M. Perek-Polnik, C. C. Haberler, O. Delattre, S. S. Reynaud, F. F. Doz, S. S. Pernet-Fattet, B. K. Cho, S. K. Kim, K. C. Wang, W. Scheurlen, C. G. Eberhart, M. Fèvre-Montange, A. Jouvet, I. F. Pollack, X. Fan, K. M. Muraszko, G. Y. Gillespie, C. Di Rocco, L. Massimi, E. M. Michiels, N. K. Kloosterhof, P. J. French, J. M. Kros, J. M. Olson, R. G. Ellenbogen, K. Zitterbart, L. Kren, R. C. Thompson, M. K. Cooper, B. Lach, R. E. McLendon, D. D. Bigner, A. Fontebasso, S. Albrecht, N. Jabado, J. C. Lindsey, S. Bailey, N. Gupta, W. A. Weiss, L. Bognár, A. Klekner, T. E. Van Meter, T. Kumabe, T. Tominaga, S. K. Elbaaba, J. R. Leonard, J. B. Rubin, L. M. Liau, E. G. Van Meir, M. Fouladi, H. Nakamura, G. Cinalli, M. Garami, P. Hauser, A. G. Saad, A. Iolascon, S. Jung, C. G. Carlotti, R. Vibhakar,

Y. S. Ra, S. Robinson, M. Zollo, C. C. Faria, J. A. Chan, M. L. Levy, P. H. Sorensen, M. Meyerson, S. L. Pomeroy, Y. J. Cho, G. D. Bader, U. Tabori, C. E. Hawkins, E. Bouffet, S. W. Scherer, J. T. Rutka, D. Malkin, S. C. Clifford, S. J. Jones, J. O. Korbel, S. M. Pfister, M. A. Marra, M. D. Taylor, Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).

20. H.-C. Cheng, R. Z. Qi, H. Paudel, H.-J. Zhu, Regulation and function of protein kinases and phosphatases. *Enzyme Res.* **2011**, 794089 (2011).

21. S. Sarno, H. Reddy, F. Meggio, M. Ruzzene, S. P. Davies, A. Donella-Deana, D. Shugar, L. A. Pinna, Selectivity of 4,5,6,7-tetrabromobenzotriazole, an ATP site-directed inhibitor of protein kinase CK2 ('casein kinase-2'). *FEBS Lett.* **496**, 44–48 (2001).

22. M. A. Pagano, J. Bain, Z. Kazimierczuk, S. Sarno, M. Ruzzene, G. Di Maira, M. Elliott, A. Orzeszko, G. Cozza, F. Meggio, L. A. Pinna, The selectivity of inhibitors of protein kinase CK2: An update. *Biochem. J.* **415**, 353–365 (2008).

23. G. Cozza, M. Mazzorana, E. Papinutto, J. Bain, M. Elliott, G. di Maira, A. Gianoncelli, M. A. Pagano, S. Sarno, M. Ruzzene, R. Battistutta, F. Meggio, S. Moro, G. Zagotto, L. A. Pinna, Quinalizarin as a potent, selective and cell-permeable inhibitor of protein kinase CK2. *Biochem. J.* **421**, 387–395 (2009).

24. H. Brehme, T. Kirschstein, R. Schulz, R. Köhling, In vivo treatment with the casein kinase 2 inhibitor 4,5,6,7-tetrabromotriazole augments the slow afterhyperpolarizing potential and prevents acute epileptiform activity. *Epilepsia* **55**, 175–183 (2014).

25. A. V Ljubimov, S. Caballero, A. M. Aoki, L. A. Pinna, M. B. Grant, R. Castellon, Involvement of protein kinase CK2 in angiogenesis and retinal neovascularization. *Invest. Ophthalmol. Vis. Sci.* **45**, 4583–4591 (2004).

26. A. Sudarov, A. L. Joyner, Cerebellum morphogenesis: The foliation pattern is orchestrated by multi-cellular anchoring centers. *Neural Dev.* **2**, 26 (2007).

27. J. Xie, R. L. Johnson, X. Zhang, J. W. Bare, F. M. Waldman, P. H. Cogen, A. G. Menon, R. S. Warren, L.-C. Chen Jr., M. P. Scott, E. H. Epstein, Mutations of the *PATCHED* gene in several types of sporadic extracutaneous tumors. *Cancer Res.* **57**, 2369–2372 (1997).

28. L. Milenkovic, M. P. Scott, R. Rohatgi, Lateral transport of Smoothened from the plasma membrane to the membrane of the cilium. *J. Cell Biol.* **187**, 365–374 (2009).

29. L. Milenkovic, L. E. Weiss, J. Yoon, T. L. Roth, Y. S. Su, S. J. Sahl, M. P. Scott, W. E. Moerner, Single-molecule imaging of Hedgehog pathway protein Smoothened in primary cilia reveals binding events regulated by Patched1. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8320–8325 (2015).

30. E. W. Humke, K. V. Dorn, L. Milenkovic, M. P. Scott, R. Rohatgi, The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* **24**, 670–682 (2010).

31. D. Y. Lou, I. Dominguez, P. Toselli, E. Landesman-Bollag, C. O'Brien, D. C. Seldin, The alpha catalytic subunit of protein kinase CK2 is required for mouse embryonic development. *Mol. Cell. Biol.* **28**, 131–139 (2008).

32. T. Buchou, M. Vernet, O. Blond, H. H. Jensen, H. Pointu, B. B. Olsen, C. Cochet, O. G. Issinger, B. Boldyreff, Disruption of the regulatory beta subunit of protein kinase CK2 in mice leads to a cell-autonomous defect and early embryonic lethality. *Mol. Cell. Biol.* **23**, 908–915 (2003).

33. R. Battistutta, G. Cozza, Fabrice Pierre, E. Papinutto, G. Lolli, S. Sarno, S. E. O'Brien, A. Siddiqui-Jain, M. Haddach, K. Anderes, D. M. Ryckman, F. Meggio, L. A. Pinna, Unprecedented selectivity and structural determinants of a new class of protein kinase CK2 inhibitors in clinical trials for the treatment of cancer. *Biochemistry* **50**, 8478–8488 (2011).

34. J. K. Chen, J. Taipale, K. E. Young, T. Maiti, P. A. Beachy, Small molecule modulation of Smoothened activity. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14071–14076 (2002).

35. C. M. Rudin, C. L. Hann, J. Laterra, R. L. Yauch, C. A. Callahan, L. Fu, T. Holcomb, J. Stinson, S. E. Gould, B. Coleman, P. M. LoRusso, D. D. Von Hoff, F. J. de Sauvage, J. A. Low, Treatment of medulloblastoma with hedgehog pathway inhibitor GDC-0449. *N. Engl. J. Med.* **361**, 1173–1178 (2009).

36. M.-H. Chen, C. W. Wilson, Y.-J. Li, K. K. L. Law, C.-S. Lu, R. Gacayan, X. Zhang, C.-c. Hui, P.-T. Chuang, Cilium-independent regulation of Gli protein function by Sufu in Hedgehog signaling is evolutionarily conserved. *Genes Dev.* **23**, 1910–1928 (2009).

37. H. Jia, Y. Liu, R. Xia, C. Tong, T. Yue, J. Jiang, J. Jia, Casein kinase 2 promotes Hedgehog signaling by regulating both Smoothened and Cubitus interruptus. *J. Biol. Chem.* **285**, 37218–37226 (2010).

38. S. Zhang, Y. Wang, J.-H. Mao, D. Hsieh, I.-J. Kim, L.-M. Hu, Z. Xu, H. Long, D. M. Jablons, L. You, Inhibition of CK2α down-regulates Hedgehog/Gli signaling leading to a reduction of a stem-like side population in human lung cancer cells. *PLOS ONE* **7**, e38996 (2012).

39. E. Roessler, A. N. Ermilov, D. K. Grange, A. Wang, M. Grachtchouk, A. A. Dlugosz, M. Muenke, A previously unidentified amino-terminal domain regulates transcriptional activity of wild-type and disease-associated human GLI2. *Hum. Mol. Genet.* **14**, 2181–2188 (2005).

40. S. Buonamici, J. Williams, M. Morrissey, A. Wang, R. Guo, A. Vattay, K. Hsiao, J. Yuan, J. Green, B. Ospina, Q. Yu, L. Ostrom, P. Fordjour, D. L. Anderson, J. E. Monahan, J. F. Kelleher, S. Peukert, S. Pan, X. Wu, S.-M. Maira, C. García-Echeverría, K. J. Briggs, D. Neil Watkins, Y.-m. Yao, C. Lengauer, M. Warmuth, W. R. Sellers, M. Dorsch, Interfering with resistance to Smoothened antagonists by inhibition of the PI3K pathway in medulloblastoma. *Sci. Transl. Med.* **2**, 51ra70 (2010).

41. X. Zhao, T. Ponomaryov, K. J. Ornell, P. Zhou, S. K. Dabral, E. Pak, W. Li, S. X. Atwood, R. J. Whitson, A. L. S. Chang, J. Li, A. E. Oro, J. A. Chan, J. F. Kelleher, R. A. Segal, RAS/MAPK activation drives resistance to Smo inhibition, metastasis, and tumor evolution in Shh pathway–dependent tumors. *Cancer Res.* **75**, 3623–3635 (2015).

42. X. Zhao, E. Pak, K. J. Ornell, M. F. Pazyra-Murphy, E. L. MacKenzie, E. J. Chadwick, T. Ponomaryov, J. F. Kelleher, R. A. Segal, A transposon screen identifies loss of primary cilia as a mechanism of resistance to SMO inhibitors. *Cancer Discov.* **7**, 1436–1449 (2017).

43. Y. Pei, K. W. Liu, J. Wang, A. Garancher, R. Tao, L. A. Esparza, D. L. Maier, Y. T. Udaka, N. Murad, S. Morrissy, H. Seker-Cin, S. Brabetz, L. Qi, M. Kogiso, S. Schubert, J. M. Olson, Y. J. Cho, X. N. Li, J. R. Crawford, M. L. Levy, M. Kool, S. M. Pfister, M. D. Taylor, R. J. Wechsler-Reya, HDAC and PI3K antagonists cooperate to inhibit growth of MYC-driven medulloblastoma. *Cancer Cell* **29**, 311–323 (2016).

44. S. N. Brun, S. L. Markant, L. A. Esparza, G. Garcia, D. Terry, J.-M. Huang, M. S. Pavlyukov, X.-N. Li, G. A. Grant, J. R. Crawford, M. L. Levy, E. M. Conway, L. H. Smith, I. Nakano, A. Berezov, M. I. Greene, Q. Wang, R. J. Wechsler-Reya, Survivin as a therapeutic target in Sonic hedgehog-driven medulloblastoma. *Oncogene* **34**, 3770–3779 (2015).

45. R. J. Vanner, M. Remke, M. Gallo, H. J. Selvadurai, F. Coutinho, L. Lee, M. Kushida, R. Head, S. Morrissy, X. Zhu, T. Aviv, V. Voisin, I. D. Clarke, Y. Li, A. J. Mungall, R. A. Moore, Y. Ma, S. J. M. Jones, M. A. Marra, D. Malkin, P. A. Northcott, M. Kool, S. M. Pfister, G. Bader, K. Hochedlinger, A. Korshunov, M. D. Taylor, P. B. Dirks, Quiescent SOX2⁺ cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell* **26**, 33–47 (2014).

46. R. L. Yauch, G. J. P. Dijkgraaf, B. Alicke, T. Januario, C. P. Ahn, T. Holcomb, K. Pujara, J. Stinson, C. A. Callahan, T. Tang, J. F. Bazan, Z. Kan, S. Seshagiri, C. L. Hann, S. E. Gould, J. A. Low, C. M. Rudin, F. J. de Sauvage, *Smoothened* mutation confers resistance to a Hedgehog pathway inhibitor in medulloblastoma. *Science* **326**, 572–574 (2009).

47. J. Rodon, H. A. Tawbi, A. L. Thomas, R. G. Stoller, C. P. Turtschi, J. Baselga, J. Sarantopoulos, D. Mahalingam, Y. Shou, M. A. Moles, L. Yang, C. Granvil, E. Hurh, K. L. Rose, D. D. Amakye, R. Dummer, A. C. Mita, A phase I, multicenter, open-label, first-in-human, dose-escalation study of the oral smoothened inhibitor sonidegib (LDE225) in patients with advanced solid tumors. *Clin. Cancer Res.* **20**, 1900–1909 (2014).

48. P. M. LoRusso, C. M. Rudin, J. C. Reddy, R. Tibes, G. J. Weiss, M. J. Borad, C. L. Hann, J. R. Brahmer, I. Chang, W. C. Darbonne, R. A. Graham, K. L. Zerivitz, J. A. Low, D. D. Von Hoff, Phase I trial of hedgehog pathway inhibitor vismodegib (GDC-0449) in patients with refractory, locally advanced or metastatic solid tumors. *Clin. Cancer Res.* **17**, 2502–2511 (2011).

49. K. D. Robarge, S. A. Brunton, G. M. Castanedo, Y. Cui, M. S. Dina, R. Goldsmith, S. E. Gould, O. Guichert, J. L. Gunzner, J. Halladay, W. Jia, C. Khojasteh, M. F. T. Koehler, K. Kotkow, H. La, R. L. LaLonde, K. Lau, L. Lee, D. Marshall, J. C. Marsters Jr., L. J. Murray, C. Qian, L. L. Rubin, L. Salphati, M. S. Stanley, J. H. A. Stibbard, D. P. Sutherlin, S. Ubhayaker, S. Wang, S. Wong, M. Xie, GDC-0449—A potent inhibitor of the hedgehog pathway. *Bioorg. Med. Chem. Lett.* **19**, 5576–5581 (2009).

50. R. F. Marschke, M. J. Borad, R. W. McFarland, R. H. Alvarez, J. K. Lim, C. S. Padgett, D. D. Von Hoff, S. E. O'Brien, D. W. Northfelt, Findings from the phase I clinical trials of CX-4945, an orally available inhibitor of CK2. *J. Clin. Oncol.* **29**, 3087–3087 (2011).

51. J. Kim, B. T. Aftab, J. Y. Tang, D. Kim, A. H. Lee, M. Rezaee, J. Kim, B. Chen, E. M. King, A. Borodovsky, G. J. Riggins, E. H. Epstein, P. A. Beachy, C. M. Rudin, Itraconazole and arsenic trioxide inhibit Hedgehog pathway activation and tumor growth associated with acquired resistance to smoothened antagonists. *Cancer Cell* **23**, 23–34 (2013).

52. K. Niefind, B. Guerra, I. Ermakowa, O.-G. Issinger, Crystal structure of human protein kinase CK2: Insights into basic properties of the CK2 holoenzyme. *EMBO J.* **20**, 5320–5331 (2001).

53. S. Sarno, S. Moro, F. Meggio, G. Zagotto, D. D. Ben, P. Ghisellini, R. Battistutta, G. Zanotti, L. A. Pinna, Toward the rational design of protein kinase casein kinase-2 inhibitors. *Pharmacol. Ther.* **93**, 159–168 (2002).

54. Y. Bian, J. Han, V. Kannabiran, S. Mohan, H. Cheng, J. Friedman, L. Zhang, C. VanWaes, Z. Chen, Mek inhibitor PD-0325901 overcomes resistance to CK2 inhibitor CX-4945 and exhibits anti-tumor activity in head and neck cancer. *Int. J. Biol. Sci.* **11**, 411–422 (2015).

55. L. Ferrer-Font, L. Villamañan, N. Arias-Ramos, J. Vilardell, M. Plana, M. Ruzzene, L. A. Pinna, E. Itarte, A. P. Candiota, Targeting protein kinase CK2: Evaluating CX-4945 potential for GL261 glioblastoma therapy in immunocompetent mice. *Pharmaceuticals* **10**, E24 (2017).

56. S. Surade, T. L. Blundell, Structural biology and drug discovery of difficult targets: The limits of ligandability. *Chem. Biol.* **19**, 42–50 (2012).

57. J. T. Romer, H. Kimura, S. Magdaleno, K. Sasai, C. Fuller, H. Baines, M. Connelly, C. F. Stewart, S. Gould, L. L. Rubin, T. Curran, Suppression of the Shh pathway using a small molecule inhibitor eliminates medulloblastoma in *Ptc1⁺/⁻p53⁻/⁻* mice. *Cancer Cell* **6**, 229–240 (2004).

58. H. Y. Lee, L. A. Greene, C. A. Mason, M. C. Manzini, Isolation and culture of post-natal mouse cerebellar granule neuron progenitor cells and neurons. *J. Vis. Exp.*, 990 (2009).

59. X. Ge, L. Milenkovic, K. Suyama, T. Hartl, T. Purzner, A. Winans, T. Meyer, M. P. Scott, Phosphodiesterase 4D acts downstream of Neuropilin to control Hedgehog signal transduction and the growth of medulloblastoma. *eLife* **4**, e07068 (2015).

60. M. P. Weekes, P. Tomasec, E. L. Huttlin, C. A. Fielding, D. Nusinow, R. J. Stanton, E. C. Y. Wang, R. Aicheler, I. Murrell, G. W. G. Wilkinson, P. J. Lehner, S. P. Gygi, Quantitative temporal viromics: An approach to investigate host-pathogen interaction. *Cell* **157**, 1460–1472 (2014).

61. D. Schwartz, S. P. Gygi, An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398 (2005).

62. M. F. Chou, D. Schwartz, Biological sequence motif discovery using *motif-x*. *Curr. Protoc. Bioinformatics* **35**, 15–24 (2011).

63. E. A. Lumpkin, T. Collisson, P. Parab, A. Omer-Abdalla, H. Haeberle, P. Chen, A. Doetzlhofer, P. White, A. Groves, N. Segil, J. E. Johnson, Math1-driven GFP expression in the developing nervous system of transgenic mice. *Gene Expr. Patterns* **3**, 389–395 (2003).

64. S. D. Cappell, M. Chung, A. Jaimovich, S. L. Spencer, T. Meyer, Irreversible APC[Cdh1] inactivation underlies the point of no return for cell-cycle entry. *Cell* **166**, 167–180 (2016).

65. C. Fraley, A. E. Raftery, Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **97**, 611–631 (2002).

66. G. Cozza, C. Girardi, A. Ranchio, G. Lolli, S. Sarno, A. Orzeszko, Z. Kazimierczuk, R. Battistutta, M. Ruzzene, L. A. Pinna, Cell-permeable dual inhibitors of protein kinases CK2 and PIM-1: Structural features and pharmacological potential. *Cell. Mol. Life Sci.* **71**, 3173–3185 (2014).

67. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

68. R. Costa, G. Arrigoni, G. Cozza, G. Lolli, R. Battistutta, J. Carlos, I. Belmonte, L. A. Pinna, S. Sarno, The lysine-specific demethylase 1 is a novel substrate of protein kinase CK2. *Biochim. Biophys. Acta* **1844**, 722–729 (2014).

69. A. Venerando, C. Franchin, N. Cant, G. Cozza, M. A. Pagano, K. Tosoni, A. Al-Zahrani, G. Arrigoni, R. C. Ford, A. Mehta, L. A. Pinna, Detection of phospho-sites generated by protein kinase CK2 in CFTR: Mechanistic aspects of Thr1471 phosphorylation. *PLOS ONE* **8**, e74232 (2013).

Citation: T. Purzner, J. Purzner, T. Buckstaff, G. Cozza, S. Gholamin, J. M. Rusert, T. A. Hartl, J. Sanders, N. Conley, X. Ge, M. Langan, V. Ramaswamy, L. Ellis, U. Litzenburger, S. Bolin, J. Theruvath, R. Nitta, L. Qi, X.-N. Li, G. Li, M. D. Taylor, R. J. Wechsler-Reya, L. A. Pinna, Y.-J. Cho, M. T. Fuller, J. E. Elias, M. P. Scott, Developmental phosphoproteomics identifies the kinase CK2 as a driver of Hedgehog signaling and a therapeutic target in medulloblastoma. *Sci. Signal.* **11**, eaau5147 (2018).

CANCER

The HDAC3–SMARCA4–miR-27a axis promotes expression of the *PAX3:FOXO1* fusion oncogene in rhabdomyosarcoma

Narendra Bharathy[1], Noah E. Berlow[1], Eric Wang[2], Jinu Abraham[3], Teagan P. Settelmeyer[1], Jody E. Hooper[4], Matthew N. Svalina[1], Yoshihiro Ishikawa[5], Keith Zientek[5], Zia Bajwa[1,6], Martin W. Goros[7], Brian S. Hernandez[7], Johannes E. Wolff[8], Michelle A. Rudek[9,10,11], Linping Xu[9,10], Nicole M. Anders[9,10], Ranadip Pal[12], Alexandra P. Harrold[1], Angela M. Davies[13], Arya Ashok[13], Darnell Bushby[13], Maria Mancini[13], Christopher Noakes[13], Neal C. Goodwin[13], Peter Ordentlich[14], James Keck[15], Douglas S. Hawkins[16], Erin R. Rudzinski[16], Bishwanath Chatterjee[17], Hans Peter Bächinger[5,18], Frederic G. Barr[17], Jennifer Liddle[19], Benjamin A. Garcia[19], Atiya Mansoor[6], Theodore J. Perkins[20,21], Christopher R. Vakoc[2], Joel E. Michalek[7], Charles Keller[1]*

Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma of childhood with an unmet clinical need for decades. A single oncogenic fusion gene is associated with treatment resistance and a 40 to 45% decrease in overall survival. We previously showed that expression of this *PAX3:FOXO1* fusion oncogene in alveolar RMS (aRMS) mediates tolerance to chemotherapy and radiotherapy and that the class I–specific histone deacetylase (HDAC) inhibitor entinostat reduces PAX3:FOXO1 protein abundance. Here, we established the antitumor efficacy of entinostat with chemotherapy in various preclinical cell and mouse models and found that HDAC3 inhibition was the primary mechanism of entinostat-induced suppression of PAX3:FOXO1 abundance. HDAC3 inhibition by entinostat decreased the activity of the chromatin remodeling enzyme SMARCA4, which, in turn, derepressed the microRNA miR-27a. This reexpression of miR-27a led to *PAX3:FOXO1* mRNA destabilization and chemotherapy sensitization in aRMS cells in culture and in vivo. Furthermore, a phase 1 clinical trial (ADVL1513) has shown that entinostat is tolerable in children with relapsed or refractory solid tumors and is planned for phase 1B cohort expansion or phase 2 clinical trials. Together, these results implicate an HDAC3–SMARCA4–miR-27a–PAX3:FOXO1 circuit as a driver of chemoresistant aRMS and suggest that targeting this pathway with entinostat may be therapeutically effective in patients.

INTRODUCTION

Soft tissue sarcomas are among the most common and deadliest childhood cancers (*1*, *2*). Despite improved survival for other childhood cancers, progress for particularly metastatic sarcoma has been minimal, and therapeutic options remain limited (*2*, *3*). Rhabdomyosarcoma (RMS) is the most common soft tissue sarcoma in children (*2*). Patients with diagnosed metastatic RMS have dismal prognoses for survival that have not improved in years, if not several decades (*4*–*6*). Although ~85% of local lesions can be excised by surgery, undetectable circulating tumor cells and recurrence at the primary site from occult microscopic residual disease are presumed to act as a gateway to relapse and subsequent regional and distal metastasis (*7*).

Alveolar RMS (aRMS) is the more aggressive RMS subtype and is generally not survivable long term once metastatic (<8%) (*4*, *6*, *8*). This disease does respond to chemotherapy in most cases, but the clinical challenge is preventing or overcoming recurrence after chemotherapy and/or radiation (*9*). Unique to aRMS is the expression of a fusion protein called PAX3:FOXO1, which results from a chromosomal translocation and comprises the transcription factors PAX3 (paired box 3) and FOXO1 (forkhead box 1). PAX3:FOXO1 is seen in some patients with aRMS. We have previously reported that native, non-constitutive, G_2 cell cycle phase-enriched expression of the PAX3:FOXO1 oncogene in aRMS facilitates checkpoint adaptation (meaning a high tolerance of DNA breaks or mitotic catastrophe) [(*10*) and reviewed in (*11*)]. In terms of clinical impact, this PAX3:FOXO1-mediated treatment resistance causes a notable 40 to 45% difference (drop) in survival over 10 years compared to PAX3:FOXO1-negative cases (*9*). In experimental studies, genetic knockdown of PAX3:FOXO1 abundance improves chemotherapy and radiation sensitivity and reduces tumor reestablishment (*10*),

[1]Children's Cancer Therapy Development Institute, Beaverton, OR 97005, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [3]Department of Pediatrics, Oregon Health & Science University, Portland, OR 97239, USA. [4]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA. [5]Research Center, Shriners Hospital for Children, Portland, OR 97239, USA. [6]Department of Pathology, Oregon Health & Science University, Portland, OR 97239, USA. [7]Department of Epidemiology and Biostatistics, University of Texas Health Science Center, San Antonio, TX 78229, USA. [8]Department of Pediatric Hematology Oncology and Blood and Marrow Transplantation, Cleveland Clinic Children's, Cleveland, OH 44195, USA. [9]Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD 21231, USA. [10]Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21224, USA. [11]Division of Clinical Pharmacology, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21231, USA. [12]Electrical and Computer Engineering, Texas Tech University, Lubbock, TX 79409, USA. [13]Champions Oncology, Rockville, MD 20850, USA. [14]Syndax Pharmaceuticals, Waltham, MA 02451, USA. [15]The Jackson Laboratory, Sacramento, CA 95838, USA. [16]Seattle Children's Hospital, Seattle, WA 98105, USA. [17]Cancer Molecular Pathology Section, Laboratory of Pathology, National Cancer Institute, Bethesda, MD 20892-1500, USA. [18]Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, OR 97239, USA. [19]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA. [20]Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa K1H 8L6, Canada. [21]Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa K1H 8M5, Canada.
*Corresponding author. Email: charles@cc-tdi.org

but drugging transcription factors such as PAX3:FOXO1 in patients is a daunting task. However, in this study, we found that PAX3:FOXO1 can be pharmacologically silenced at both the mRNA and protein levels by entinostat (ENT), a class I histone deacetylase (HDAC) inhibitor (HDACi) (with weak HDAC10/HDAC11 inhibition) that was granted a U.S. Food and Drug Administration breakthrough designation for certain breast cancers in adults.

In combination with the chemotherapy vincristine (VCR), ENT had strong antitumor activity in a genetically engineered mouse model (GEMM) of aRMS and in both orthotopic allograft and patient-derived xenograft (PDX) mouse models of aRMS. Pretreatment with ENT also improved radiation sensitivity. We further established the comparative efficacy of ENT to other HDACis at reducing the expression of PAX3:FOXO1 and that of another, related fusion protein also found in aRMS, PAX7:FOXO1. Mechanistically, we determined that, in aRMS cells, HDAC3 inhibition by ENT decreases SMARCA4, which, in turn, derepresses miR-27a, which then silences PAX3:FOXO1; by blocking HDAC3, ENT suppressed the abundance and activity of PAX3:FOXO1 and the growth of fusion-positive aRMS tumors in mice. Together, these preclinical and mechanistic studies have led to the Children's Oncology Group pediatric phase 1 clinical trial, ADVL1513, and a planned phase 1B cohort expansion for patients with RMS for this same trial.

RESULTS
ENT in combination with VCR slows aRMS tumor growth in vivo
In preclinical studies using RNA interference, depletion of PAX3:FOXO1 abundance in aRMS cells is shown to be critical to overcome checkpoint adaptation, a process whereby tumor cells survive chemotherapy and radiation in late phases of the cell cycle (summarized in Fig. 1A) (*10*). To affirm genetically that silencing *PAX3:FOXO1* increases chemosensitivity of aRMS to VCR in vitro, we first revalidated two derivative cell lines from the aRMS GEMM-derived primary tumor culture U23674 that did or did not have stable knockdown of PAX3:FOXO1 (shYFP or shCtrl, respectively) (*10*). PAX3:FOXO1 was then also restored by retroviral-mediated transduction into the shYFP-U23674 line. Immunoblot analysis confirmed stable knockdown of PAX3:FOXO1 and/or restored expression of PAX3:FOXO1, respectively (Fig. 1B). When treated with VCR for 24 hours, shYFP (PAX3:FOXO1 knockdown) cells alone had an absolute half maximal inhibitory concentration (IC_{50}) of 33 nM. However, control cells did not reach an absolute IC_{50} (Fig. 1C). Restoring expression of PAX3:FOXO1 in the stable knockdown cells (shYFP + PAX3:FOXO1) rescued the cells from the cytotoxic effects of VCR, in that no absolute IC_{50} was reached (Fig. 1C). Although PAX3:FOXO1 was tumor protective against chemotherapy, PAX3:FOXO1 knockdown was only cytostatic, which is consistent with previous reports of murine and human aRMS (*10*, *12*).

Our previous studies examined whether cell of origin influenced PAX3:FOXO1 expression and whether expression could be pharmacologically altered by HDACis or other epigenetic agents (*13*). To test whether the narrow-spectrum class I HDACi ENT could sensitize PAX3:FOXO1-positive aRMS to the chemotherapies most often used in high-risk or relapsed disease, we first tested the efficacy of ENT and VCR as single agents and in combination in orthotopic allograft mouse models of aRMS. This aRMS model was generated by injecting murine aRMS primary cell cultures into the cardiotoxin-

preinjured gastrocnemius muscle of severe combined immunodeficient hairless outbred (SHO) mice. The tumor-bearing mice were treated with ENT at a daily dose of 5 mg/kg administered intraperitoneally or with VCR at 1 mg/kg administered intraperitoneally once a week, or in combination. Unlike monotherapy of either agent, the combination of ENT and VCR was effective in delaying tumor growth in the aRMS mice (Fig. 1, D and E). Histologically, no cytodifferentiation was seen (table S1).

To determine whether pretreatment with ENT potentiates the effect of radiation on RMS cells, we pretreated a murine primary tumor cell culture (U23674) with ENT or dimethyl sulfoxide (DMSO) for 24 hours followed by irradiation [10 Gy (gray)] (fig. S1A). After irradiation, 500,000 viable cells were injected into the cardiotoxin-preinjured gastrocnemius muscle of SHO mice. Tumor cells treated with DMSO alone or ENT alone were used as controls. After injection of tumor cells, mice were monitored for tumor development. Mice that received tumor cells pretreated with ENT before irradiation showed increased latency of tumor development compared to mice that received cells pretreated with DMSO only (fig. S1B), suggesting that ENT treatment sensitizes RMS cells to radiation. With respect to chemotherapy, we also confirmed that the biochemical effect of PAX3:FOXO1 silencing occurs—and with synergy—at clinically achievable concentrations of ENT and VCR in human RMS cultures (fig. S2, A to C).

To further assess the clinical feasibility of ENT in a clinical setting, we investigated whether ENT had a dose-dependent effect and/or whether ENT was unique among HDACis for its suppressive effect on PAX3:FOXO1. First, we observed a dose-dependent effect of ENT on *PAX3:FOXO1* expression in Rh30 cells in response to doses from 0.1 to 2 μM, when examining the mRNA abundance at 24 hours (fig. S3A) and the protein abundance at 72 hours (fig. S3B). A PDX-derived explant aRMS culture, CF-001, also exhibited a dose-dependent response (fig. S3C), as did the murine aRMS culture, U23674 (fig. S3D).

To examine the relative efficacy of other HDACis at reducing PAX3:FOXO1 when compared to ENT (an HDAC1-3 inhibitor), we tested SAHA (suberoylanilide hydroxamic acid, a class I and class II HDACi), panobinostat (PAN), CUDC-907, and CUDC-101 in human and murine aRMS using the highest reported maximum drug serum concentration (C_{max}) in humans. Where C_{max} was unknown, we treated the cell cultures at the measured drug IC_{25} (25% inhibiting concentration). ENT at its highest achievable C_{max} (1000 nM) (*14*) significantly reduced *PAX3:FOXO1* expression in human aRMS (Rh30 and Rh41) cells (fig. S4A), which was reflected by near-complete reduction at the protein level in the human aRMS cells and in murine aRMS (U23674) cells (fig. S4B). In comparison, PAN (a broad-spectrum HDACi that targets HDAC1-11) at its C_{max} (45 nM) (*15*) did not consistently diminish PAX3:FOXO1 at either the mRNA or protein level (fig. S4, C and D), and SAHA (an HDAC1/2/3/6/8 inhibitor) was only somewhat effective at reducing PAX3:FOXO1 abundance at its C_{max} (1000 nM) (fig. S4, E and F) (*16*) but to a lesser degree than ENT. Near-complete cytotoxicity was observed in Rh41 cells with SAHA at the end of 72 hours of treatment, preventing inclusion of Rh41 in these biochemical studies. In addition to PAN and SAHA, CUDC-907 [a dual phosphatidylinositol 3-kinase (PI3K) and HDAC1/2/3/10 inhibitor] and CUDC-101 (a class I and class II HDACi) were tested but were found to be less efficient at reducing PAX3:FOXO1 compared to ENT at their IC_{25} (fig. S4, G and H). To examine whether these HDACis can have an effect on the related fusion protein, PAX7:FOXO1, the human
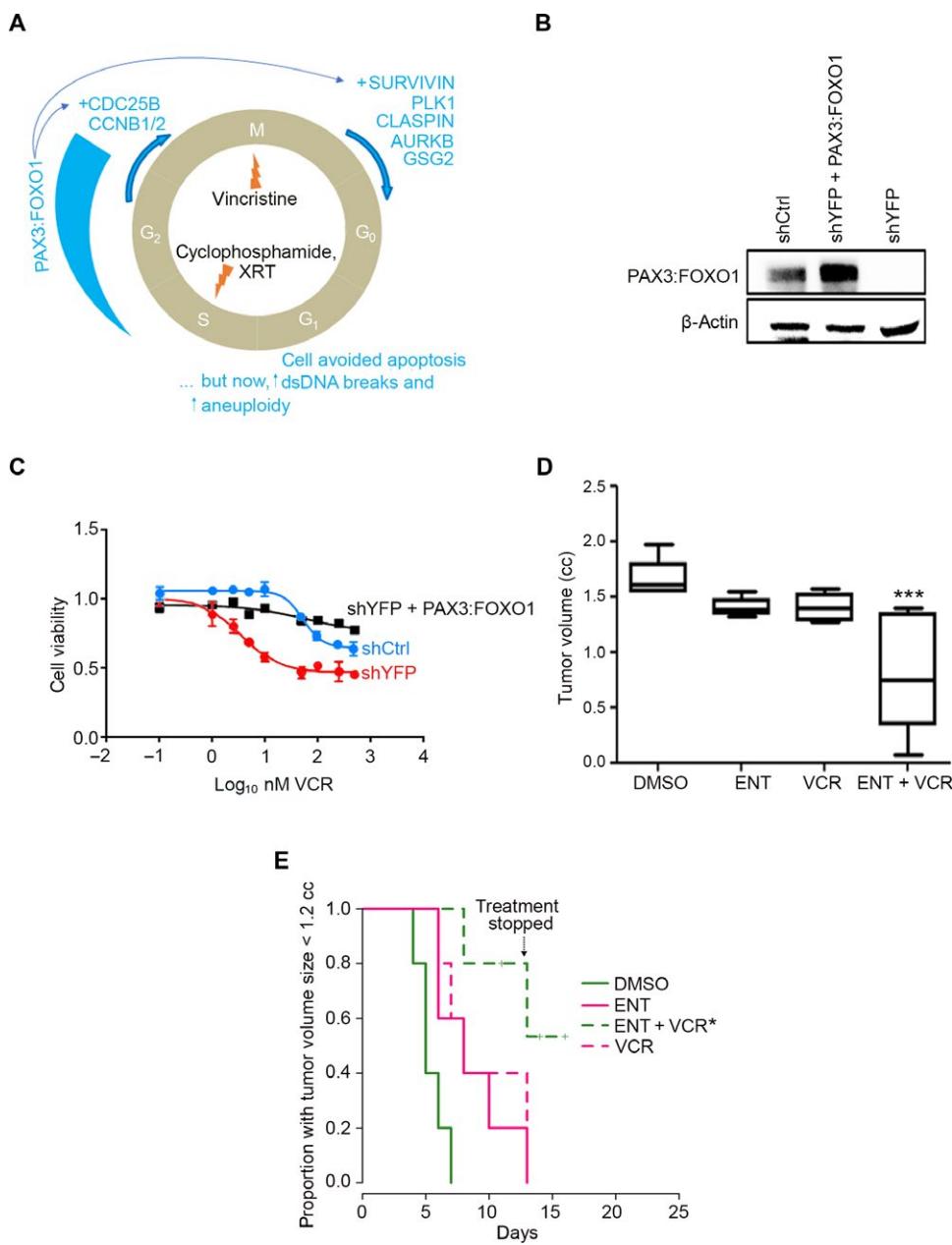
Fig. 1. ENT treatment of aRMS in vivo. (A) Diagrammatic representation of checkpoint adaptation, a process whereby tumor cells survive chemotherapy and radiation in late phases of the cell cycle. XRT, x-ray telescope radiotherapy; PLK1, polo-like kinase 1; AURKB, aurora kinase B; GSG2, genomic structure of haspin; CDC25B, cell division cycle 25b; CCNB1/2, cyclin B1/2; dsDNA, double-stranded DNA. (B) Basal PAX3:FOXO1 protein expression in murine U23674 aRMS cells transfected with control shRNA (shCtrl), PAX3:FOXO1-targeted shRNA (shYFP), or targeted shRNA plus a lentivirus expressing PAX3:FOXO1 (shYFP + PAX3:FOXO1). Blots are representative of $n = 3$ biological replicates. (C) Viability of the U23674 cells described in (B) exposed to VCR at 4 nM for 24 hours. Graph plotted using GraphPad Prism. Data are means ± SD of $n = 3$ independent experiments. (D) Box-and-whisker plot showing the tumoristatic efficacy of ENT or VCR, alone and in combination, in aRMS mice at day 13 (DMSO versus ENT + VCR). Treatment with ENT at a daily dose of 5 mg/kg by intraperitoneal injection or with VCR at a dose of 1 mg/kg weekly by intraperitoneal injection, or a combination of both. Data are means ± SEM ($n = 5$ mice per cohort). ***$P < 0.001$ by log-rank test. (E) Kaplan-Meier plot of the proportion of mice with tumors smaller than 1.2 cc after treatment with ENT at a daily dose of 5 mg/kg by intraperitoneal injection or with VCR at a dose of 1 mg/kg weekly by intraperitoneal injection, or a combination of both. Treatment was stopped after day 13 because body weight loss approached 10 to 15%. Data are means ± SEM ($n = 5$ mice per cohort). *$P < 0.05$ by log-rank test. In this experiment, treatment was stopped for all mice after day 13 because body weight loss approached 10 to 15%.

aRMS cell line CW9019 harboring the t(1;13) translocation was treated with HDACis at their respective C_{max} or IC_{25} concentration. ENT and all other HDACis tested were effective at reducing PAX7:FOXO1 protein abundance to an almost undetectable level (fig. S4I).

ENT plus VCR has efficacy in aRMS PDXs

We next investigated the antitumor efficacy of ENT and VCR as single agents or in combination for various biologically independent PDX models of aRMS. The dosing details are given in table S2, and PDX model characteristics are given in tables S3 and S4. Seven of the eight models were from recurrent and/or metastatic tumors taken from biopsy or autopsy. All these contemporary models were established after 2010. Seven of eight models carried the *PAX3:FOXO1* oncogene, whereas one model (J99873/CF-2) was established from an autopsy tumor harboring *PAX7:FOXO1*. In each case, tumor growth inhibition by combined ENT plus VCR was superior to either drug alone (Fig. 2, A to G) except for one model (CTG-1008), in which tumor growth inhibition by ENT alone was better than combined treatment (Fig. 2H). Waterfall plots summarizing the tumor growth inhibition (%) for the eight models (Fig. 2I and fig. S5, A to C) showed that, in five of eight cases, ENT had single-agent activity relative to the control. No difference was seen between different treatment groups in terms of rhabdomyoblast differentiation (fig. S5, D and E, and table S5). Biochemical analysis of tumor lysates from J77636 PDXs treated with ENT showed pharmacodynamic downregulation of PAX3:FOXO1 protein abundance, which was consistent with results seen in two-dimensional cultures of human and murine aRMS cell lines (Fig. 2J). Statistical summaries of the eight aRMS PDX models are provided in tables S6 to S13.

Mechanisms underlying ENT treatment converge on known downstream genes

ENT targets class I HDACs (HDAC1, HDAC2, and HDAC3) but that at higher concentrations also inhibits the class II member HDAC10. To determine further the mechanism by which ENT silences *PAX3:FOXO1* expression, we
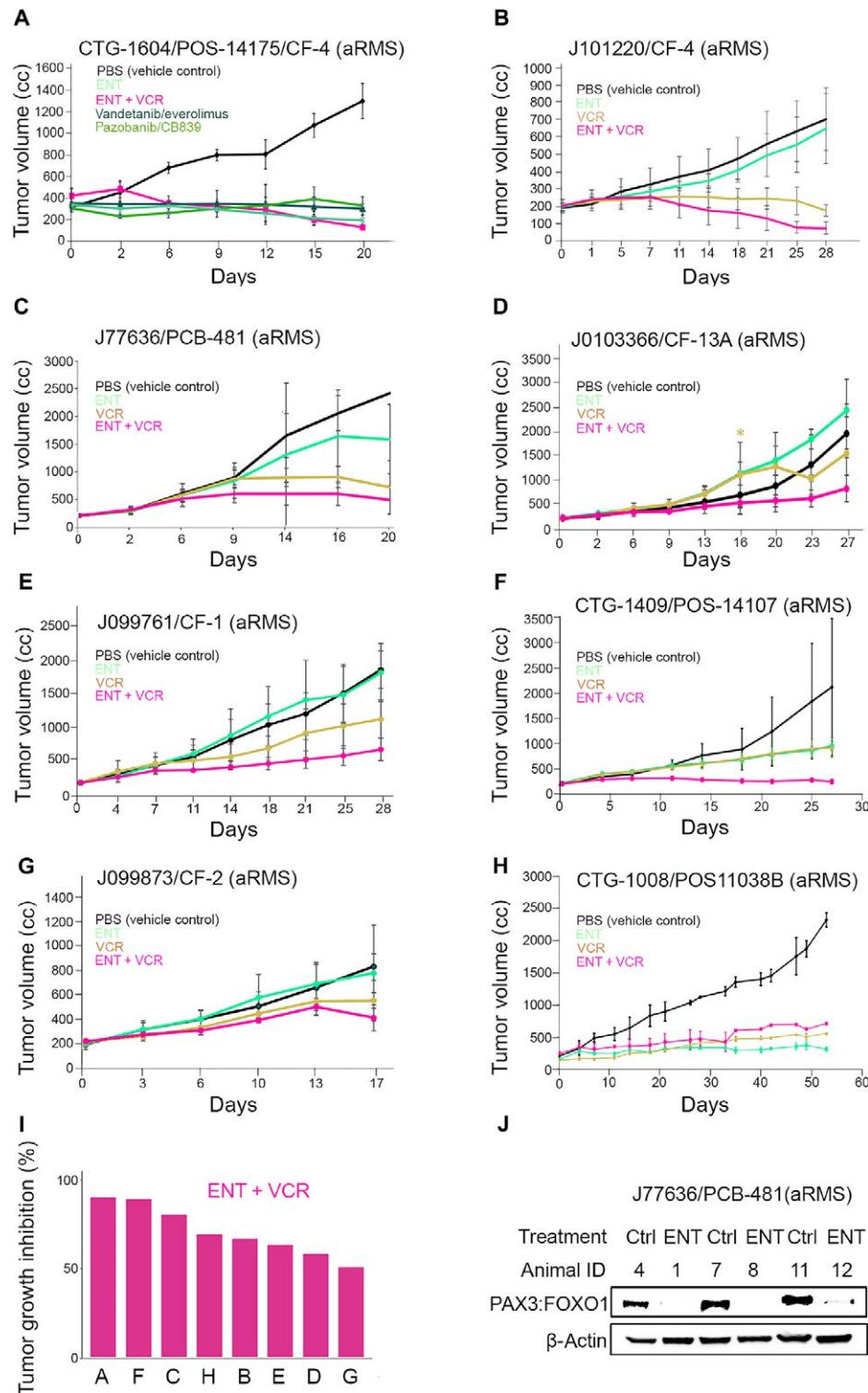
Fig. 2. In vivo evaluation of ENT and VCR in aRMS PDX. (A to H) Graphical analysis of tumor volume inhibition by either ENT alone or in combination with VCR in eight different PDX aRMS mouse models (Champions Oncology and The Jackson Laboratory) established from clinical biopsies, recurrent aRMS, or autopsies. Demographic features of these models are given in tables S3 and S4; treatment schedules are given in table S2. Statistical analyses are given in tables S6 to S13. **(I)** Waterfall plot showing tumor growth inhibition (%) for combination treatment (ENT + VCR) in the eight model models shown in (A) to (H), as labeled. **(J)** Pharmacodynamic assessment of tumor lysates from J77636 treated with ENT for PAX3:FOXO1 protein expression. Blots show *n* = 3 biological replicates (*n* represents tumor lysates from three mice per cohort). Ctrl, control.

reduced the endogenous expression of HDAC1, HDAC2, and/or HDAC3 in Rh30 cells using small interfering RNA (siRNA) (siHDAC1/2/3) alongside a scrambled siRNA control (Fig. 3A). PAX3:FOXO1 abundance was decreased to some extent by each individual HDAC siRNA, but depletion of HDAC2 or HDAC3 was the most potent (Fig. 3A). Depleting Rh30 cells of HDAC10, albeit somewhat less efficient than HDAC1/2/3 depletion, had no detectable impact on the protein abundance of PAX3:FOXO1 (Fig. 3B). These results indicated the importance of class I HDACs, specifically HDAC2 and HDAC3, for PAX3:FOXO1 expression and suggest that inhibition of specifically these HDACs is the mechanism through which ENT represses the fusion protein.

Our RNA sequencing (RNA-seq) and chromatin immunoprecipitation–exonuclease (ChIP-exo) data identified four classes of cell-autonomous gene expression changes related to ENT treatment: These subclasses were PAX3:FOXO1 binding with or without HDAC binding (Fig. 3C); HDAC1, HDAC2, HDAC3, or HDAC11 binding only (Fig. 3D); indirect targets of PAX3:FOXO1 (Fig. 3E); and otherwise-regulated genes (Fig. 3F) (*13*). Although HDAC-related and HDAC-unrelated changes in expression of the known PAX3:FOXO1 direct and indirect targets were expected, neither these nor those of the otherwise-regulated genes could explain altered transcription of *PAX3:FOXO1* with respect to *PAX3* regulatory elements.

To understand the mechanisms of action for ENT and PAN, we also analyzed the expression of wild-type PAX3 and FOXO1 in U23674 and Rh30 aRMS cells, as well as C2C12 mouse myoblasts, human skeletal muscle myoblasts (HSMMs), and CureFast-1 (CF-1) PDX explant cell culture each treated separately with ENT and PAN. Consistent with previous results (*13*), PAX3:FOXO1 was significantly reduced by ENT but not by PAN in U23674 cells (Fig. 4, A and B). These cells do not express detectably sufficient wild-type PAX3 for analysis (Fig. 4A); however, in C2C12 cells, which do so, ENT significantly reduced wild-type PAX3 but not wild-type FOXO1 abundance (Fig. 4, A and C), and in both cell lines, neither ENT nor PAN altered FOXO1 (Fig. 4, A, D, and E).
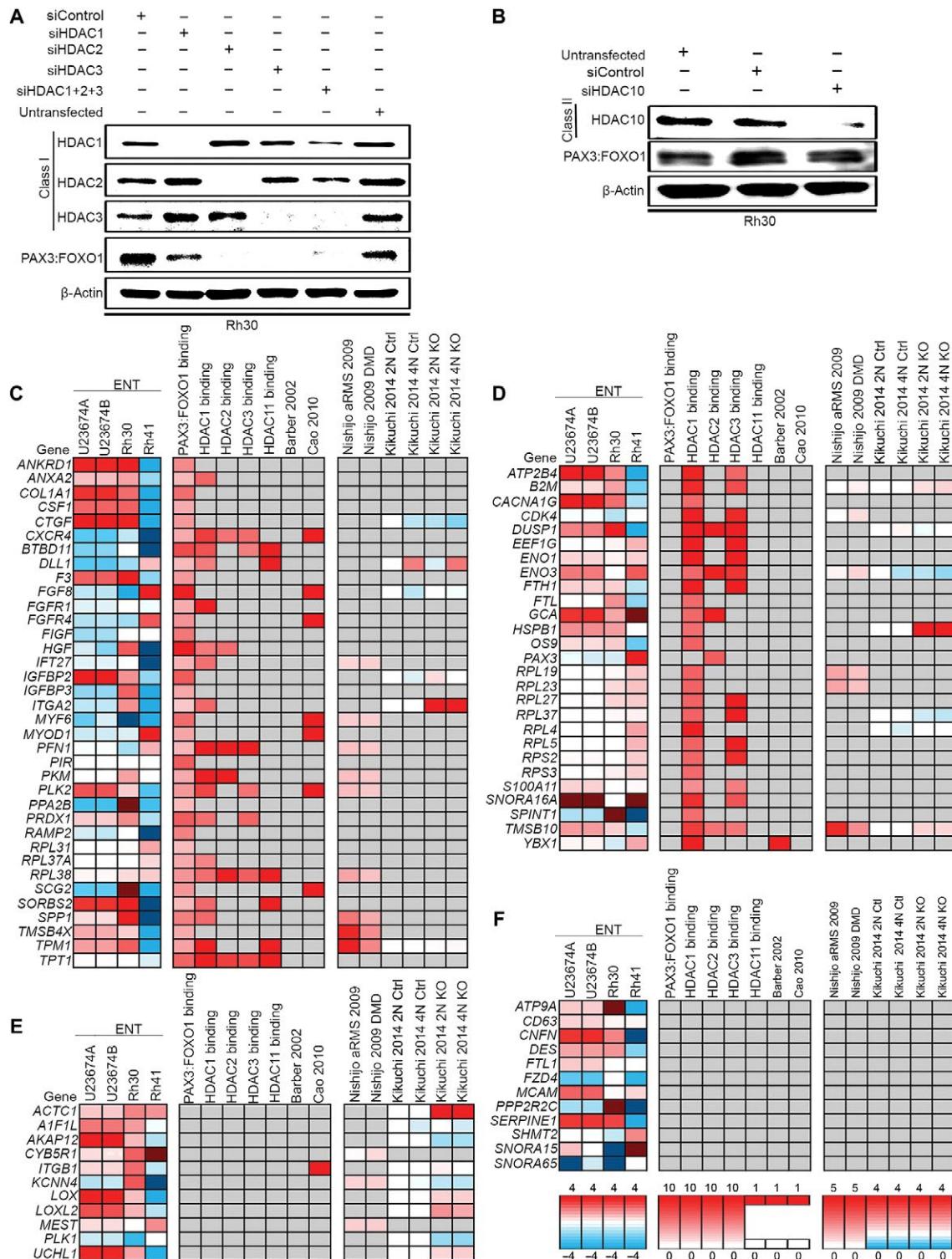
Fig. 3. siRNA-mediated knockdown of HDAC and gene expression and PAX3:FOXO1 binding data for key aRMS gene targets. (**A** and **B**) PAX3:FOXO1 expression in Rh30 cells transfected with siRNA at 100 nM for 72 hours, targeting HDAC1, HDAC2, and/or HDAC3 (A), as well as HDAC10 (B). Blots are representative of n = 3 independent experiments. (**C** to **F**) Three aRMS samples [U23674 in two replicates (U23674A and U23674B), Rh30, and Rh41] were sequenced after treatment with ENT at 2 μM for 72 hours or DMSO. Featured genes were decreased (log$_2$ ratio of ENT-induced expression divided by control expression > 1) in all samples. Additional expression data for these key targets were curated from previous publications (10, 33), and PAX3:FOXO1 binding data were curated from the literature as indicated (31, 32) or were generated through chromatin immunoprecipitation sequencing (ChIP-seq) experiments. Four subclasses of ENT-induced gene expression were identified as (C) PAX3:FOXO1 binding with or without HDAC binding; (D) HDAC1, HDAC2, HDAC3, or HDAC11 binding only; (E) indirect targets of PAX3:FOXO1; and (F) otherwise-regulated genes.
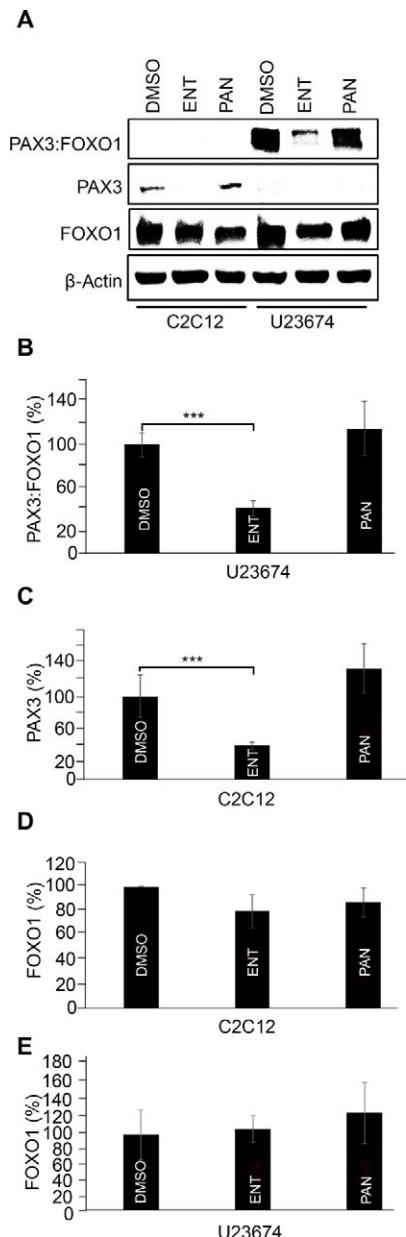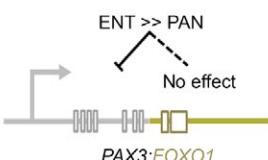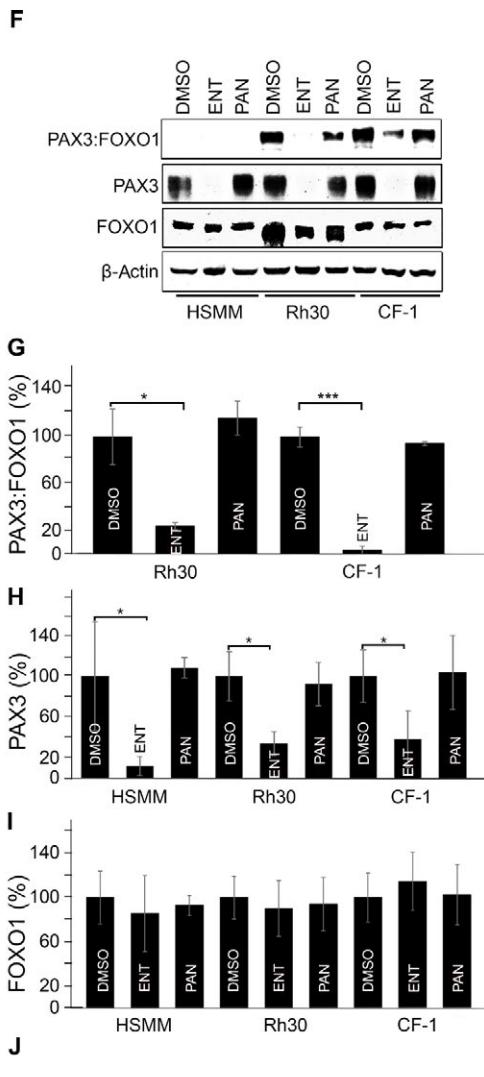
Fig. 4. Wild-type PAX3 expression altered by ENT and not by PAN. (**A**) Immunoblots of PAX3:FOXO1, wild-type PAX3, and wild-type FOXO1 in murine C2C12 myoblasts and U23674 aRMS cells upon 72 hours of treatment with ENT (1 μM) and PAN (45 nM). Blots are representative of n = 3 independent experiments. Note that, in the murine aRMS GEM culture U23674, both wild-type *PAX3* alleles are homozygously replaced by *PAX3:FOXO1*, but both *FOXO1* alleles are intact (homozygous wild type). (**B** to **E**) Densitometric analysis of (A) PAX3:FOXO1, wild-type PAX3, and wild-type FOXO1 expression upon treatment with ENT and PAN. Data are means ± SD (n = 3 independent experiments). ***P < 0.001, two-sided Student's t test. (**F**) Western blots of PAX3:FOXO1, wild-type PAX3, and wild-type FOXO1 in HSMMs, Rh30, and CF-1 cells upon 72 hours of treatment with ENT (1 μM) or PAN (45 nM). Blots are representative of n = 3 biological replicates. In contrast to murine aRMS cultures, *PAX3* is heterozygous (wild type/PAX3:FOXO1) and *FOXO1* is heterozygous (wild type/null) in human aRMS cultures, and both are homozygous wild type in HSMMs. (**G** to **I**) Densitometric analysis of PAX3:FOXO1, wild-type PAX3, and wild-type FOXO1 protein expression upon treatment with ENT. Data are means ± SD (n = 3 independent experiments). *P < 0.05 and ***P < 0.001, two-sided Student's t tests. (**J**) A diagrammatic representation attributing the effect of ENT on PAX3:FOXO1 expression to *PAX3* cis-regulatory elements.

Similar results were observed in Rh30, CF-1, and HSMM cells (Fig. 4, F to I). These results suggest that, for *PAX3:FOXO1* in aRMS tumor cells, the *PAX3* cis-regulatory elements are affected differently by ENT treatment versus PAN treatment, perhaps attributable to

secondary effects on *PAX3* cis-elements or *PAX3* mRNA stability factors (Fig. 4J).

Loss of SMARCA4 expression or activity increases, or derepresses, the expression of the microRNA miR-27a

To further understand how ENT silences the *PAX3:FOXO1* gene, whereas PAN does not, RNA-seq analysis of ENT-treated versus PAN-treated human and murine aRMS cells was carried out with an emphasis on assessing the expression of transcription factors and chromatin remodeling complexes. Differential expression across samples after ENT treatment versus after PAN treatment identified key up-regulated and down-regulated genes in the ENT-specific effects (Fig. 5A). Many epigenetic targets, including chromatin-modifying enzymes and remodeling complexes, were indeed decreased by ENT treatment and increased by PAN treatment in Rh30 and U23674 cells (Fig. 5A). Specifically, ENT treatment had a strong effect on the expression of many of the SWI/SNF (switch/sucrose non-fermentable) complexes, which included that of *SMARCA4/ BRG1*, which was markedly reduced by ENT but not by PAN treatment (Fig. 5A). At the protein level, only ENT reduced SMARCA4 protein abundance in both cells, notably to an undetected level (Fig. 5B). Consistent with ENT-mediated pharmacological inhibition of class I HDACs, siRNA-mediated knockdown of HDAC3 decreased SMARCA4 protein abundance to an almost undetectable level; knockdown of HDAC2 had markedly less of an effect (Fig. 5C). This finding concurs with a previous report indicating a noncanonical role of HDAC3 as a transcriptional activator (*17*). Furthermore, the abundance of the PAX3:FOXO1 fusion protein was decreased in Rh30 cells upon reduction of SMARCA4 by siRNA (siSMARCA4) (Fig. 5D), altogether suggesting an HDAC3-SMARCA4 circuit in promoting PAX3:FOXO1 fusion protein abundance in aRMS cells.

Upon searching the miR search engine miRBase for microRNA (miRNA) that bound the *PAX3* promoter or cis-elements, we observed that miR-27a was unique in that, in both mouse and human cells, it bound *PAX3* upstream (5′) of the *PAX3: FOXO1* breakpoint (in intron 4). miR-27a-5p reportedly targets *PAX3*, as well as *PAX7*, in developing and adult muscle cells (*18*).
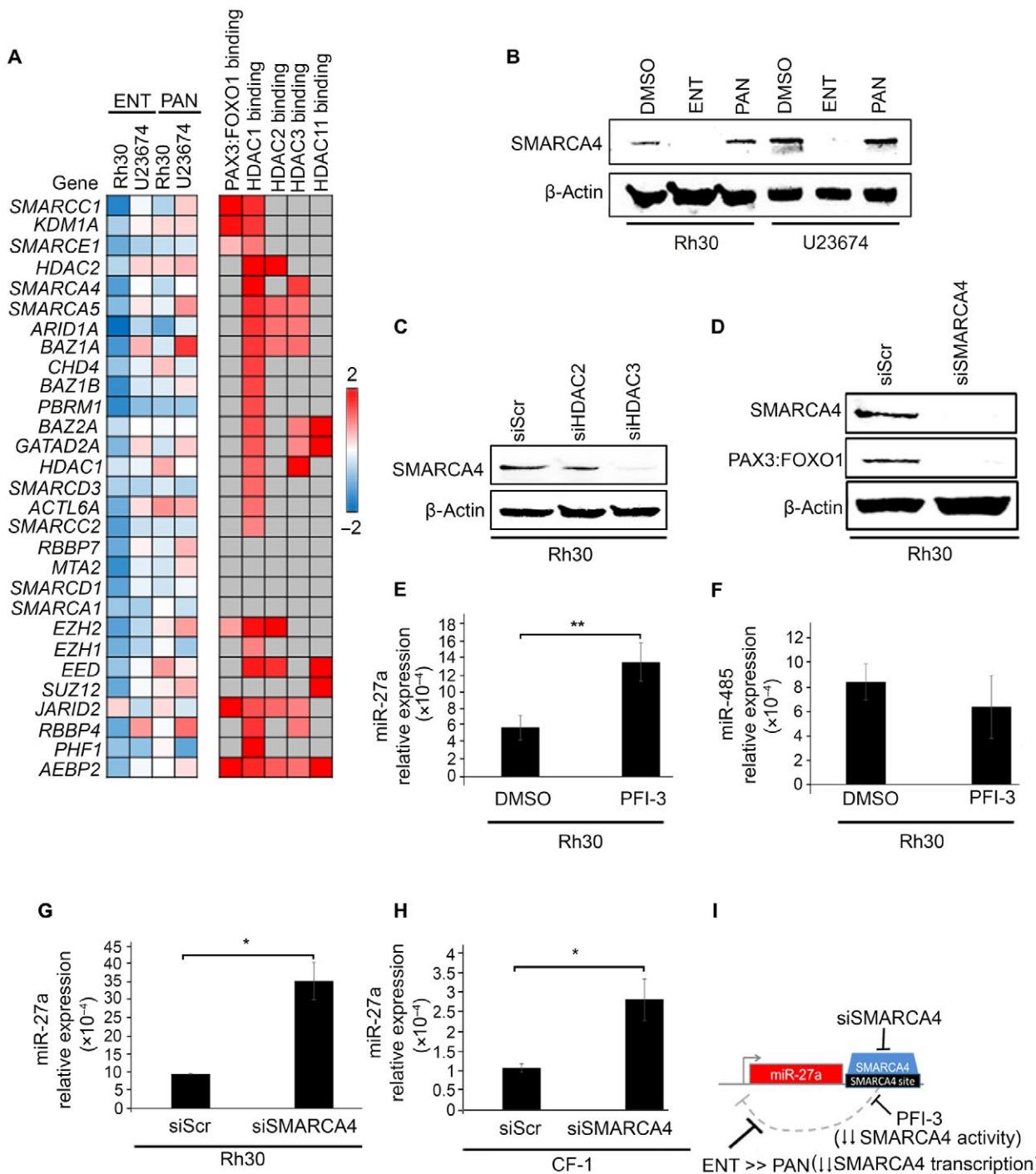
Fig. 5. Loss of SMARCA4 expression or activity derepresses miR-27a. (**A**) Differential expression analysis by RNA-seq for genes associated with chromatin-modifying complexes in Rh30 and U23674 cells after treatment for 72 hours with ENT (1 μM) or PAN (45 nM) relative to cells treated with DMSO. SMARCA4 average transcripts per million for DMSO-treated cells, ENT-treated cells, and PAN-treated cells were 86.86, 56.15, and 72.19, respectively. (**B**) SMARCA4 protein abundance in Rh30 and U23674 aRMS cultures treated with ENT or PAN. Blots are representative of $n = 3$ independent experiments. (**C**) SMARCA4 protein abundance in Rh30 cells transfected with HDAC2- or HDAC3-targeted siRNA (100 nM for 72 hours). Blots are representative of $n = 3$ independent experiments. (**D**) PAX3:FOXO1 protein abundance in Rh30 cells transfected with SMARCA4-targeted siRNA (100 nM for 72 hours). Blots are representative of $n = 3$ independent experiments. (**E**) Quantitative polymerase chain reaction (qPCR) analysis of miR-27a expression in Rh30 cells treated with the SMARCA4 bromodomain inhibitor PFI-3 (10 μM for 24 hours). (**F**) qPCR analysis of miR-485 in Rh30 cells treated with PFI-3. **$P < 0.01$, two-sided Student's t test. (**G** and **H**) qPCR analysis of miR-27a expression in Rh30 and CF-1 cells transfected with SMARCA4-targeted siRNA (100 nM for 72 hours). Data were normalized to the expression of U6snRNA. Gene expression was quantified using the $2^{-\Delta Ct}$ method. Data are means ± SD ($n = 3$ independent experiments each in triplicate). *$P < 0.05$, two-sided Student's t test. (**I**) Diagrammatic representation of how ENT and PFI-3 disrupt SMARCA4-mediated interference of *miR-27a* transcription.
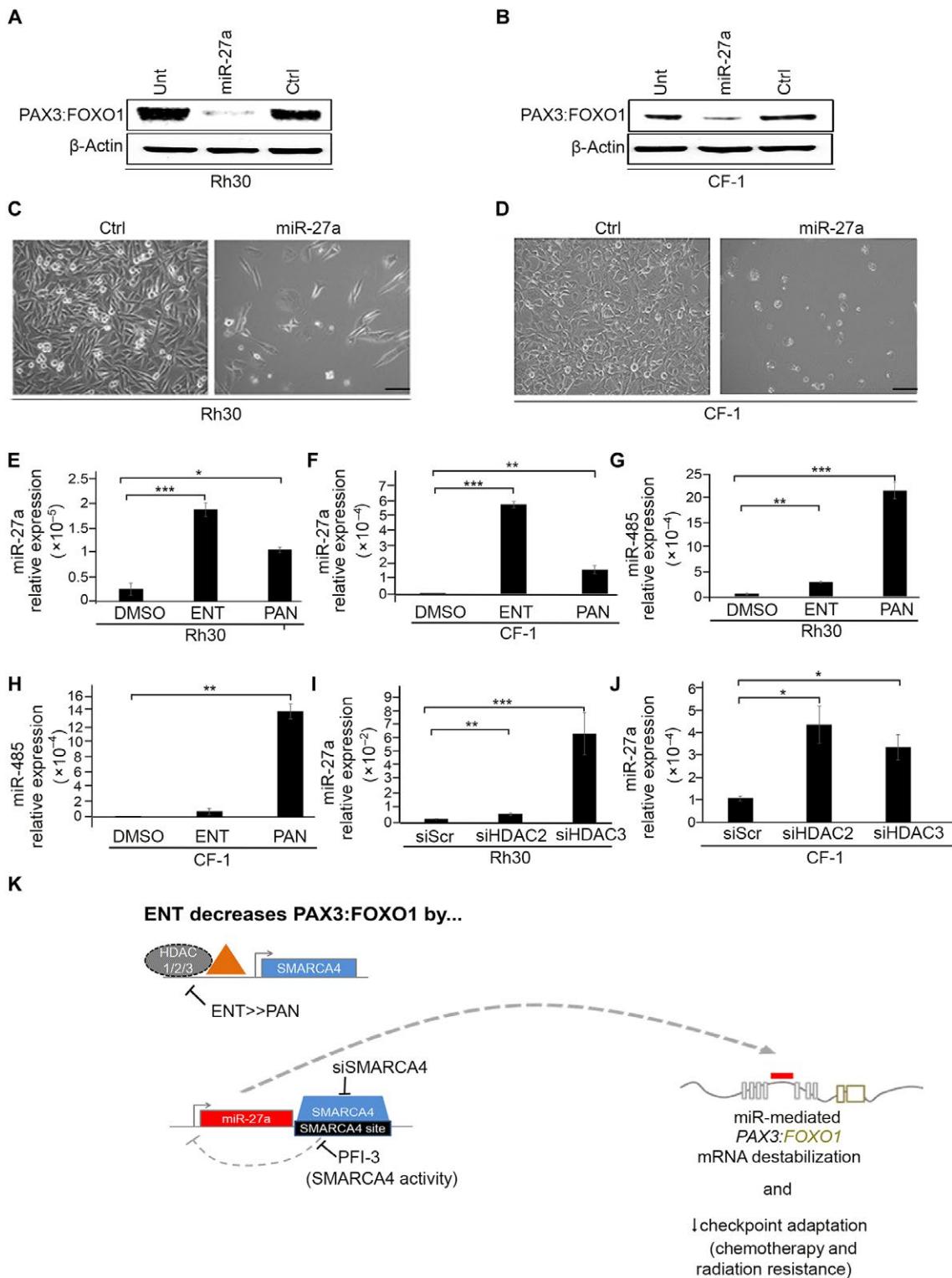
Fig. 6. miR-27a overexpression silences PAX3:FOXO1. (**A** and **B**) Western blot of PAX3:FOXO1 protein abundance in Rh30 and CF-1 cells transfected with mimics of miR-27a (10 µM for 72 hours) or a negative control (Ctrl). Also shown is blotting in lysates from untransfected control cells (Unt). Blots are representative of $n = 3$ independent experiments. (**C** and **D**) Light microscopy images of Rh30 and CF-1 cells transfected with mimics of miR-27a. Images are representative of $n = 3$ independent experiments. Scale bars, 50 µM. (**E** to **H**) qPCR analysis of miR-27a (E and F) or miR-485p (G and H) expression in Rh30 and CF-1 cells treated for 72 hours with ENT (1 µM) or PAN (45 nM). (**I** and **J**) qPCR of miR-27a expression in Rh30 and CF-1 cells transfected with HDAC2- or HDAC3-targeted siRNA (100 nM). Data were normalized to U6snRNA expression. Gene expression was quantified using the $2^{-\Delta Ct}$ method. Data are means ± SD ($n = 3$ independent experiments each in triplicate). *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, two-sided Student's t tests. (**K**) Summary of the HDAC3–SMARCA4–miR-27a–PAX3:FOXO1 regulatory circuit targeted by ENT.

For comparison, we chose miR-485-5p as a conserved miRNA binding the *PAX3* 3′ untranslated region (3′UTR). Referencing the University of California, Santa Cruz (UCSC) genome browser visualization of ChIP-seq data from the Encyclopedia of DNA Elements (ENCODE) predicted that *SMARCA4* binds to the 3′UTR of miR-27a. We then measured miR-27a expression upon pharmacological inhibition of SMARCA4 bromodomain activity using the small-molecule inhibitor PFI-3 at a concentration of 10 µM. PFI-3 inhibits 95% of the SMARCA4 bromodomain activity at this concentration according to analysis by time-resolved fluorescence energy transfer (TR-FRET; fig. S6). Inhibition of SMARCA4 bromodomain activity increased miR-27a expression (Fig. 5E) but not that of miR-485 (Fig. 5F). We then carried out siRNA-mediated knockdown of SMARCA4 in Rh30 cells and CF-1 cells, which showed an increase in miR-27a compared to scrambled siRNA control, consistent with the effect seen on miR-27a upon pharmacological inhibition of SMARCA4 (Fig. 5, G and H). Together, these results suggested that a SMARCA4–miR-27a regulatory circuitry controls the expression of PAX3:FOXO1 (Fig. 5I).

ENT-induced reexpression of miR-27a silences PAX3:FOXO1

To understand whether ENT silences *PAX3:FOXO1* transcript via miR-27a–mediated transcription interference, we reexpressed miR-27a in Rh30 and CF-1 cells by transient transfection of miR-27a mimics and examined PAX3:FOXO1 protein expression. Compared to a control oligonucleotide (a "negative mimic"), miR-27a reexpression decreased PAX3:FOXO1 abundance in both Rh30 and CF-1 cells (Fig. 6, A and B), which also exhibited altered cell morphology (Fig. 6, C and D). qPCR analysis then showed a more potent increase in the expression of miR-27a in ENT-treated than in PAN-treated Rh30 and CF-1 cells (Fig. 6, E and F). However, PAN increased miR-485 expression more potently than did ENT (Fig. 6, G and H). Because our earlier studies showed that ENT reduced PAX3:FOXO1 expression by inhibiting HDAC2/3 (Fig. 3A), we silenced HDAC2/3 by siRNA in Rh30 and CF-1 cells, which resulted in increased miR-27a expression relative to that in scrambled siRNA control cells (Fig. 6, I and J). These results altogether suggest that, by inhibiting HDAC3 and maybe in part HDAC2, ENT blocks epigenetic suppression of an miRNA, thereby enabling its restriction of PAX3:FOXO1 expression and reducing the growth of aRMS tumors (Fig. 6K).

DISCUSSION

Our previous studies uncovered that cell of origin conveys an epigenetic memory to tumor cells, leading to differences in histological phenotype and drug sensitivity, and that cell of origin also epigenetically influences the transcription of the *PAX3:FOXO1* oncogene in tumor cells (*13*). Apropos to those findings, mechanistic studies that we performed here suggest that HDAC2/HDAC3 inhibition is the primary mechanism of ENT-mediated repression of PAX3:FOXO1 in aRMS. We also found that ENT and PAN, despite both being HDACis, are very different drugs with different global actions and specifically different activity with respect to SMARCA4, the inhibition of which we have now shown is critical to miR-27a expression and *PAX3:FOXO1* mRNA destabilization.

Deregulated expression and function of miRNAs has been reported in many cancers, including RMS. HDACis and ENT specifically have been previously reported to alter miRNA expression (*19*, *20*). In this study, we found an miRNA that inhibited *PAX3:*

FOXO1 mRNA translation and/or its stability. Our focus was on miRNAs that could bind either the *PAX3* promoter and exons 1 to 8 or the *FOXO1* exons 1 to 3 and its 3′UTR, as well as those that were conserved across humans and mice, given that ENT reduced both human and mouse PAX3:FOXO1. miR-27a-5p targets both *PAX3* and *PAX7* (*18*); miR-27a's own promoter is regulated in muscle by the myogenic transcription factors MYF6, MYOD, and MEF2C (which aligns with RMS features) (*21*); and SMARCA4 is predicted by ENCODE to bind the 3′UTR of miR-27a. Knockdown of SMARCA4 or pharmacological inhibition of its bromodomain activity increased the expression of miR-27a. Although SMARCA4 is typically characterized as a transcriptional activator, reports suggest that SMARCA4 can also act as a repressor (*22*, *23*). How precisely SMARCA4 represses miR-27a requires further investigation and may reveal insight into its function in other contexts.

From a clinical point of view, ENT in combination with the chemotherapy agent VCR showed strong antitumor activity in aRMS orthotopic mouse models at clinically achievable adult drug concentrations. For the purposes of clinical trial planning, we have also established the comparative efficacy of ENT to other HDACis at reducing PAX3:FOXO1 and PAX7:FOXO1 expression. ENT not only reduced PAX3:FOXO1 abundance across species in both murine and human aRMS cell lines/cultures and PDX explant cell cultures but also inhibited PAX3:FOXO1 levels more effectively than the other HDACis tested. We note that, in the time since our studies were performed, Malempati *et al.* (*24*) have reported the phase 1 clinical trial results for ENT in pediatric patients, showing higher drug exposures and decreased clearance without any additional toxicity for children versus adults (C_{max} of 140.8 nM at a dose of 4 mg/m^2 and a half-life of 45 hours in children). Thus, our preclinical studies may be overly stringent with respect to clinical modeling.

In summary, our preclinical data with strong mechanical evidence suggest that targeting HDAC3–SMARCA4–miR-27a–PAX3:FOXO1 regulatory circuitry in aRMS may provide real therapeutic benefits for patients with aRMS. The preclinical efficacy in vivo and overall biochemical performance at PAX3:FOXO1 suppression make ENT a promising drug candidate for combining with chemotherapy and/or radiation in clinical trials for aRMS. Our findings directly support the pediatric phase 1B clinical trial ADVL1513 concept under review for the use of ENT as a single agent in RMS and possible phase 2 chemotherapy-ENT trials to follow.

MATERIALS AND METHODS
Cell culture

Murine primary tumor cell cultures (U23674) were generated as described previously (*25*). Human aRMS cell lines Rh30 and Rh41 were cultured in growth medium (GM) RPMI 1640 (Thermo Fisher Scientific, catalog no. 11875-093) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific, catalog no. 26140079) and 1% penicillin/streptomycin (Thermo Fisher Scientific, catalog no. 15140-122). CW9019 was cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma, catalog no. 11965-084) supplemented with 10% FBS and 1% penicillin/streptomycin. The mouse myoblast cell line C2C12 was obtained from the American Type Culture Collection (ATCC) and was cultured in DMEM supplemented with 20% FBS and 1% penicillin/streptomycin. Primary HSMMs (Lonza Inc., catalog no. CC-2580) were cultured in GM (Cell Applications, catalog no. 151-500). Phoenix-Amphotropic cells

(a second-generation retrovirus producer cell line) (ATCC, catalog no. CRL-3213) were cultured in DMEM supplemented with 10% FBS and 1% antibiotics. PDX explant cell culture CF-1 was cultivated in RPMI 1640 supplemented with 10% FBS and 1% antibiotics. All cells were incubated at 37°C and 5% CO_2. Rh30 and Rh41 were obtained from the Children's Oncology Group (www.COGcell.org) and were authenticated by Short Tandem Repeat validation assay through Biosynthesis. The murine cell culture U23674 was authenticated by PCR validation by Transnetyx. CF-001 was authenticated by analyzing the expression of PAX3:FOXO1.

VCR chemotherapy, HDAC inhibition, and SMARCA4 inhibition

The chemotherapy agent VCR sulfate, used here in the treatment of Rh30 and U23674 cells, was obtained from Sigma (product no. V8879). HDACis used in this study were purchased from Selleckchem: ENT (also known as MS-275, a class I HDACi; catalog no. S1053), PAN (a broad-spectrum HDACi; catalog no. S1030), SAHA (catalog no. S1047), CUDC-907 (a PI3K inhibitor and a class I and class II HDACi; catalog no. S2759), and CUDC-101 (a class I and class II HDACi and an inhibitor of epidermal growth factor receptor family members EGFR and HER2; catalog no. S1194). aRMS cell lines and primary tumor cell cultures were treated with these drugs at their clinically relevant C_{max} (maximum plasma concentrations) or, where C_{max} is not yet reported, their determined IC_{25}.

TR-FRET assay of the treatment of cells with the SMARCA4 inhibitor

The SMARCA4 bromodomain activity inhibitor, PFI-3 (S7315), was purchased from Selleckchem. Rh30 cells were treated with PFI-3 at a concentration of 10 μM. PFI-3 inhibits 95% of the SMARCA4 bromodomain activity at 10 μM based on the TR-FRET assay carried out by a commercial vendor, BPS Bioscience. Briefly, the assay was performed by TR-FRET technology using recombinant bromodomain and BET ligand. The TR-FRET signal from the assay is correlated with the amount of ligand binding to the bromodomain. Binding experiments were performed in duplicate at each concentration. The TR-FRET data were analyzed using the computer software GraphPad Prism (GraphPad Software Inc.). In the absence of the compound in wells containing BET ligand, the TR-FRET signal (Ft) in each dataset was defined as 100% activity. Select wells where control inhibitors were more than 100-fold the IC_{50} were used to define the TR-FRET signal (Fb) as 0% activity. The percent activity in the presence of each compound was calculated according to the following equation: % activity = $[(F - Fb)/(Ft - Fb)] \times 100$, where F = the TR-FRET signal in the presence of the compound. The percent inhibition was calculated according to the following equation: % inhibition = 100 − % activity.

RNA-extraction and RT-PCR

Rh30 (human aRMS) cell line and murine aRMS primary tumor cell cultures (U23674) were treated with 0.1, 0.2, 0.4, 0.8, 1, and 2 μM ENT for 24 hours. DMSO treatment was used as a control. After treatment with ENT, total RNA was extracted and complementary DNA (cDNA) was synthesized as previously described (13). Expression of *PAX3:FOXO1* was determined by real-time PCR (RT-PCR) using custom Taqman primers and probe (catalog nos. 4304970 and 4316034) on a StepOnePlus RT-PCR machine (Applied Biosystems).

Rh30 and Rh41 cell lines were treated with 1 μM ENT, 45 nM PAN, and 1 μM SAHA for 24 hours. DMSO-treated cells were used as control. Total RNA was extracted from cells using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. cDNA was prepared from RNA using a high-capacity cDNA reverse transcription kit (Applied Biosystems) with a ribonuclease inhibitor. qPCR was performed using TaqMan Universal Master Mix, no AmpErase Uracil-N glycosylase on the ABI 7900HT Fast Real-Time PCR System (Applied Biosystems). Primers used were Gapdh-Hs02758991_g1 and PAX3:FOXO1-Hs03024825_ft. Gene expression was quantified using the $2^{-\Delta Ct}$ method.

Immunoblotting

Human Rh30 and Rh41 cell lines and murine aRMS primary tumor cell cultures (U23674) were treated with 1 μM ENT, 45 nM PAN, 1 μM SAHA, 150 nM CUDC907, or 150 nM CUDC-101 for 72 hours. DMSO (vehicle)–treated cells were used as control. After 72 hours, lysates were collected using radioimmunoprecipitation assay lysis buffer with Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, catalog no. 78440) and analyzed for PAX3:FOXO1 expression using an anti-PAX3 antibody (1:400; R&D Systems, catalog no. MAB2457), PAX7:FOXO1 expression using an anti-FKHR antibody (H-128, 1:300; Santa Cruz Biotechnology), and SMARCA4 expression using an anti-SMARCA4 antibody (G-7, 1:300; Santa Cruz Biotechnology) and matched for protein expression using an anti–β-actin antibody (1:10,000; Abcam, catalog no. ab8227). All blots were visualized using a FluorChem Q system (ProteinSimple). To analyze endogenous wild-type FOXO1 expression, recombinant FOXO1 (NM_002015) (Origene) was used. FOXO3 (NM_001455) was used as negative control to choose the appropriate antibody (fig. S7). Anti-FOXO1 antibody (1:500; Millipore, clone 2H8.2) detects recombinant FOXO1 without background compared to other antibodies and subsequently used for all immunoblots (fig. S7).

Orthotopic allograft studies

All animal studies were conducted with Institutional Animal Care and Use Committee (IACUC) approval at the Oregon Health & Science University. An orthotopic allograft mouse model of aRMS (U23674, genotype *Myf6Cre, PAX3:FOXO1, p53*) was generated as previously described (13). Mice were treated with ENT at a daily dose of 5 mg/kg by intraperitoneal injection or with VCR sulfate at a dose of 1 mg/kg weekly by intraperitoneal injection, or a combination of both. Treatment was started once the tumors reached 0.25 cc and ended when the tumors reached 1.5 cc. During treatment, mice that experienced body weight loss (10 to 15%) were euthanized early. For the radiation studies, murine aRMS primary tumor cell cultures (U23674) were pretreated with either DMSO or 2 μM ENT and then subjected to 10 Gy radiation before injecting 500,000 viable cells into each mouse. Failure was defined as an event for tumor size greater than or equal to 1.2 cc.

PDX models at Champions Oncology

The Champions Personalized Tumorgraft chemosensitivity tests were conducted using a TumorGraft model established from two independent RMS biopsy specimens. The explants were received and immediately implanted into immunodeficient mice. The antitumor activity of ENT and VCR was tested in a low-passage immune-compromised female mice (Harlan; *nu/nu*) between 5 and 8 weeks of age housed

on irradiated papertwist-enriched $\frac{1}{8}$-inch corncob bedding (Shepherd) in individual HEPA (high-efficiency particulate air) ventilated cages (Innocage IVC, Innovive, USA) on a 12-hour light/12-hour dark cycle at 68° to 74°F (20° to 23°C) and 30 to 70% humidity. Animals were fed water ad libitum (reverse osmosis, 2 parts per million Cl2) and an irradiated test rodent diet (Teklad 2919) consisting of 19% protein, 9% fat, and 4% fiber. All compounds were formulated according to the manufacturer's specifications. Beginning day 0, tumor dimensions were measured twice weekly by a digital caliper, and data including individual and mean estimated tumor volumes (mean TV ± SEM) were recorded for each group; tumor volume was calculated using the following formula (1): $TV = width^2 \times length \times 0.52$. All studies were done with the approval of Champions Oncology IACUC.

PDX models at The Jackson Laboratory

To establish each PDX model, NSG (NOD.Cg-*Prkdcscid IL2rgtm1Wjl/SzJ*) mice were obtained by The Jackson Laboratory. Tumor explants were obtained from the patients and immediately implanted into the rear flanks of recipient female NSG (JAX, stock no. 005557) mice using a trochar. Once tumors reached about 2000 mm^3, they were collected and passaged for serial transplantation in NSG mice to create low-passage fragments or cohort for future studies. The tumor volume range for enrollment was 150 to 250 mm^3. All studies were done with the approval of The Jackson Laboratory IACUC. Mice were treated with vehicle or ENT/VCR as a single agent or a combination at doses and routes of administration provided in table S2 until tumors reached 2000 mm^3 or reached study day 28. The antitumor activity of ENT and VCR was tested. All compounds were formulated according to the manufacturer's specifications. Beginning day 0, tumor dimensions were measured twice weekly by a digital caliper, and data including individual and mean estimated tumor volumes (mean TV ± SEM) were recorded for each group; tumor volume was calculated using the following formula: $TV = width^2 \times length/2$.

siRNA-mediated silencing of class I and class II HDACs

For silencing of class I HDACs (1–3), Rh30 cells were transfected with 100 nM HDAC1 (L-003493-00-0005, 5 nmol), HDAC2 (L-003495-02-0005, 5 nmol), HDAC3 (L-003496-00-0005, 5 nmol), or HDAC10 (L-004072-00-0005, 5 nmol) SMARTpool siRNA reagent (a pool of four siRNA duplexes all designed to target distinct sites within the specific gene of interest) (Dharmacon) versus scrambled siRNA (Dharmacon) using Lipofectamine RNAimax (Invitrogen, catalog no. 13778030). siRNA cell lysates were subjected to Western blotting using anti-HDAC1 (H-51, 1:400; Santa Cruz Biotechnology), anti-HDAC2 (C-8, 1:400; Santa Cruz Biotechnology), anti-HDAC3 (B-12, 1:400; Santa Cruz Biotechnology), and anti-HDAC10 (E-2, 1:400; Santa Cruz Biotechnology) antibodies and PAX3:FOXO1 expression using anti-PAX3 antibody (R&D Systems).

Retroviral infection

To restore expression of PAX3:FOXO1 in U23674 with stable knockdown of PAX3:FOXO1 (U23674 shYFP), retroviral vector pk1-PAX3:FOXO1 was transfected into Phoenix-AMPHO packaging cells using Lipofecatmine Plus reagent (Invitrogen, catalog no. 15338100). Medium was harvested after 24, 48, and 72 hours and filtered with a 0.45-μm syringe filter. Retroviral supernatant with polybrene was used for transduction of U23674 shYFP culture.

siRNA-mediated silencing of SMARCA4

For silencing of SMARCA4, Rh30 cells were transfected with 100 nM SMARCA4 (L-010431-00-0005, 5 nmol) SMARTpool siRNA reagent (Dharmacon) versus scrambled siRNA (Dharmacon) using Lipofectamine RNAimax (Invitrogen). Cell lysates were subjected to Western blotting using anti-SMARCA4 (G-7, 1:300; Santa Cruz Biotechnology).

miRNA overexpression

For overexpression of *miR-27a-5p*, Rh30 and CF-1 cells were transfected with 10 μM miR-27-a-5p mimics (assay name, hsa-miR-27a-5p; accession no., MI0000085) or miR-27a-5p negative control (5 nmol; Thermo Fisher Scientific, catalog no. 4464058) using Lipofectamine RNAimax (Invitrogen). Cell lysates were collected at the end of 72 hours and subjected to Western blotting for analysis of PAX3:FOXO1 expression using an antibody against PAX3 (R&D Systems).

miRNA isolation and RT-PCR

miRNA isolation was performed as per the manufacturer's protocol (Invitrogen, catalog no. K157001). Briefly, Rh30 and CF-1 cells were seeded at a density of 0.3 million cells and, the next day, were treated with ENT (1 μM) or DMSO (vehicle). At the end of 24 hours of treatment, miRNA was isolated as per the manufacturer's standard protocol. The recovery tube with RNA was then quantified using a BioTek microvolume plate. Complementary DNA was prepared from 5 ng of miRNA using a TaqMan miRNA reverse transcriptase kit (Thermo Fisher Scientific, catalog no. 4366596). qPCR was performed using TaqMan Universal Master Mix, no AmpErase UNG (Thermo Fisher Scientific, catalog no. 4440040) on the Bio-Rad Thermocycler Real-Time PCR System. Primers used were hsa-miR-27a and hsa-miR-485 (TaqMan miRNA assays). U6snRNA (TaqMan miRNA assays) was used as loading control. miRNA expression was quantified using the $2^{-\Delta Ct}$ method.

RNA sequencing

For the identification of transcriptional changes (Fig. 3, C to F) in aRMS cells after ENT treatment, each of the three aRMS cultures (human cell lines Rh30 and Rh41 and mouse cell culture U23674) was treated with ENT alongside a paired untreated sample for a fixed time period. All samples were treated with ENT for 72 hours, except for Rh41, which was treated for 24 hours because of higher sensitivity to ENT. All cells were cultured on 10-cm dishes, and treatment began when plates were 60% confluent. Passages lower than 7 were used for all mouse cultures. RNA isolation and sequencing were performed by the commercial service provider, Beijing Genomics Institute (BGI). Differential expression for a single sample was defined as post-ENT treatment expression divided by vehicle-treated expression. These criteria identified 348 overexpressed and 358 underexpressed genes in aRMS. Differential expression of genes associated with chromatin modification and remodeling after ENT and PAN treatment (72 hours) (Fig. 5A) was analyzed by one more set of RNA-seq with RNA isolation and sequencing performed by BGI. Log$_2$ scaled ratios of ENT-treated versus vehicle-treated RH30 and U23674 and PAN-treated versus vehicle-treated RH30 and U23674 were organized by change in regulation.

Bioinformatic analysis of RNA-seq

The paired-end raw reads for all RNA-seq data samples were aligned using TopHat version 2.0.9. Up to two mismatches in the alignment

were permitted before a read alignment was discarded. The reads for human samples were aligned to the UCSC hg36 human reference genome. The reads for mouse samples were aligned to the UCSC mm19 mouse reference genome. The aligned reads were assembled into transcripts using Cufflinks version 2.2.1. Differential comparisons were performed by the Cuffdiff function of Cufflinks version 2.2.1. Differential comparisons were made between samples treated with ENT and untreated control samples. For the differential analysis of treated versus untreated samples, the standard Cuffdiff parameters were used. Reported in all figures are $\log_2$ scaled differential values. Genes for which treated samples had nonzero expression and untreated samples had zero expression fixed were set to the minimum of the treated expression and 1024, before $\log_2$ scaling, to prevent divide-by-zero errors during downstream analysis. Only genes with quantified expression across all samples (total of 14,575 genes) were considered. $\log_2$ scaled mean differential expression across all samples of a common subtype was used to identify overexpression and underexpression.

Chromatin immunoprecipitation–exonuclease

For identification of ENT-related gene expression changes for the known targets of PAX3:FOXO1, ChIP-exo studies were performed by Peconic Genomics using anti-PAX3 antibody (R&D Systems, catalog no. MAB2457) with the murine U23674 aRMS culture as previously described (*26*). Paired-end, 40–base pair (bp) reads were generated. Putative binding sites (peaks) were identified using the Model-based Analysis of ChIP-seq (MACS) program version 2.1.0.20140616, with default parameters without a control sample (in which case MACS uses regional read levels for background) (*27*). Overall data quality and peak quality were verified by a combination of methods, including cross-correlation of positive- and negative-strand reads (which showed a peak expected from punctate binding at a plausible offset of about 15 bp) (*26*) and the R package ChIPQC (which showed good numbers of reads in peaks) (*28*). ChIP-seq data for additional ENT targets HDAC1, HDAC2, HDAC3, and HDAC11 were taken from published sources (*29*, *30*). HDAC11 was used as a surrogate for HDAC10, because HDAC10 data were not available. Additional data were gathered from published sources for PAX3:FOXO1 binding identified by cyclic amplification and selection of targets (CASTing) (*31*) and by previously published *PAX3: FOXO1* ChIP-seq data (*32*). Published expression data (*33*) and cell cycle–dependent expression data (*10*) were also integrated to infer indirect targets of *PAX3:FOXO1*. Differential expression across samples after ENT treatment was used to identify key up-regulated and down-regulated genes in aRMS. Bioinformatic analysis of PAX3:FOXO1 and HDACs binding sites was then cross-referenced for RNA-seq data for ENT-treated cells.

Statistical analysis

Bioinformatics and computational methods are described above. Continuously distributed outcomes were summarized with the mean ± 1 SD. Treatment groups were contrasted on the mean with analysis of variance in log units. Time-to-event distributions were summarized with Kaplan-Meier curves, and significance of variation with treatment group was assessed with log-rank tests. Corrections for multiple comparisons were made with the Dunnett method for analysis of variance (ANOVA) and the Bonferroni method for log-rank testing. Statistical testing on means and time to event was two sided with a nominal significance level of 5% and was carried out with R. For PDX mouse models, the significance of variation in tumor volume with treatment was assessed with a repeated-measures linear model with an autoregressive order 1 autocorrelation matrix and a Tukey correction for multiple comparisons in terms of treatment, day, and the treatment × day interaction. All analyses were carried out in $\log_{10}$ units, and all statistical testing was two sided with a 5% experiment-wise significance level. SAS version 9.4 for Windows (SAS Institute) was used throughout. GraphPad Prism was used for statistical analysis of cell viability assay and TR-FRET assay. For densitometric analysis (% change), significance was determined by a two-tailed Student's *t* test, and *P* values of <0.05 were considered to be statistically significant. Statistical significance was set at *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$. Error bars indicate mean ± SD or SEM.

SUPPLEMENTARY MATERIALS

Fig. S1. ENT treatment of aRMS in vivo.
Fig. S2. ENT reduces PAX3:FOXO1 abundance at a clinically relevant dose and time of exposure and reduces cell viability synergistically with VCR.
Fig. S3. *PAX3:FOXO1* mRNA expression is reduced by ENT in aRMS cell lines.
Fig. S4. Better than other HDACis, ENT reduces PAX3:FOXO1 abundance in murine and human aRMS cells.
Fig. S5. Representative histology of PDX mouse aRMS tissue.
Fig. S6. PFI-3–mediated inhibition of SMARCA4 bromodomain activity.
Fig. S7. Validation of commercially available antibodies for detecting endogenous FOXO1 abundance in aRMS cell lines.
Table S1. Histological markers of differentiation in aRMS orthotopic allograft PDX mice.
Table S2. Treatment schedules for the CF-4/PCB-513 PDX mice.
Table S3. Patient history of the Champions Oncology PDX aRMS models.
Table S4. Patient history of The Jackson Laboratory PDX aRMS models.
Table S5. Histological scoring of markers of differentiation in PDX aRMS mice.
Table S6. Statistical analysis of CTG-1604/POS-14175 data.
Table S7. Statistical analysis of J101220/CF-4 data.
Table S8. Statistical analysis of J77636/PCB-481 data.
Table S9. Statistical analysis of J0103366/CF-13A data.
Table S10. Statistical analysis of J099761/CF-1 data.
Table S11. Statistical analysis of CTG-1409/POS 14107 data.
Table S12. Statistical analysis of J099873/CF-2 data.
Table S13. Statistical analysis of CTG-1008 data.

REFERENCES AND NOTES

1. L. E. Davis, C. Keller, "Integrative biology of rhabdomyosarcoma using genetic murine models," *AACR 2012 Annual Meeting Educational Book* (2012).
2. M. A. Smith, S. F. Altekruse, P. C. Adamson, G. H. Reaman, N. L. Seibel, Declining childhood and adolescent cancer mortality. *Cancer* **120**, 2497–2506 (2014).
3. National Institutes of Health, National Cancer Institute, U.S. Department of Health and Human Services, National Cancer Institute, LiveStrong Young Adult Alliance, "Closing the gap: Research and care imperatives for adolescents and young adults with cancer report of the Adolescent and Young Adult Oncology Progress Review Group" (NIH Publication No. 06-6067, National Cancer Institute, 2006).
4. J. C. Breneman, E. Lyden, A. S. Pappo, M. P. Link, J. R. Anderson, D. M. Parham, S. J. Qualman, M. D. Wharam, S. S. Donaldson, H. M. Maurer, W. H. Meyer, K. S. Baker, C. N. Paidas, W. M. Crist, Prognostic factors and clinical outcomes in children and adolescents with metastatic rhabdomyosarcoma—A report from the Intergroup Rhabdomyosarcoma Study IV. *J. Clin. Oncol.* **21**, 78–84 (2003).
5. S. Malempati, D. S. Hawkins, Rhabdomyosarcoma: Review of the Children's Oncology Group (COG) soft-tissue Sarcoma committee experience and rationale for current COG studies. *Pediatr. Blood Cancer* **59**, 5–10 (2012).
6. B. A. Williams, K. M. Williams, J. Doyle, D. Stephens, M. Greenberg, D. Malkin, A. S. Pappo, Metastatic rhabdomyosarcoma: A retrospective review of patients treated at the hospital for sick children between 1989 and 1999. *J. Pediatr. Hematol. Oncol.* **26**, 243–247 (2004).
7. A. S. Pappo, J. R. Anderson, W. M. Crist, M. D. Wharam, P. P. Breitfeld, D. Hawkins, R. B. Raney, R. B. Womer, D. M. Parham, S. J. Qualman, H. E. Grier, Survival after relapse in children and adolescents with rhabdomyosarcoma: A report from the Intergroup Rhabdomyosarcoma Study Group. *J. Clin. Oncol.* **17**, 3487–3493 (1999).
8. E. R. Rudzinski, J. R. Anderson, Y.-Y. Chi, J. M. Gastier-Foster, C. Astbury, F. G. Barr, S. X. Skapek, D. S. Hawkins, B. J. Weigel, A. Pappo, W. H. Meyer, M. A. Arnold, L. A. Teot,

D. M. Parham, Histology, fusion status, and outcome in metastatic rhabdomyosarcoma: A report from the Children's Oncology Group. *Pediatr. Blood Cancer* **64**, e26645 (2017).

9. E. Missiaglia, D. Williamson, J. Chisholm, P. Wirapati, G. Pierron, F. Petel, J.-P. Concordet, K. Thway, O. Oberlin, K. Pritchard-Jones, O. Delattre, M. Delorenzi, J. Shipley, *PAX3/FOXO1* fusion gene status is the key prognostic molecular marker in rhabdomyosarcoma and significantly improves current risk stratification. *J. Clin. Oncol.* **30**, 1670–1677 (2012).

10. K. Kikuchi, S. Hettmer, M. I. Aslam, J. E. Michalek, W. Laub, B. A. Wilky, D. M. Loeb, B. P. Rubin, A. J. Wagers, C. Keller, Cell-cycle dependent expression of a translocation-mediated fusion oncogene mediates checkpoint adaptation in rhabdomyosarcoma. *PLOS Genet.* **10**, e1004107 (2014).

11. E. Y. Chen, M. T. DeRan, M. S. Ignatius, K. B. Grandinetti, R. Clagg, K. M. McCarthy, R. M. Lobbardi, J. Brockmann, C. Keller, X. Wu, D. M. Langenau, Glycogen synthase kinase 3 inhibitors induce the canonical WNT/β-catenin pathway to suppress growth and self-renewal in embryonal rhabdomyosarcoma. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5349–5354 (2014).

12. K. Kikuchi, K. Tsuchiya, O. Otabe, T. Gotoh, S. Tamura, Y. Katsumi, S. Yagyu, S. Tsubai-Shimizu, M. Miyachi, T. Iehara, H. Hosoi, Effects of PAX3-FKHR on malignant phenotypes in alveolar rhabdomyosarcoma. *Biochem. Biophys. Res. Commun.* **365**, 568–574 (2008).

13. J. Abraham, Y. Nuñez-Álvarez, S. Hettmer, E. Carrió, H.-I. H. Chen, K. Nishijo, E. T. Huang, S. I. Prajapati, R. L. Walker, S. Davis, J. Rebeles, H. Wiebush, A. T. McCleish, S. T. Hampton, C. R. R. Bjornson, A. S. Brack, A. J. Wagers, T. A. Rando, M. R. Capecchi, F. C. Marini, B. R. Ehler, L. A. Zarzabal, M. W. Goros, J. E. Michalek, P. S. Meltzer, D. M. Langenau, R. D. LeGallo, A. Mansoor, Y. Chen, M. Suelves, B. P. Rubin, C. Keller, Lineage of origin in rhabdomyosarcoma informs pharmacological response. *Genes Dev.* **28**, 1578–1591 (2014).

14. D. A. Yardley, S. F. Jones, P. Ordentlich, C. J. Braun, H. S. Pentikis, J. L. Hasapidis, M. Rees, H. A. Burris, A phase 1 study to assess the food effect on the pharmacokinetics (PK) of entinostat. *J. Clin. Oncol.* **32**, 591–591 (2014).

15. S. Bauer, R. A. Hilger, T. Mühlenberg, F. Grabellus, J. Nagarajah, M. Hoiczyk, A. Reichardt, M. Ahrens, P. Reichardt, S. Grunewald, M. E. Scheulen, A. Pustowka, E. Bock, M. Schuler, D. Pink, Phase I study of panobinostat and imatinib in patients with treatment-refractory metastatic gastrointestinal stromal tumors. *Br. J. Cancer* **110**, 1155–1162 (2014).

16. E. H. Rubin, N. G. B. Agrawal, E. J. Friedman, P. Scott, K. E. Mazina, L. Sun, L. Du, J. L. Ricker, S. R. Frankel, K. M. Gottesdiener, J. A. Wagner, M. Iwamoto, A study to determine the effects of food and multiple dosing on the pharmacokinetics of vorinostat given orally to patients with advanced cancer. *Clin. Cancer Res.* **12**, 7039–7045 (2006).

17. X.-F. Lu, X.-Y. Cao, Y.-J. Zhu, Z.-R. Wu, X. Zhuang, M.-Y. Shao, Q. Xu, Y.-J. Zhou, H.-J. Ji, Q.-R. Lu, Y.-J. Shi, Y. Zeng, H. Bu, Histone deacetylase 3 promotes liver regeneration and liver cancer cells proliferation through signal transducer and activator of transcription 3 signaling pathway. *Cell Death Dis.* **9**, 398 (2018).

18. C. G. Crist, D. Montarras, G. Pallafacchina, D. Rocancourt, A. Cumano, S. J. Conway, M. Buckingham, Muscle stem cell behavior is modified by microRNA-27 regulation of Pax3 expression. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13383–13387 (2009).

19. S. Wang, J. Huang, H. Lyu, C.-K. Lee, J. Tan, J. Wang, B. Liu, Functional cooperation of miR-125a, miR-125b, and miR-205 in entinostat-induced downregulation of erbB2/erbB3 and apoptosis in breast cancer cells. *Cell Death Dis.* **4**, e556 (2013).

20. D. Sampath, C. Liu, K. Vasan, M. Sulda, V. K. Puduvalli, W. G. Wierda, M. J. Keating, Histone deacetylases mediate the silencing of miR-15a, miR-16, and miR-29b in chronic lymphocytic leukemia. *Blood* **119**, 1162–1172 (2012).

21. F. Hernandez-Torres, A. E. Aranega, D. Franco, Identification of regulatory elements directing miR-23a–miR-27a–miR-24-2 transcriptional regulation in response to muscle hypertrophic stimuli. *Biochim. Biophys. Acta* **1839**, 885–897 (2014).

22. X. Zhang, B. Li, W. Li, L. Ma, D. Zheng, L. Li, W. Yang, M. Chu, W. Chen, R. B. Mailman, J. Zhu, G. Fan, T. K. Archer, Y. Wang, Transcriptional repression by the BRG1-SWI/SNF complex affects the pluripotency of human embryonic stem cells. *Stem Cell Reports* **3**, 460–474 (2014).

23. L. Ooi, N. D. Belyaev, K. Miyake, I. C. Wood, N. J. Buckley, BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *J. Biol. Chem.* **281**, 38974–38980 (2006).

24. S. Malempati, B. H. Chang, J. M. Reid, X. Liu, C. G. Minard, C. Keller, E. Fox, B. Weigel, ADVL1513: Results of a phase 1 trial of entinostat, an oral histone deacetylase inhibitor, in pediatric patients with recurrent or refractory solid tumors. *J. Clin. Oncol.* **36**, 10556–10556 (2018).

25. M. I. Aslam, J. Abraham, A. Mansoor, B. J. Druker, J. W. Tyner, C. Keller, PDGFRβ reverses EphB4 signaling in alveolar rhabdomyosarcoma. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6383–6388 (2014).

26. P. Ramachandran, G. A. Palidwor, C. J. Porter, T. J. Perkins, MaSC: Mappability-sensitive cross-correlation for estimating mean fragment length of single-end short-read sequencing data. *Bioinformatics* **29**, 444–450 (2013).

27. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based Analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).

28. T. S. Carroll, Z. Liang, R. Salama, R. Stark, I. de Santiago, Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front. Genet.* **5**, 75 (2014).

29. N. Nègre, C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison, J. G. Henikoff, X. Feng, K. Ahmad, S. Russell, R. A. H. White, L. Stein, S. Henikoff, M. Kellis, K. P. White, A comprehensive map of insulator elements for the *Drosophila* genome. *PLOS Genet.* **6**, e1000814 (2010).

30. Z. Wang, C. Zang, K. Cui, D. E. Schones, A. Barski, W. Peng, K. Zhao, Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019–1031 (2009).

31. T. D. Barber, M. C. Barber, O. Tomescu, F. G. Barr, S. Ruben, T. B. Friedman, Identification of target genes regulated by PAX3 and PAX3-FKHR in embryogenesis and alveolar rhabdomyosarcoma. *Genomics* **79**, 278–284 (2002).

32. L. Cao, Y. Yu, S. Bilke, R. L. Walker, L. H. Mayeenuddin, D. O. Azorsa, F. Yang, M. Pineda, L. J. Helman, P. S. Meltzer, Genome-wide identification of PAX3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer. *Cancer Res.* **70**, 6497–6508 (2010).

33. K. Nishijo, Q.-R. Chen, L. Zhang, A. T. McCleish, A. Rodriguez, M. J. Cho, S. I. Prajapati, J. A. L. Gelfond, G. B. Chisholm, J. E. Michalek, B. J. Aronow, F. G. Barr, R. L. Randall, M. Ladanyi, S. J. Qualman, B. P. Rubin, R. D. LeGallo, C. Wang, J. Khan, C. Keller, Credentialing a preclinical mouse model of alveolar rhabdomyosarcoma. *Cancer Res.* **69**, 2902–2911 (2009).

CANCER

Noncoding regions are the main source of targetable tumor-specific antigens

Céline M. Laumont[1,2]*, Krystel Vincent[1,2]*, Leslie Hesnard[1,2], Éric Audemard[1], Éric Bonneil[1], Jean-Philippe Laverdure[1], Patrick Gendron[1], Mathieu Courcelles[1], Marie-Pierre Hardy[1], Caroline Côté[1], Chantal Durette[1], Charles St-Pierre[1,2], Mohamed Benhammadi[1,2], Joël Lanoix[1], Suzanne Vobecky[3], Elie Haddad[3], Sébastien Lemieux[1,4], Pierre Thibault[1,5]†, Claude Perreault[1,2]†‡

Tumor-specific antigens (TSAs) represent ideal targets for cancer immunotherapy, but few have been identified thus far. We therefore developed a proteogenomic approach to enable the high-throughput discovery of TSAs coded by potentially all genomic regions. In two murine cancer cell lines and seven human primary tumors, we identified a total of 40 TSAs, about 90% of which derived from allegedly noncoding regions and would have been missed by standard exome-based approaches. Moreover, most of these TSAs derived from nonmutated yet aberrantly expressed transcripts (such as endogenous retroelements) that could be shared by multiple tumor types. Last, we demonstrated that, in mice, the strength of antitumor responses after TSA vaccination was influenced by two parameters that can be estimated in humans and could serve for TSA prioritization in clinical studies: TSA expression and the frequency of TSA-responsive T cells in the preimmune repertoire. In conclusion, the strategy reported herein could considerably facilitate the identification and prioritization of actionable human TSAs.

INTRODUCTION

CD8+ T cells are the main mediators of naturally occurring and therapeutically induced immune responses to cancer. Accordingly, the abundance of CD8+ tumor-infiltrating lymphocytes (TILs) positively correlates with response to immune checkpoint inhibitors and favorable prognosis (1–3). Because CD8+ T cells recognize major histocompatibility complex class I (MHC I)–associated peptides, the most important unanswered question is the nature of the specific peptides recognized by CD8+ TILs (4). Knowing that the abundance of CD8+ TILs correlates with the mutation load of tumors, the dominant paradigm holds that CD8+ TILs recognize mutated tumor-specific antigens (mTSAs), commonly referred to as neoantigens (2, 5, 6). The superior immunogenicity of mTSAs is ascribed to their selective expression on tumors, which minimizes the risk of immune tolerance (7). Nonetheless, some TILs have been shown to recognize cancer-restricted nonmutated MHC peptides (8) that we will refer to as aberrantly expressed TSAs (aeTSAs). aeTSAs can derive from a variety of cis- or trans-acting genetic and epigenetic changes that lead to the transcription and translation of genomic sequences normally not expressed in cells, such as endogenous retroelements (EREs) (9–11).

Considerable efforts are being devoted to discovering actionable TSAs that can be used in therapeutic cancer vaccines. The most common strategy hinges on reverse immunology, in which exome sequencing is performed on tumor cells to identify mutations, and MHC-binding prediction software tools are used to identify which mutated peptides might be good MHC binders (12, 13). Although reverse immunology can enrich for TSA candidates, 90% of these candidates are false positives (6, 14) because available computational

methods may predict MHC binding, but they cannot predict other steps involved in MHC peptide processing (15, 16). To overcome this limitation, a few studies have included mass spectrometry (MS) analyses in their TSA discovery pipeline (17), thereby providing a rigorous molecular definition of several TSAs (18, 19). However, the yield of these approaches has been meager: In melanoma, one of the most mutated tumor types, an average of two TSAs per individual tumors has been validated by MS (20), whereas only a handful of TSAs has been found for other cancer types (15). The paucity of TSAs is puzzling because injection of TILs or immune checkpoint inhibitors would not cause tumor regression if tumors did not express immunogenic antigens (21). We surmised that approaches based on exonic mutations have failed to identify TSAs because they did not take into account two crucial elements. First, these approaches focus only on mTSAs and neglect aeTSAs, essentially because there is currently no method for high-throughput identification of aeTSAs. This represents a major shortcoming because, whereas mTSAs are private antigens (that is, unique to a given tumor), aeTSAs would be preferred targets for vaccine development because they can be shared by multiple tumors (8, 10). Second, focusing on the exome as the only source of TSAs is very restrictive. Of particular relevance to TSA discovery, 99% of cancer mutations are located in noncoding regions (22). Moreover, the exome (all protein-coding sequences) is only 2% of the human genome, whereas up to 75% of the genome can be transcribed and potentially translated (23). Hence, many allegedly noncoding regions are protein coding, and translation of noncoding regions has been shown to generate numerous MHC peptides (24, 25), some of which were retrospectively identified as targets of TILs and autoreactive T cells (26, 27).

With these considerations in mind, we developed a proteogenomic strategy designed to discover mTSAs and aeTSAs coded by all genomic regions. We used this approach to study two well-characterized murine cancer cell lines, CT26 and EL4, as well as seven primary human samples comprising four B-lineage acute lymphoblastic leukemias (B-ALLs) and three lung cancers. Our main objectives were to determine whether noncoding regions contribute to the TSA landscape and which parameters may influence TSA immunogenicity.

[1]Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montreal, Quebec H3C 3J7, Canada. [2]Department of Medicine, Université de Montréal, Montreal, Quebec H3C 3J7, Canada. [3]CHU Sainte-Justine, Université de Montréal, Montreal, Quebec H3T 1C5, Canada. [4]Department of Computer Science and Operations Research, Université de Montréal, Montreal, Quebec H3C 3J7, Canada. [5]Department of Chemistry, Université de Montréal, Montreal, Quebec H3C 3J7, Canada.
*These authors contributed equally to this work.
†These authors jointly supervised this work.
‡Corresponding author. Email: claude.perreault@umontreal.ca

RESULTS

Rationale and design of a proteogenomic method for TSA discovery

Attempts to computationally predict TSAs using various algorithms are fraught with exceedingly high false discovery rates (28). Hence, a system-level molecular definition of the MHC peptide repertoire may only be achievable by high-throughput MS studies (4). Current approaches use tandem MS (MS/MS) software tools, such as Peaks (29), which rely on a user-defined protein database to match each acquired MS/MS spectrum to a peptide sequence. Because the reference proteome does not contain TSAs, MS-based TSA discovery workflows must use proteogenomic strategies to build customized databases derived from tumor RNA sequencing (RNA-seq) data (30) that should ideally contain all proteins, even unannotated ones, expressed in the considered tumor sample. Because current MS/MS software tools cannot deal with the large search space created by translating all RNA-seq reads in all reading frames (31, 32), we devised a proteogenomic strategy enriching for cancer-specific sequences to comprehensively characterize the landscape of TSAs coded by all genomic regions. The resulting database, termed a global cancer database, is composed of two customizable parts. The first part, the canonical cancer proteome (Fig. 1A), was obtained by in silico translation of expressed protein-coding transcripts in their canonical frame; it therefore contains proteins coded by exonic sequences that are normal or contain single-base mutations. The second part, the cancer-specific proteome (Fig. 1B), was generated using an alignment-free RNA-seq workflow called k-mer profiling because current mappers and variant callers poorly identify structural variants. This second dataset enabled the detection of peptides encoded by any reading frame of any genomic origin (including structural variants), as long as they were cancer specific (that is, absent from normal cells). Here, we elected to use MHC II^{hi} medullary thymic epithelial cells (mTEC^{hi}) cells as a "normal control" because they express most known genes and orchestrate T cell selec-

tion to induce central tolerance to MHC peptides coded by their vast transcriptome (fig. S1A) (33). Thus, to identify RNA sequences that were cancer specific, we chopped cancer RNA-seq reads into 33-nucleotide-long sequences, called k-mers (34), from which we removed k-mers present in syngeneic mTEC^{hi} cells (fig. S2, A and B). Redundancy inherent to the k-mer space was removed by assembling overlapping cancer-specific k-mers into longer sequences, called contigs, which were 3-frame translated in silico (Fig. 1B and fig. S2, C and D). We then concatenated the canonical and cancer-specific proteomes to create a global cancer database, one for each analyzed sample (table S1A). Using these optimized databases, we identified MHC peptides eluted from two well-characterized mouse tumor cell lines that we sequenced by MS, namely CT26, a colorectal carcinoma from a Balb/c mouse, and EL4, a T-lymphoblastic lymphoma from a C57BL/6 mouse (Fig. 1C and table S2A) (35, 36).

Noncoding regions as a major source of TSAs

We identified 1875 MHC peptides on CT26 cells and 783 on EL4 cells (tables S3 and S4). Among these, peptides absent from the mTEC^{hi} proteome were considered TSA candidates (i) if their 33-nucleotide-long peptide-coding sequence derived from a full cancer-restricted 33-nucleotide-long k-mer and was absent from the mTEC^{hi} transcriptome or (ii) if their 24- to 30-nucleotide-long peptide-coding sequence, derived from a truncated version of a cancer-restricted 33-nucleotide-long k-mer, was overexpressed by at least 10-fold in the transcriptome of cancer cells versus mTEC^{hi} cells (fig. S3A). Because no error estimation was used in our study, we manually validated the MS spectra of our TSA candidates. Before assigning peptides a genomic location, we also removed any indistinguishable isoleucine/leucine variants (figs. S3, B and C, and S4) and ended up with a total of 6 mTSAs and 15 aeTSA candidates: 14 presented by CT26 cells and 7 by EL4 cells (Fig. 2, A and B). MHC peptides that were both mutated and aberrantly expressed were included in the mTSA category. All of these peptides
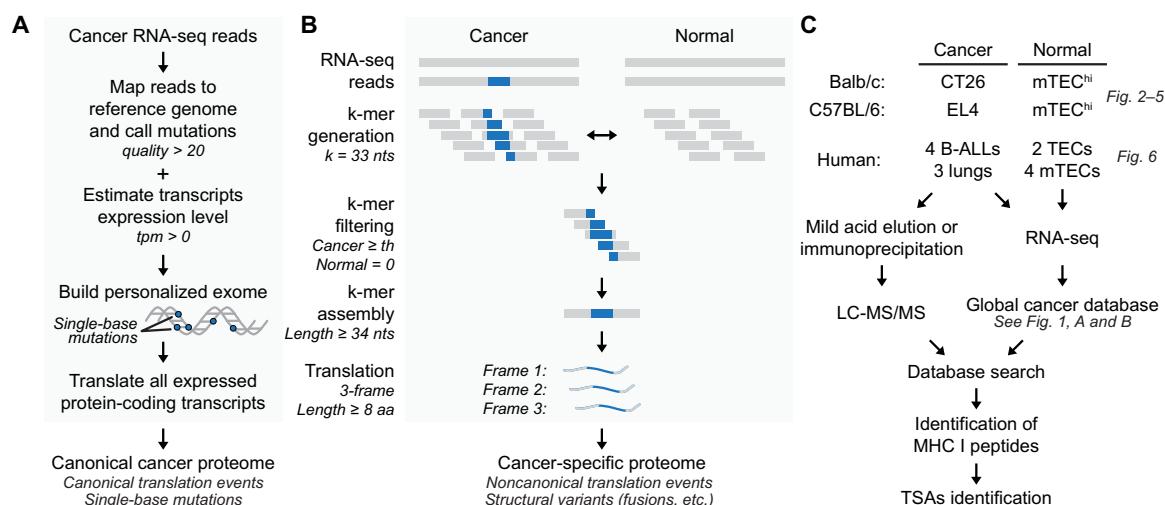

Fig. 1. Proteogenomic workflow for the identification of TSAs. (**A** and **B**) Schematic detailing how the canonical cancer proteome (A) and cancer-specific proteome (B) were built for each analyzed sample. In (A), "quality" refers to the Phred score; a score of >20 means that the accuracy of the nucleobase call is at least 99%. (**C**) The combination of the above two proteomes, termed the global cancer database, was then used to identify MHC peptides, and more specifically TSAs, sequenced by liquid chromatography–MS/MS (LC-MS/MS). We analyzed two well-characterized murine cell lines, CT26 and EL4, and seven human primary samples, namely, four B-ALLs and three lung tumor biopsies (n = 2 to 4 per sample). Statistics regarding each part of the global cancer database can be found in table S1, and implementation details of building the cancer-specific proteome by k-mer profiling are presented in fig. S2. aa, amino acids; nts, nucleotides; th, sample-specific threshold for k-mer occurrence; tpm, transcripts per million.

were absent from the Immune Epitope Database (37), except for one: the AH1 peptide (SPSYVYHQF), the sole aeTSA previously identified on CT26 cells using reverse immunology (10, 38).

To assess the stringency of our database-building strategy based on the removal of mTEC[hi] k-mers from cancer k-mers, we evaluated the peripheral expression of RNAs coding for aeTSAs across a panel of 22 tissues (table S5) (39, 40). Four of the 15 aeTSA candidates had an expression profile similar to that of previously reported "overexpressed" tumor-associated antigens (41, 42), as their peptide-coding sequences were expressed in most or all tissues (Fig. 2C). These four peptides were therefore excluded from the TSA list. In contrast, 11 peptides were considered genuine aeTSAs because their source transcripts were either totally absent or present at trace amounts in a few tissues (Fig. 2C). We note that detection of low transcript amounts is less relevant because MHC peptides preferentially derive from highly abundant transcripts (43, 44). This concept is illustrated by the AH1 TSA, which elicits strong antitumor responses devoid of adverse effects (10, 38), despite the weak expression of its peptide-coding sequence in the liver, thymus, and urinary bladder (Fig. 2C). These results demonstrate that subtracting mRNA sequences found in mTEC[hi] strongly enriches for cancer-restricted peptide-coding sequences. When we consider our entire murine TSA dataset (6 mTSAs and 11 aeTSAs), we find that most of them derive from atypical translation events: the out-of-frame translation of a coding exon or the translation of noncoding regions (Fig. 2D). We also noticed that any type of noncoding region can generate TSAs (table S6): intergenic and intronic sequences, noncoding exons, untranslated region (UTR)/exon junctions, and EREs, which appear to be a particularly rich source of TSAs (eight aeTSAs and one mTSA). Last, our approach efficiently captured at least one structural variant as we identified an antigen, VTPVYQHL, derived from a very large intergenic deletion (~7500 bp) in EL4 cells (table S6B). Together, these observations confirm that noncoding regions are the main source of TSAs and that they have the potential to considerably expand the TSA landscape of tumors.
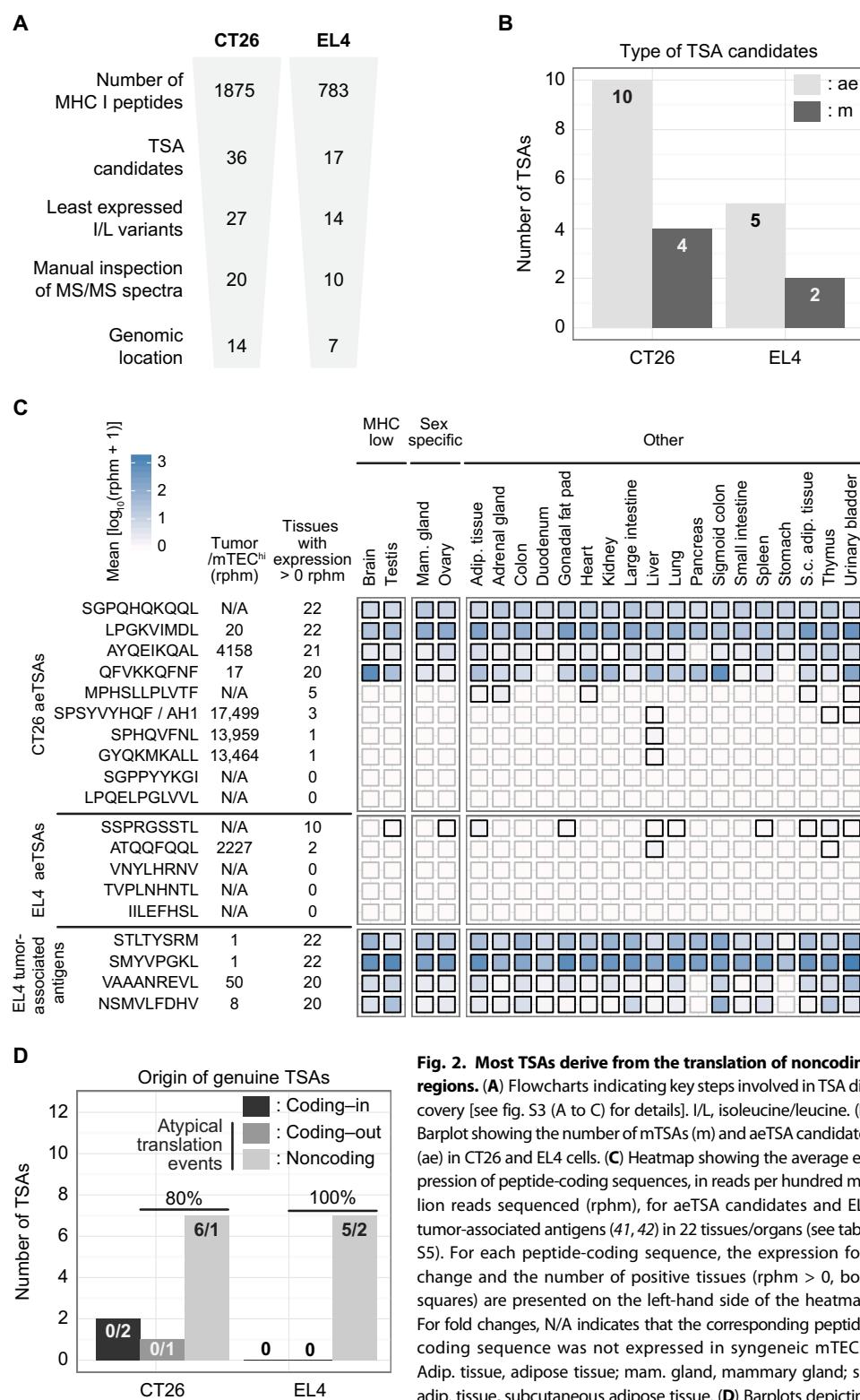
Fig. 2. Most TSAs derive from the translation of noncoding regions. (A) Flowcharts indicating key steps involved in TSA discovery [see fig. S3 (A to C) for details]. I/L, isoleucine/leucine. (B) Barplot showing the number of mTSAs (m) and aeTSA candidates (ae) in CT26 and EL4 cells. (C) Heatmap showing the average expression of peptide-coding sequences, in reads per hundred million reads sequenced (rphm), for aeTSA candidates and EL4 tumor-associated antigens (41, 42) in 22 tissues/organs (see table S5). For each peptide-coding sequence, the expression fold change and the number of positive tissues (rphm > 0, bold squares) are presented on the left-hand side of the heatmap. For fold changes, N/A indicates that the corresponding peptide-coding sequence was not expressed in syngeneic mTEC[hi]. Adip. tissue, adipose tissue; mam. gland, mammary gland; s.c. adip. tissue, subcutaneous adipose tissue. (D) Barplots depicting the number of TSAs derived from the translation of noncoding regions (noncoding) and of coding exons in-frame (coding–in) or out-of-frame (coding–out). The number of aeTSAs/mTSAs is reported within bars. The proportion of TSAs derived from atypical translation events is shown above bars. Features of CT26 and EL4 TSAs can be found in table S6 (A and B, respectively).

Differential protection against EL4 cells after immunization against individual TSAs

We then performed detailed studies on some of the TSAs that seemed most therapeutically promising: those presented by EL4 cells and whose peptide-coding sequence was not expressed by any normal tissue (Fig. 2C and tables S6B and S7). To assess immunogenicity, C57BL/6 mice were immunized twice with either unpulsed (control group) or TSA-pulsed dendritic cells (DCs) before being challenged with live EL4 cells. Priming against IILEFHSL or TVPLNHNTL prolonged survival for 10% of mice, with only one TVPLNHNTL-immunized mouse surviving up to day 150 (Fig. 3A). In contrast, the other three TSAs showed superior efficacy, with day 150 survival rates of 20% (VNYIHRNV), 30% (VTPVYQHL), and 100% (VNYLHRNV) (Fig. 3, B and C). To evaluate the long-term efficacy of TSA vaccination, surviving mice were rechallenged with live EL4 cells at day 150 and monitored for signs of disease. The two VNYIHRNV-immunized survivors died of leukemia within 50 days, whereas all others (immunized against TVPLNHNTL, VTPVYQHL, or VNYLHRNV) survived the rechallenge (Fig. 3). We conclude that immunization against individual TSAs confers different degrees of protection against EL4 cells (0 to 100%) and that, in most cases, this protection is long-lasting.

Frequency of TSA-responsive T cells in naïve and immunized mice

In various models, the strength of in vivo immune responses is regulated by the number of antigen-reactive T cells (45, 46). We therefore assessed the frequency of TSA-responsive T cells in naïve and immunized mice using a tetramer-based enrichment protocol (47, 48), for which the gating strategy and one representative experiment can be found in fig. S5 (A to C). As positive controls, we used tetramers to detect $CD8^+$ T cells specific for three immunodominant viral epitopes (gp-33, M45, and B8R). We confirmed that these T cells had a high abundance and that their frequency was similar to that observed in previous studies (Fig. 4A) (46). In naïve mice, $CD8^+$ T cells specific for TVPLNHNTL, VTPVYQHL, and IILEFHSL were rare (less than one tetramer$^+$ cell per 10^6 $CD8^+$ T cells), whereas $CD8^+$ T cells specific for the ERE TSAs (VNYIHRNV and VNYLHRNV) displayed frequencies similar to those of our viral controls (Fig. 4A and fig. S6A). Accordingly, in mice immunized with TSA-pulsed DCs, we found that the T cell frequencies against the two ERE TSAs, as assessed by tetramer staining or interferon-γ (IFN-γ) enzyme-linked immunospot (ELISpot) assays (figs. S1B, S5, C and D, and S6A), were higher than that of TVPLNHNTL, VTPVYQHL, and IILEFHSL (Fig. 4, B and C). Moreover, in both naïve and immunized mice, results obtained with tetramer staining and IFN-γ ELISpot correlated with each other (fig. S7). Last, we estimated that the functional avidity of T cells specific for VNYIHRNV and VNYLHRNV was similar to that of T cells specific for two highly immunogenic nonself antigens: the minor histocompatibility antigens H7[a] and H13[a] (Fig. 4D). Hence, these TSAs, derived from allegedly noncoding regions, were recognized by highly abundant T cells with a high functional avidity. This is particularly noteworthy for the VNYLHRNV aeTSA because it has an unmutated germline sequence.

Together, our results show that the frequency of TSA-responsive T cells was a crucial parameter for TSA immunogenicity. However, VTPVYQHL was an outsider: It afforded the second-best protection against EL4 challenge although its cognate T cells were present at a very low frequency (Figs. 3 and 4, A to C). To better evaluate the importance of T cell expansion in leukemia protection, we estimated the frequency of tetramer$^+$ $CD8^+$ T cells in long-term survivors after
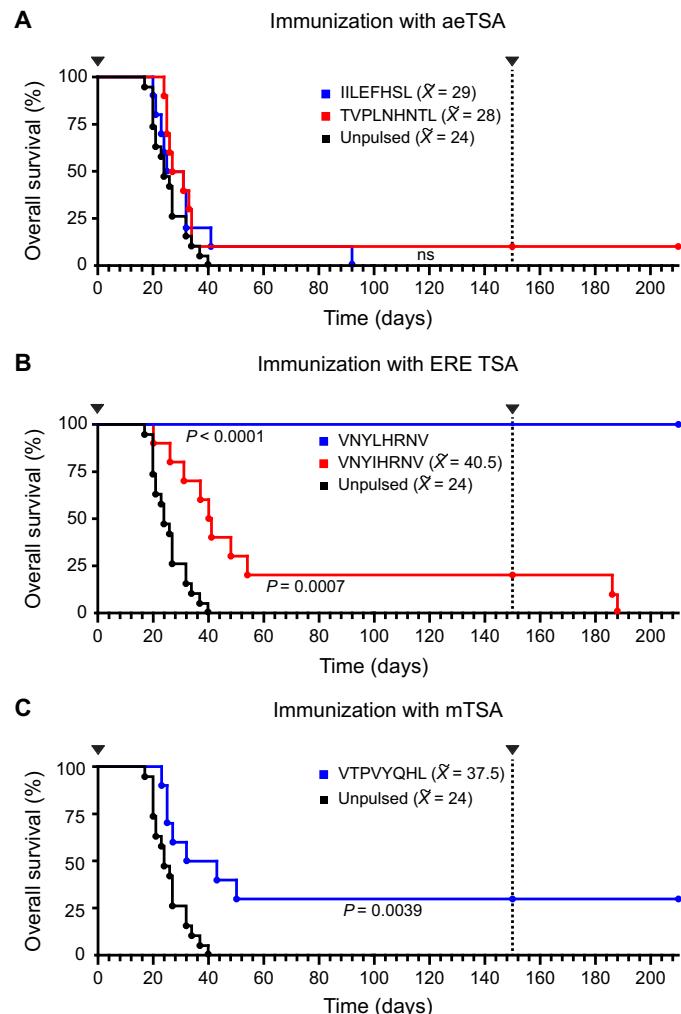
Fig. 3. Immunization against individual TSAs confers different degrees of protection against EL4 cells. C57BL/6 mice were immunized twice with DCs pulsed with individual TSAs: (**A**) two aeTSAs, (**B**) two ERE TSAs (one aeTSA or one mTSA), or (**C**) one mTSA. Mice were injected intravenously with 5×10^5 live EL4 cells (arrowheads) on day 0, and all surviving mice were rechallenged on day 150. Control groups were immunized with unpulsed DCs (solid black line). $\tilde{X}$ represents the median survival. Statistical significance of immunized group versus control group was calculated using a log-rank test, where ns stands for not significant ($P > 0.05$). $n = 10$ mice per group for peptide-specific immunization, $n = 19$ mice for control group.

rechallenge with EL4 cells on day 150 (Fig. 3). These analyses were performed on day 210 or at the time of sacrifice (in the case of VNYIHRNV-primed mice). Two points can be made from these analyses (fig. S6, B and C). First, all long-term survivors, including VTPVYQHL-immunized mice, showed a discernable population of TSA-responsive (tetramer$^+$) $CD8^+$ T cells. Second, although VNYIHRNV was recognized by a particularly large population of tetramer$^+$ cells, it was the only TSA that did not protect mice upon rechallenge. Together, our results suggest that expansion of TSA-responsive T cells was necessary for protection against EL4 cells but was insufficient in the case of VNYIHRNV.

## The importance of antigen expression for protection against EL4 cells
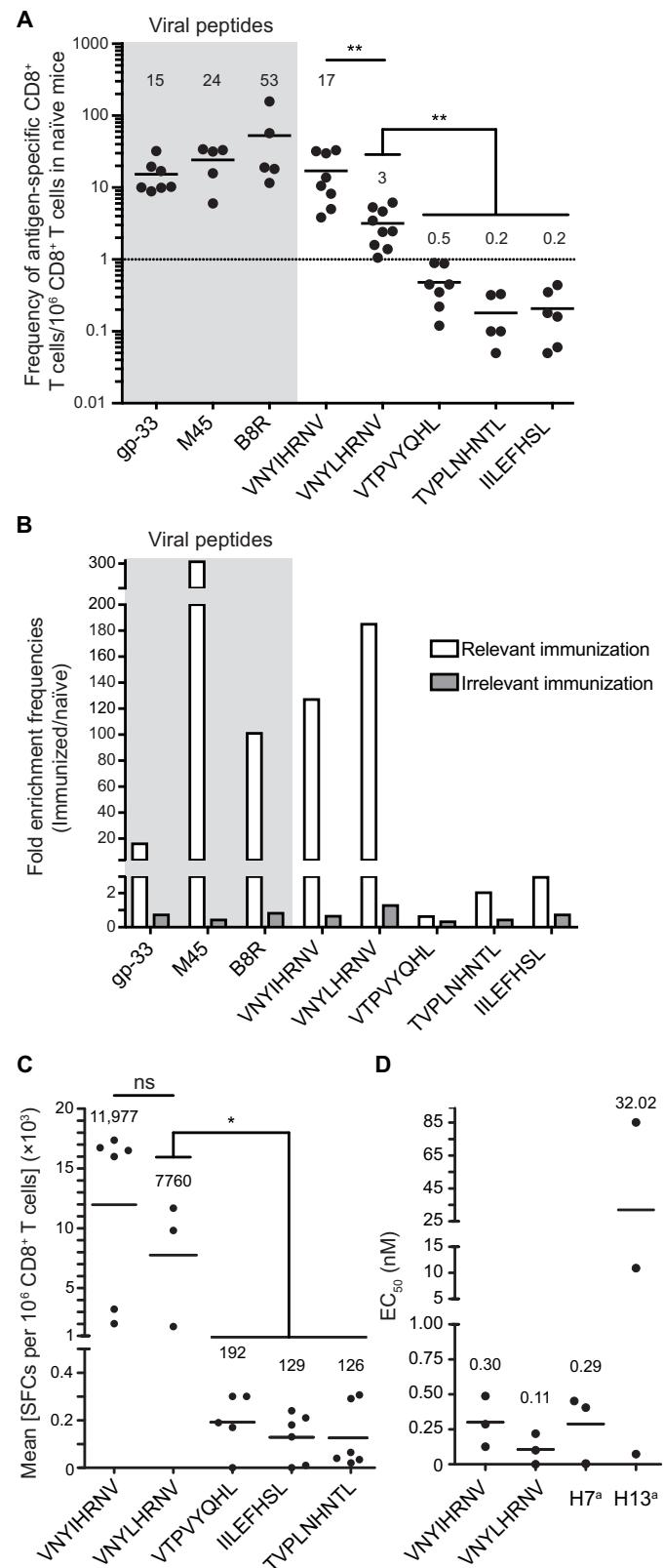
Next, we evaluated the impact of antigen expression on immunogenicity by assessing the abundance of TSAs at the RNA level in the EL4 cell

Fig. 4. Frequency of and IFN-γ secretion by TSA-responsive T cells in naïve and immunized mice. (**A**) Number of tetramer$^+$ CD8$^+$ T cells per 10^6 CD8$^+$ T cells in naïve mice. Circles, one mouse (n = 5 to 9 mice per group); dotted line, frequency of 1 tetramer$^+$ T cell per 10^6 CD8$^+$ T cells. (**B**) Fold enrichment of tetramer$^+$ CD8$^+$ T cells after immunization with relevant (white bars) or irrelevant (gray bars) peptides ($^{\text{Mean frequency}_{immunized}}$/$_{\text{Mean frequency}_{naïve}}$). (**C**) The number of spot-forming cells (SFCs), measured by an IFN-γ ELISpot assay, averaged across technical replicates (circles) after being converted to SFCs per 10^6 CD8$^+$ T cells: $\left[^{(\text{SFCs}_{immunized} - \text{SFCs}_{naïve})}/_{\text{Number of T cells plated}}\right] \times 10^6$. (**D**) The functional avidity of T cells recognizing specific TSAs and two previously reported positive controls [H7[a] and H13[a] (42)] was estimated by calculating a half maximal effective concentration (EC$_{50}$), corresponding to the peptide concentration where half of plated antigen-specific T cells secreted IFN-γ. (B to D) Three independent experiments. On relevant panels, full horizontal lines and numbers above each condition represent mean values. Viral peptides used as control are highlighted in gray. *$P \leq 0.05$ and **$P \leq 0.01$ (two-sided Wilcoxon rank sum test with the Benjamini-Hochberg correction).

population that was injected on day 0 (Fig. 3). The sequence encoding the TSA conferring the best protection against EL4 cells (VNYLHRNV) was expressed more than the other TSA-coding sequences (Fig. 5A). This suggests that VNYLHRNV is likely "clonal" (expressed by all EL4 cells) and highly expressed, whereas the other TSAs are subclonal and/or expressed at low amounts. Next, using parallel reaction monitoring (PRM) MS, we analyzed the TSA copy number per cell in the EL4 cell population used for rechallenge (day 150; Fig. 5B). As expected (41), there was no linear relationship between TSA abundance at the RNA and peptide levels (Fig. 5, A and B). The most protective TSA, VNYLHRNV, was one of the two most abundant TSAs (>500 copies per cell), whereas VNYIHRNV, which offered no protection upon rechallenge (Fig. 3B), was no longer detected on EL4 cells. This observation suggests that VNYIHRNV was a subclonal TSA and that antigen loss most likely explained the lack of protection upon rechallenge. Last, we noted that TSAs were immunogenic when presented by DCs but not when presented by EL4 cells: Injection of live EL4 cells without prior immunization did not induce an expansion of TSA-responsive T cells (Fig. 5C and fig. S6D), and immunization with irradiated EL4 cells did not confer any protection against live EL4 cells (Fig. 5D). This suggests that, in the absence of immunization, highly immunogenic TSAs (such as VNYLHRNV) were ignored likely because they were not efficiently cross-presented by DCs, highlighting the importance of efficient T cell priming in cancer immunotherapy.

Impact of noncoding regions on the TSA landscape of human primary tumors

Having established that noncoding regions are a major source of TSAs in two murine cell lines, we applied our proteogenomic approach to seven human primary tumor samples: four B-ALLs and three lung cancers (fig. S8 and tables S8 to S15). Rather than using RNA-seq data from murine syngeneic mTEC^{hi}, we sequenced the transcriptome of total TECs (n = 2) and purified mTECs (n = 4) from six unrelated donors undergoing corrective cardiovascular surgery (table S2B). We found minimal interindividual differences and demonstrated that this cohort size was sufficient to cover almost the full breadth of the mTEC transcriptomic landscape, as computing the cumulative number of detected transcripts showed that minimal gains would be achieved by adding more samples (fig. S9). Using these RNA-seq data as the repertoire of normal k-mers to generate global cancer databases (table S1B), as described in Fig. 1, we identified 2 mTSAs and 27 aeTSA candidates (Fig. 6A). After validating their assignment to a single genomic location and the quality of their MS spectra, we also

ensured that mTSAs did not intersect with known germline polymorphisms (figs. S3, 10, and 11). To further validate the status of aeTSA candidates, as we did for murine aeTSAs (Fig. 2C), we analyzed the
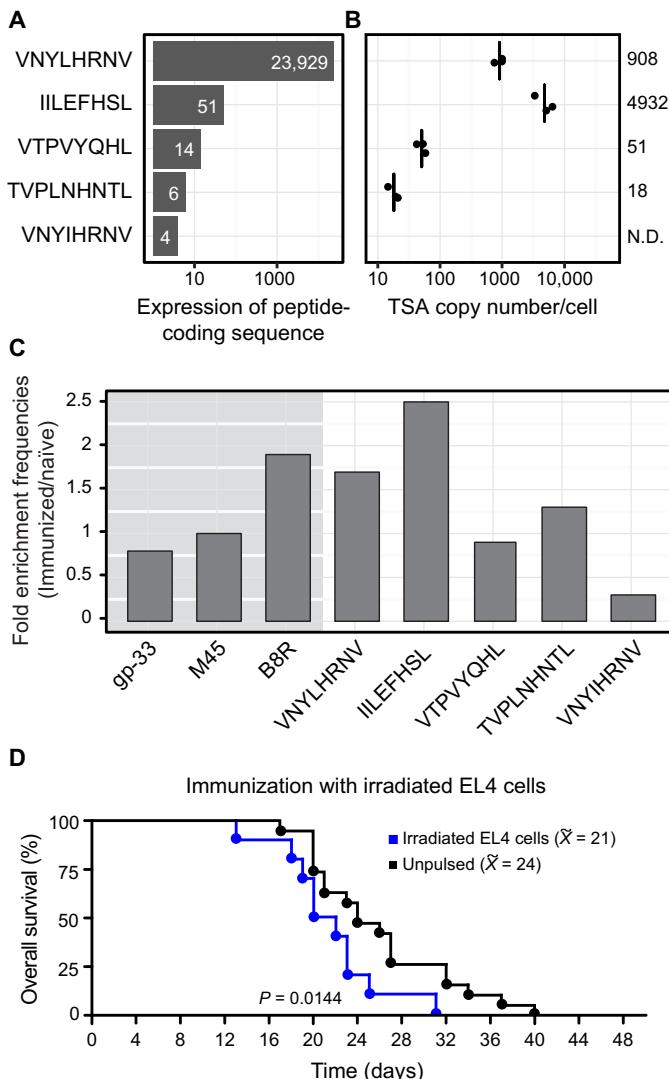
Fig. 5. High expression of EL4 TSAs is necessary but not sufficient to induce antileukemic responses. (**A** and **B**) Analysis of TSA expression at the RNA and peptide levels was performed on EL4 cells injected into mice at day 0 or day 150, respectively. (A) The number of RNA-seq reads fully overlapping the RNA sequences encoding each TSA. (B) TSA copy number per cell was estimated by PRM MS using ^{13}C-synthetic peptide analogs of the TSAs (three replicates). Black lines represent the mean TSA copy number per cell (also indicated on the left-hand side of the graph). N.D., not detected. (**C**) Fold enrichment for tetramer$^+$ CD8$^+$ T cells after injection with live EL4 cells without prior immunization (Mean frequency$_{EL4-injected}$ / Mean frequency$_{naive}$). Fold enrichment for T cells recognizing viral peptides is shown as negative controls and is highlighted in gray. Three independent experiments were performed. (**D**) Overall survival of C57BL/6 female mice immunized twice with irradiated (10,000 cGy) EL4 cells (blue line, n = 10 mice) or unpulsed DCs (black line, n = 19 mice) and then injected intravenously with 5×10^5 live EL4 cells. $\tilde{X}$ represents the median survival. Statistical significance of immunized group versus control group was calculated using a log-rank test.

expression of aeTSA-coding sequences in RNA-seq data from 28 tissues (6 to 50 individuals per tissue; Fig. 6B and table S16). On the basis of these data, we excluded six aeTSA candidates: three that were widely expressed, like most previously reported overexpressed tumor-associated antigens (49), and three that were expressed at substantial amounts in a
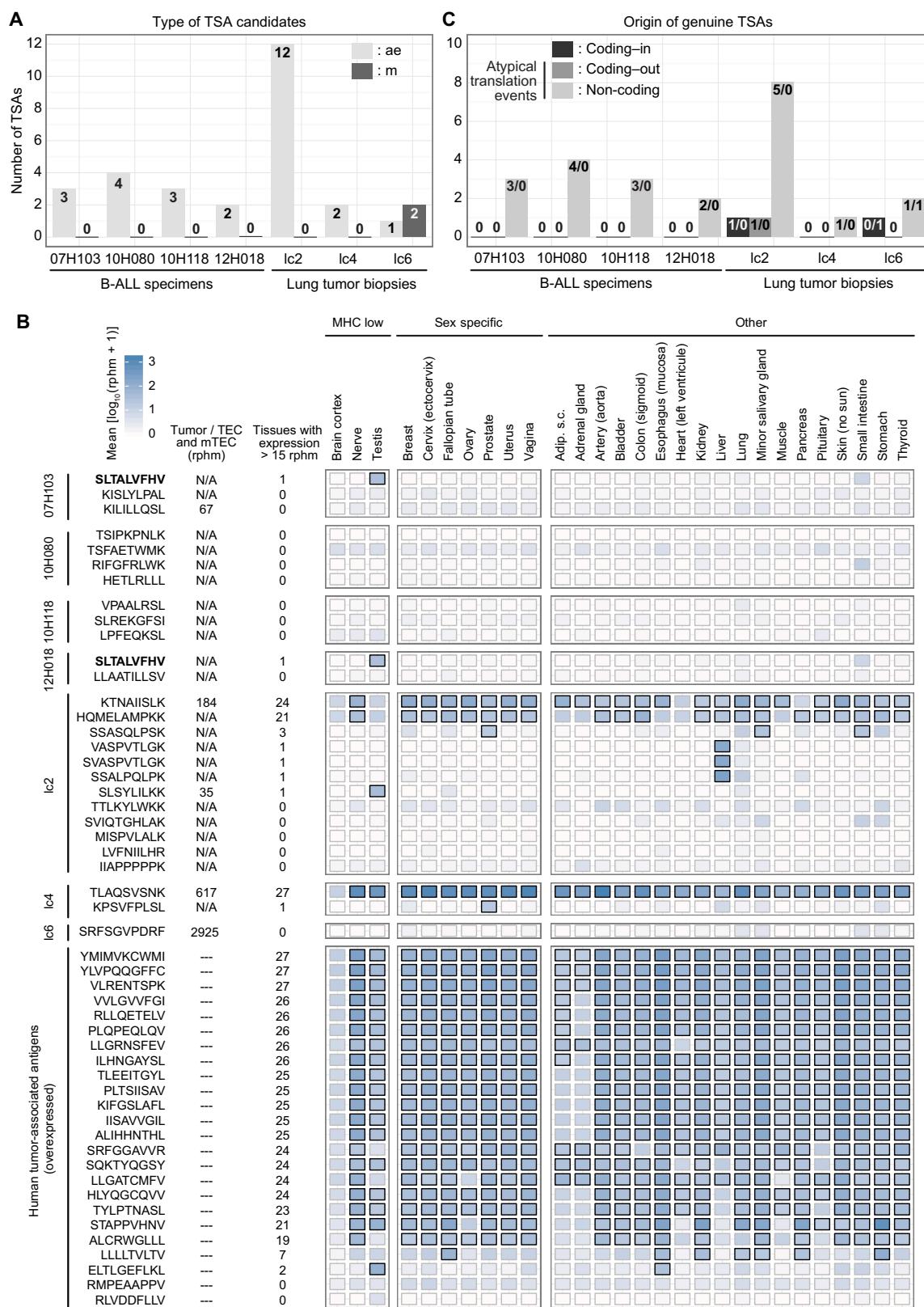
single organ, the liver (Fig. 6B). We therefore ended up with a total of two mTSAs and 20 nonredundant aeTSAs (Fig. 6C and tables S17 and S18). Of note, the SLTALVFHV aeTSA was shared by our two HLA-A*02:01–positive B-ALLs (table S17). This aeTSA derives from the 3′UTR of *TCL1A*, a gene implicated in lymphoid malignancies. Together, our results show that our proteogenomic approach can characterize the repertoire of mTSAs and aeTSAs on individual tumors in about 2 weeks.

DISCUSSION

To explore the global landscape of TSAs, we developed a proteogenomic approach that incorporates two features in the construction of databases for MS analyses: alignment-free k-mer profiling of RNA-seq data and subtraction of mTEC k-mers. In a context where TSA discovery is a critical unmet medical need, our approach led us to discover that the TSA landscape is much larger than previously anticipated. Thirty-five of the 39 nonredundant TSAs reported here derived from atypical translation events: 2 from the out-of-frame translation of coding exons and 33 from allegedly noncoding regions. Hence, ~90% of our TSAs would have been missed by standard approaches focusing on exonic mutations. In addition to MHC peptides derived from RNAs containing single-base mutations, our approach efficiently captured peptides generated by complex structural variants, as exemplified by VTPVYQHL, which derived from a large intergenic deletion (~7500 bp) in EL4 cells. Subtraction of mTEC k-mers was critical for the high-throughput identification of immunogenic aeTSAs including unmutated peptides that are not constitutively presented by mTEC^{hi} to thymocytes during the establishment of central tolerance. This is well-illustrated by VNYLHRNV, an unmutated TSA absent from mTEC^{hi} and other peripheral tissues, although strongly expressed in EL4 cells, that is, recognized by highly abundant CD8$^+$ T cells with a high functional avidity. Nonetheless, a few peptide-coding transcripts undetectable in mTEC^{hi} were detected in peripheral tissues, suggesting that k-mers from both mTECs and peripheral tissues should be used to identify genuine TSAs. In the present study, we chose to use peripheral expression as an a posteriori validation step. An alternative approach would be to remove all k-mers expressed in peripheral tissues when building the database for MS.

TSAs derived from noncoding regions present a number of peculiar and highly relevant features. First, it is evident that EREs are a rich source of TSAs; they generated 9 of the 17 TSAs found in murine cell lines and 4 of the 23 TSAs in human tumors. The difference in the proportion of ERE TSAs that we identified in mouse cell lines versus primary human tumors might be ascribed to the fact that in vitro culture conditions do not recapitulate the immune pressure exerted on developing tumors, that ERE expression greatly varies across tumor types (50), or that human EREs are more degenerated and therefore less likely to be translated than murine EREs (9). Nonetheless, ERE TSAs remain particularly relevant to the development of cancer vaccines because both oncogenic viruses and viral-like sequences in the human genome appear to be particularly immunogenic (51, 52). Second, most TSAs derived from noncoding regions do not overlap mutations and are therefore, by definition, aeTSAs. Such aeTSAs, which include EREs, present a major advantage over mTSAs (of both coding or noncoding origin): Whereas mTSAs are private antigens, aeTSAs can be shared by multiple tumors (8). We were able to identify such shared aeTSA (SLTALVFHV) in humans, whereas Probst et al. (10) showed that mice immunized against the AH1 aeTSA (SPSYVYHQF), which we identified

Fig. 6. Most TSAs detected in human primary tumors derive from the translation of noncoding regions. (**A**) Barplot showing the number of aeTSAs candidates (ae) and mTSAs (m) in each primary sample analyzed. (**B**) Heatmap showing the average expression of peptide-coding sequences, in rphm, for aeTSAs and overexpressed tumor-associated antigens obtained from the Cancer Immunity Peptide database (49) across a panel of 28 tissues (see table S16). For each peptide-coding sequence, the expression fold change (tumor/TEC and mTEC) and the number of positive tissues (rphm > 15, bold squares) are shown on the left-hand side of the heatmap. For fold changes, N/A, and—indicate that the corresponding peptide-coding sequence was not detected in TEC/mTEC samples or not computed, respectively. Adip. s.c., adipose subcutaneous. (**C**) Barplots depicting the number of TSAs derived from the translation of noncoding regions (noncoding) or from coding exons translated in-frame (coding–in) or out-of-frame (coding–out). The number of aeTSAs/mTSAs is shown within bars. Features of human TSAs identified in each sample can be found in tables S17 and S18.

by MS on CT26 cells, survived the challenge with three different cancer cell lines: the WEHI-164 fibrosarcoma and the C51 and CT26 colorectal cancers.

Together, our MS-based discovery of TSAs in primary human B-ALL and lung cancer demonstrates that the impetus to develop TSA vaccines should not be limited to cancers with a high mutational

load. Because B-ALLs harbor very few exonic mutations, it was presumed that they might not express any TSA (5). Our data show that TSAs can be found in B-ALL provided that the search strategy encompasses aeTSAs. Moreover, two elements argue that TSAs derived from atypical translation events represent promising targets for T cell–based cancer immunotherapy: (i) They outnumber TSAs derived from coding regions and (ii) they are mostly unmutated, which increases their potential to be shared between patients.

We acknowledge that our study presents several limitations for which solutions can be envisioned. First, because our approach is not compatible with the computation of classical false discovery rates, TSAs must undergo meticulous validation by manual inspection of MS spectra and, ideally, confirmation using synthetic peptide analogs. Second, our approach relies on shotgun MS, which suffers from a limited dynamic range. Consequently, we only detected the most abundant TSAs and are likely underestimating the extent of aeTSA sharing between patients. Because shared aeTSAs may represent promising actionable targets for cancer immunotherapy (8), aeTSA frequency across patients and/or tumor types could be further evaluated by targeted MS analyses that have a more limited scope but are 10 times more sensitive than shotgun MS and can provide quantitative data such as the number of TSA copies per cell (53). Third, because TSA immunogenicity cannot be predicted (54), it has to be tested experimentally for each TSA. This issue is being addressed by several research groups that are currently developing artificial platforms requiring less material than the conventional IFN-γ ELISpot assays used for such testing (8, 55, 56).

In practice, how could we prioritize TSAs for clinical trials? In our EL4 tumor model, the efficacy of TSA immunization was largely determined by two criteria: TSA abundance and the frequency of TSA-responsive T cells. TSA abundance can be assessed by targeted MS analyses, and the frequency of TSA-responsive T cells in peripheral blood mononuclear cells of a cohort of subjects could be estimated using MHC peptide multimers or functional assays. Widely shared, highly abundant TSAs recognized by high-frequency T cells could then be selected for clinical trials. These optimal aeTSAs could then potentially be combined in a single vaccine using already available synthesis and delivery platforms (13, 57).

MATERIALS AND METHODS
Study design
The purpose of this study was to develop a proteogenomic approach that would enable the identification of TSAs derived from any region of the genome and to identify features that influenced TSA immunogenicity. To do so, we first characterized the TSA landscape of two murine cell lines, the EL4 T-lymphoblastic lymphoma cell line and the CT26 colorectal cancer cell line that were both obtained from the American Type Culture Collection. As a normal control, we used thymi isolated from 5- to 8-week-old C57BL/6 or Balb/c mice obtained from the Jackson Laboratory. Mice were housed under specific pathogen-free conditions, and all experimental protocols were approved by the Comité de Déontologie de l'Expérimentation sur des Animaux of Université de Montréal. We also applied our approach to seven human primary cancer samples from treatment-naïve patients. These included three lung tumor biopsies (lc2, lc4, and lc6) purchased from Tissue Solutions and four primary leukemic samples (B-ALL specimens 07H103, 10H080, 10H118, and 12H018) that were collected and cryopreserved at the Banque de Cellules Leucémiques du Québec at Hôpital Maisonneuve-Rosemont. The project was approved by the

Comité d'éthique de la recherche de l'Hôpital Maisonneuve-Rosemont (CÉR 12100). NOD Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) mice were used to expand our B-ALL specimens. These mice were purchased and housed as described for C57BL/6 and Balb/c mice. As a normal control, we used thymi obtained from 3-month-old to 7-year-old patients undergoing corrective cardiovascular surgery (CHU Sainte-Justine Research Ethic Board, protocol and biobank no. 2126). No statistical method was used to predetermine sample size. One replicate was sequenced for all RNA-seq experiments. For MS, at least two replicates were analyzed. To assess the immunogenicity of EL4-derived TSAs, we measured the frequency and antigen avidity of T cells recognizing TSAs. In addition, we estimated the survival of 8- to 12-week-old C57BL/6 female mice that were immunized or not with individual TSAs. Investigators were not blinded during sample preparation or during data collection and analysis. For all in vitro and in vivo experiments described in this manuscript, at least three replicates were analyzed and found to be concordant with each other. No data were excluded from the analyses, and values are reported in table S7. The number of mice used, numbers of replicates, and statistical values (where applicable) are provided in the figure legends. For information regarding original RNA-seq and MS data, see the Data and materials availability section in Acknowledgements and table S2.

Statistical analysis
Procedures to evaluate statistical significance are described in the relevant figure legends. Overall, a log-rank test was used for survival curves, a Wilcoxon rank sum test with the Benjamini-Hochberg correction for multiple testing was used to compare T cell frequencies as estimated by tetramer and ELISpot, and a one-sided Wilcoxon rank sum test was used to compare T cell frequencies between immunized and rechallenged mice. $P \leq 0.05$ was considered significant.

Table S17. Features of human TSAs detected in B-ALL specimens.
Table S18. Features of human TSAs detected in lung tumor biopsies.
References (60–68)

REFERENCES AND NOTES

1. B. Mlecnik, G. Bindea, A. Kirilovsky, H. K. Angell, A. C. Obenauf, M. Tosolini, S. E. Church, P. Maby, A. Vasaturo, M. Angelova, T. Fredriksen, S. Mauger, M. Waldner, A. Berger, M. R. Speicher, F. Pagès, V. Valge-Archer, J. Galon, The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8**, 327ra326 (2016).

2. P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).

3. S. C. Wei, J. H. Levine, A. P. Cogdill, Y. Zhao, N.-A. S. Anang, M. C. Andrews, P. Sharma, J. Wang, J. A. Wargo, D. Pe'er, J. P. Allison, Distinct cellular mechanisms underlie anti-CTLA-4 and anti-PD-1 checkpoint blockade. *Cell* **170**, 1120–1133.e17 (2017).

4. W. Shao, P. G. A. Pedrioli, W. Wolski, C. Scurtescu, E. Schmid, J. A. Vizcaíno, M. Courcelles, H. Schuster, D. Kowalewski, F. Marino, C. S. L. Arlehamn, K. Vaughan, B. Peters, A. Sette, T. H. M. Ottenhoff, K. E. Meijgaarden, N. Nieuwenhuizen, S. H. E. Kaufmann, R. Schlapbach, J. C. Castle, A. I. Nesvizhskii, M. Nielsen, E. W. Deutsch, D. S. Campbell, R. L. Moritz, R. A. Zubarev, A. J. Ytterberg, A. W. Purcell, M. Marcilla, A. Paradela, Q. Wang, C. E. Costello, N. Ternette, P. A. van Veelen, C. A. C. M. van Els, A. J. R. Heck, G. A. de Souza, L. M. Sollid, A. Admon, S. Stevanovic, H.-G. Rammensee, P. Thibault, C. Perreault, M. Bassani-Sternberg, R. Aebersold, E. Caron, The SysteMHC Atlas project. *Nucleic Acids Res.* **46**, D1237–D1247 (2018).

5. S. D. Martin, G. Coukos, R. A. Holt, B. H. Nelson, Targeting the undruggable: Immunotherapy meets personalized oncology in the genomic era. *Ann. Oncol.* **26**, 2367–2374 (2015).

6. R. Marty, S. Kaabinejadian, D. Rossell, M. J. Slifker, J. van de Haar, H. B. Engin, N. de Prisco, T. Ideker, W. H. Hildebrand, J. Font-Burgada, H. Carter, MHC-I genotype restricts the oncogenic mutational landscape. *Cell* **171**, 1272–1283.e15 (2017).

7. M. Yarchoan, B. A. Johnson III, E. R. Lutz, D. A. Laheru, E. M. Jaffee, Targeting neoantigens to augment antitumour immunity. *Nat. Rev. Cancer* **17**, 209–222 (2017).

8. M. H. Gee, A. Han, S. M. Lofgren, J. F. Beausang, J. L. Mendoza, M. E. Birnbaum, M. T. Bethune, S. Fischer, X. Yang, R. Gomez-Eerland, D. B. Bingham, L. V. Sibener, R. A. Fernandes, A. Velasco, D. Baltimore, T. N. Schumacher, P. Khatri, S. R. Quake, M. M. Davis, K. C. Garcia, Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell* **172**, 549–563.e16 (2018).

9. G. Kassiotis, J. P. Stoye, Immune responses to endogenous retroelements: Taking the bad with the good. *Nat. Rev. Immunol.* **16**, 207–219 (2016).

10. P. Probst, J. Kopp, A. Oxenius, M. P. Colombo, D. Ritz, T. Fugmann, D. Neri, Sarcoma eradication by doxorubicin and targeted TNF relies upon CD8+ T-cell recognition of a retroviral antigen. *Cancer Res.* **77**, 3644–3654 (2017).

11. S. Goel, M. J. DeCristo, A. C. Watt, H. BrinJones, J. Sceneay, B. B. Li, N. Khan, J. M. Ubellacker, S. Xie, O. Metzger-Filho, J. Hoog, M. J. Ellis, C. X. Ma, S. Ramm, I. E. Krop, E. P. Winer, T. M. Roberts, H.-J. Kim, S. S. McAllister, J. J. Zhao, CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **548**, 471–475 (2017).

12. P. A. Ott, Z. Hu, D. B. Keskin, S. A. Shukla, J. Sun, D. J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T. A. Carter, S. Li, D. J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W. J. Lane, I. Javeri, K. Nellaiappan, A. M. Salazar, H. Daley, M. Seaman, E. I. Buchbinder, C. H. Yoon, M. Harden, N. Lennon, S. Gabriel, S. J. Rodig, D. H. Barouch, J. C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E. S. Lander, E. F. Fritsch, N. Hacohen, C. J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).

13. U. Sahin, E. Derhovanessian, M. Miller, B.-P. Kloke, P. Simon, M. Löwer, V. Bukur, A. D. Tadmor, U. Luxemburger, B. Schrörs, T. Omokoko, M. Vormehr, C. Albrecht, A. Paruzynski, A. N. Kuhn, J. Buck, S. Heesch, K. H. Schreeb, F. Müller, I. Ortseifer, I. Vogler, E. Godehardt, S. Attig, R. Rae, A. Breitkreuz, C. Tolliver, M. Suchan, G. Martic, A. Hohberger, P. Sorn, J. Diekmann, J. Ciesla, O. Waksmann, A.-K. Brück, M. Witt, M. Zillgen, A. Rothermel, B. Kasemann, D. Langer, S. Bolte, M. Diken, S. Kreiter, R. Nemecek, C. Gebhardt, S. Grabbe, C. Höller, J. Utikal, C. Huber, C. Loquai, Ö. Türeci, Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).

14. A.-H. Capietto, S. Jhunjhunwala, L. Delamarre, Characterizing neoantigens for personalized cancer immunotherapy. *Curr. Opin. Immunol.* **46**, 58–65 (2017).

15. The problem with neoantigen prediction. *Nat. Biotechnol.* **35**, 97 (2017).

16. H. Pearson, T. Daouda, D. P. Granados, C. Durette, E. Bonneil, M. Courcelles, A. Rodenbrock, J.-P. Laverdure, C. Côté, S. Mader, S. Lemieux, P. Thibault, C. Perreault, MHC class I–associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* **126**, 4690–4701 (2016).

17. M. Di Marco, J. K. Peper, H.-G. Rammensee, Identification of immunogenic epitopes by MS/MS. *Cancer J.* **23**, 102–107 (2017).

18. M. Yadav, S. Jhunjhunwala, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk, Z. Modrusan, I. Mellman, J. R. Lill, L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).

19. M. Bassani-Sternberg, G. Coukos, Mass spectrometry-based antigen discovery for cancer immunotherapy. *Curr. Opin. Immunol.* **41**, 9–17 (2016).

20. M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M. E. Martignoni, A. Werner, R. Hein, D. H. Busch, C. Peschel, R. Rad, J. Cox, M. Mann, A. M. Krackhardt, Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404 (2016).

21. M. M. Gubin, X. Zhang, H. Schuster, E. Caron, J. P. Ward, T. Noguchi, Y. Ivanova, J. Hundal, C. D. Arthur, W.-J. Krebber, G. E. Mulder, M. Toebes, M. D. Vesely, S. S. K. Lam, A. J. Korman, J. P. Allison, G. J. Freeman, A. H. Sharpe, E. L. Pearce, T. N. Schumacher, R. Aebersold, H.-G. Rammensee, C. J. Melief, E. R. Mardis, W. E. Gillanders, M. N. Artyomov, R. D. Schreiber, Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).

22. E. Khurana, Y. Fu, D. Chakravarty, F. Demichelis, M. A. Rubin, M. Gerstein, Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).

23. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

24. C. M. Laumont, T. Daouda, J.-P. Laverdure, É. Bonneil, O. Caron-Lizotte, M.-P. Hardy, D. P. Granados, C. Durette, S. Lemieux, P. Thibault, C. Perreault, Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat. Commun.* **7**, 10238 (2016).

25. C. M. Laumont, C. Perreault, Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy. *Cell. Mol. Life Sci.* **75**, 607–621 (2018).

26. S. A. Rosenberg, P. Tong-On, Y. Li, J. P. Riley, M. El-Gamil, M. R. Parkhurst, P. F. Robbins, Identification of BING-4 cancer antigen translated from an alternative open reading frame of a gene in the extended MHC class II region using lymphocytes from a patient with a durable complete regression following immunotherapy. *J. Immunol.* **168**, 2402–2407 (2002).

27. M. J. L. Kracht, M. van Lummel, T. Nikolic, A. M. Joosten, S. Laban, A. R. van der Slik, P. A. van Veelen, F. Carlotti, E. J. P. de Koning, R. C. Hoeben, A. Zaldumbide, B. O. Roep, Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nat. Med.* **23**, 501–507 (2017).

28. Y. Simoni, E. Becht, M. Fehlings, C. Y. Loh, S.-L. Koo, K. W. W. Teng, J. P. S. Yeong, R. Nahar, T. Zhang, H. Kared, K. Duan, N. Ang, M. Poidinger, Y. Y. Lee, A. Larbi, A. J. Khng, E. Tan, C. Fu, R. Mathew, M. Teo, W. T. Lim, C. K. Toh, B.-H. Ong, T. Koh, A. M. Hillmer, A. Takano, T. K. H. Lim, E. H. Tan, W. Zhai, D. S. W. Tan, I. B. Tan, E. W. Newell, Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).

29. J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, B. Ma, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).

30. J. A. Alfaro, A. Sinha, T. Kislinger, P. C. Boutros, Onco-proteogenomics: Cancer proteomics joins forces with genomics. *Nat. Methods* **11**, 1107–1113 (2014).

31. A. I. Nesvizhskii, Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125 (2014).

32. W. S. Noble, Mass spectrometrists should search only for peptides they care about. *Nat. Methods* **12**, 605–608 (2015).

33. Y. Takahama, I. Ohigashi, S. Baik, G. Anderson, Generation of diversity in thymic epithelial cells. *Nat. Rev. Immunol.* **17**, 295–305 (2017).

34. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).

35. J. C. Castle, M. Loewer, S. Boegel, J. de Graaf, C. Bender, A. D. Tadmor, V. Boisguerin, T. Bukur, P. Sorn, C. Paret, M. Diken, S. Kreiter, Ö. Türeci, U. Sahin, Immunomic, genomic

and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190 (2014).

36. P. Fontaine, G. Roy-Proulx, L. Knafo, C. Baron, D.-C. Roy, C. Perreault, Adoptive transfer of minor histocompatibility antigen-specific T lymphocytes eradicates leukemia cells without causing graft-versus-host disease. *Nat. Med.* **7**, 789–794 (2001).

37. R. Vita, J. A. Overton, C. J. Mungall, A. Sette, B. Peters, FAIR principles and the IEDB: Short-term improvements and a long-term vision of OBO-foundry mediated machine-actionable interoperability. *Database (Oxford)* **2018**, bax105 (2018).

38. A. Y. Huang, P. H. Gulden, A. S. Woods, M. C. Thomas, C. D. Tong, W. Wang, V. H. Engelhard, G. Pasternack, R. Cotter, D. Hunt, D. M. Pardoll, E. M. Jaffee, The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9730–9735 (1996).

39. F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L.-H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y.-C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K.-H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, B. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X.-Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren; The Mouse ENCODE Consortium, A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).

40. C. A. Sloan, E. T. Chan, J. M. Davidson, V. S. Malladi, J. S. Strattan, B. C. Hitz, I. Gabdank, A. K. Narayanan, M. Ho, B. T. Lee, L. D. Rowe, T. R. Dreszer, G. Roe, N. R. Podduturi, F. Tanaka, E. L. Hong, J. M. Cherry, ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).

41. M.-H. Fortier, É. Caron, M.-P. Hardy, G. Voisin, S. Lemieux, C. Perreault, P. Thibault, The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* **205**, 595–610 (2008).

42. K. Vincent, M.-P. Hardy, A. Trofimov, C. M. Laumont, D. Sriranganadane, S. Hadj-Mimoune, I. Salem Fourati, H. Soudeyns, P. Thibault, C. Perreault, Rejection of leukemic cells requires antigen-specific T cells with high functional avidity. *Biol. Blood Marrow Transplant.* **20**, 37–45 (2014).

43. M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, M. Mann, Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics* **14**, 658–673 (2015).

44. D. P. Granados, C. M. Laumont, P. Thibault, C. Perreault, The nature of self for T cells—A systems-level perspective. *Curr. Opin. Immunol.* **34**, 1–8 (2015).

45. W. Chen, L. C. Antón, J. R. Bennink, J. W. Yewdell, Dissecting the multifactorial causes of immunodominance in class I–restricted T cell responses to viruses. *Immunity* **12**, 83–93 (2000).

46. M. K. Jenkins, J. J. Moon, The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* **188**, 4135–4140 (2012).

47. J. J. Moon, H. H. Chu, M. Pepper, S. J. McSorley, S. C. Jameson, R. M. Kedl, M. K. Jenkins, Naive CD4+ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).

48. F. P. Legoux, J. J. Moon, Peptide:MHC tetramer-based enrichment of epitope-specific T cells. *J. Vis. Exp.* 4420 (2012).

49. N. Vigneron, V. Stroobant, B. J. Van den Eynde, P. van der Bruggen, Database of T cell-defined human tumor antigens: The 2013 update. *Cancer Immun.* **13**, 15 (2013).

50. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

51. V. P. Balachandran, M. Łuksza, J. N. Zhao, V. Makarov, J. A. Moral, R. Remark, B. Herbst, G. Askan, U. Bhanot, Y. Senbabaoglu, D. K. Wells, C. I. O. Cary, O. Grbovic-Huezo, M. Attiyeh, B. Medina, J. Zhang, J. Loo, J. Saglimbeni, M. Abu-Akeel, R. Zappasodi, N. Riaz, M. Smoragiewicz, Z. L. Kelley, O. Basturk; Australian Pancreatic Cancer Genome Initiative, M. Gönen, A. J. Levine, P. J. Allen, D. T. Fearon, M. Merad, S. Gnjatic, C. A. Iacobuzio-Donahue, J. D. Wolchok, R. P. DeMatteo, T. A. Chan, B. D. Greenbaum, T. Merghoub, S. D. Leach, Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).

52. M. J. Welters, T. C. van der Sluis, H. van Meir, N. M. Loof, V. J. van Ham, S. van Duikeren, S. J. Santegoets, R. Arens, M. L. de Kam, A. F. Cohen, M. I. van Poelgeest, G. G. Kenter, J. R. Kroep, J. Burggraaf, C. J. Melief, S. H. van der Burg, Vaccination during myeloid cell depletion by cancer chemotherapy fosters robust T cell responses. *Sci. Transl. Med.* **8**, 334ra352 (2016).

53. E. Caron, D. J. Kowalewski, C. Chiek Koh, T. Sturm, H. Schuster, R. Aebersold, Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics* **14**, 3105–3117 (2015).

54. D. Gfeller, M. Bassani-Sternberg, Predicting antigen presentation—What could we learn from a million peptides? *Front. Immunol.* **9**, 1716 (2018).

55. Y.-C. Lu, Z. Zheng, P. F. Robbins, E. Tran, T. D. Prickett, J. J. Gartner, Y. F. Li, S. Ray, Z. Franco, V. Bliskovsky, P. C. Fitzgerald, S. A. Rosenberg, An efficient single-cell RNA-seq approach to identify neoantigen-specific T cell receptors. *Mol. Ther.* **26**, 379–389 (2018).

56. Z. Hu, A. J. Anandappa, J. Sun, J. Kim, D. E. Leet, D. J. Bozym, C. Chen, L. Williams, S. A. Shukla, W. Zhang, D. Tabbaa, S. Steelman, O. Olive, K. J. Livak, H. Kishi, A. Muraguchi, I. Guleria, J. Stevens, W. J. Lane, U. E. Burkhardt, E. F. Fritsch, D. Neuberg, P. A. Ott, D. B. Keskin, N. Hacohen, C. J. Wu, A cloning and expression system to probe T-cell receptor specificity and assess functional avidity to neoantigens. *Blood* **132**, 1911–1921 (2018).

57. L. M. Kranz, M. Diken, H. Haas, S. Kreiter, C. Loquai, K. C. Reuter, M. Meng, D. Fritz, F. Vascotto, H. Hefesha, C. Grunwitz, M. Vormehr, Y. Hüsemann, A. Selmi, A. N. Kuhn, J. Buck, E. Derhovanessian, R. Rae, S. Attig, J. Diekmann, R. A. Jabulowsky, S. Heesch, J. Hassel, P. Langguth, S. Grabbe, C. Huber, Ö. Türeci, U. Sahin, Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396–401 (2016).

58. J. A. Vizcaíno, A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033 (2016).

59. J. Lanoix, C. Durette, M. Courcelles, E. Cossette, S. Comtois-Marotte, M.-P. Hardy, C. Côté, C. Perreault, P. Thibault, Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods. *Proteomics* **18**, e1700251 (2018).

60. M.-J. Kim, C. M. Miller, J. L. Shadrach, A. J. Wagers, T. Serwold, Young, proliferative thymic epithelial cells engraft and function in aging thymuses. *J. Immunol.* **194**, 4784–4795 (2015).

61. C. Stoeckle, I. A. Rota, E. Tolosa, C. Haller, A. Melms, E. Adamopoulou, Isolation of myeloid dendritic cells and epithelial cells from human thymus. *J. Vis. Exp.* e50951 (2013).

62. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

63. T. Daouda, C. Perreault, S. Lemieux, pyGeno: A Python package for precision medicine and proteogenomics. *F1000Res.* **5**, 381 (2016).

64. D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

65. E. Caron, K. Vincent, M.-H. Fortier, J.-P. Laverdure, A. Bramoullé, M.-P. Hardy, G. Voisin, P. P. Roux, S. Lemieux, P. Thibault, C. Perreault, The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol. Syst. Biol.* **7**, 533 (2011).

66. M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).

67. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

68. M. S. Bereman, J. Beri, V. Sharma, C. Nathe, J. Eckels, B. MacLean, M. J. MacCoss, An automated pipeline to monitor system performance in liquid chromatography–tandem mass spectrometry proteomic experiments. *J. Proteome Res.* **15**, 4763–4769 (2016).

CANCER

Enhanced detection of circulating tumor DNA by fragment size analysis

Florent Mouliere[1,2]*[†], Dineika Chandrananda[1,2]*, Anna M. Piskorz[1,2]*, Elizabeth K. Moore[1,2,3]*, James Morris[1,2], Lise Barlebo Ahlborn[4,5], Richard Mair[1,2,6], Teodora Goranova[1,2], Francesco Marass[1,2,7,8], Katrin Heider[1,2], Jonathan C. M. Wan[1,2], Anna Supernat[1,2,9], Irena Hudecova[1,2], Ioannis Gounaris[1,2,3], Susana Ros[1,2], Mercedes Jimenez-Linan[2,3], Javier Garcia-Corbacho[10], Keval Patel[1,2], Olga Østrup[5], Suzanne Murphy[1,2], Matthew D. Eldridge[1,2], Davina Gale[1,2], Grant D. Stewart[2,3,11], Johanna Burge[2,11], Wendy N. Cooper[1,2], Michiel S. van der Heijden[12,13], Charles E. Massie[1,2,14], Colin Watts[15], Pippa Corrie[3], Simon Pacey[3,14], Kevin M. Brindle[1,2,16], Richard D. Baird[17], Morten Mau-Sørensen[4], Christine A. Parkinson[1,2,3,18,19], Christopher G. Smith[1,2], James D. Brenton[1,2,3,18,19]‡[§], Nitzan Rosenfeld[1,2]‡[§]

Existing methods to improve detection of circulating tumor DNA (ctDNA) have focused on genomic alterations but have rarely considered the biological properties of plasma cell-free DNA (cfDNA). We hypothesized that differences in fragment lengths of circulating DNA could be exploited to enhance sensitivity for detecting the presence of ctDNA and for noninvasive genomic analysis of cancer. We surveyed ctDNA fragment sizes in 344 plasma samples from 200 patients with cancer using low-pass whole-genome sequencing (0.4×). To establish the size distribution of mutant ctDNA, tumor-guided personalized deep sequencing was performed in 19 patients. We detected enrichment of ctDNA in fragment sizes between 90 and 150 bp and developed methods for in vitro and in silico size selection of these fragments. Selecting fragments between 90 and 150 bp improved detection of tumor DNA, with more than twofold median enrichment in >95% of cases and more than fourfold enrichment in >10% of cases. Analysis of size-selected cfDNA identified clinically actionable mutations and copy number alterations that were otherwise not detected. Identification of plasma samples from patients with advanced cancer was improved by predictive models integrating fragment length and copy number analysis of cfDNA, with area under the curve (AUC) >0.99 compared to AUC <0.80 without fragmentation features. Increased identification of cfDNA from patients with glioma, renal, and pancreatic cancer was achieved with AUC > 0.91 compared to AUC < 0.5 without fragmentation features. Fragment size analysis and selective sequencing of specific fragment sizes can boost ctDNA detection and could complement or provide an alternative to deeper sequencing of cfDNA.

[1]Cancer Research UK Cambridge Institute, University of Cambridge, CB2 0RE Cambridge, UK. [2]Cancer Research UK Major Centre–Cambridge, Cancer Research UK Cambridge Institute, CB2 0RE Cambridge, UK. [3]Cambridge University Hospitals NHS Foundation Trust, CB2 0QQ Cambridge, UK. [4]Department of Oncology, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. [5]Centre for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. [6]Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, CB2 0QQ Cambridge, UK. [7]Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. [8]Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. [9]Department of Medical Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, 80-211 Gdańsk, Poland. [10]Clinical Trials Unit, Clinic Institute of Haematological and Oncological Diseases, Hospital Clinic de Barcelona, 170 08036 Barcelona, Spain. [11]Academic Urology Group, Department of Surgery, University of Cambridge, CB2 0QQ Cambridge, UK. [12]Division of Molecular Carcinogenesis, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [13]Department of Medical Oncology, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [14]Department of Oncology, University of Cambridge, CB2 0XZ Cambridge, UK. [15]Institute of Cancer Genomics Science, University of Birmingham, B15 2TT Birmingham, UK. [16]Department of Biochemistry, University of Cambridge, CB2 1QW Cambridge, UK. [17]Early Phase Clinical Trials and Breast Cancer Research Teams, Cancer Research UK Cambridge Centre, CB2 0QQ Cambridge, UK. [18]Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, CB2 0XZ Cambridge, UK. [19]NIHR Cambridge Biomedical Research Centre, CB2 0QQ Cambridge, UK.
*These authors contributed equally to this work as co-first authors.
†Present address: Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Pathology, Cancer Center Amsterdam, de Boelelaan 1117, 1081 HV Amsterdam, Netherlands.
‡These authors contributed equally to this work as co-senior authors.
§Corresponding author. Email: james.brenton@cruk.cam.ac.uk (J.D.B.); nitzan.rosenfeld@cruk.cam.ac.uk (N.R.)

INTRODUCTION

Blood plasma of patients with cancer contains circulating tumor DNA (ctDNA), but this valuable source of information is diluted by much larger quantities of DNA of noncancerous origins, such that ctDNA usually represents only a small fraction of the total cell-free DNA (cfDNA) (*1*, *2*). High-depth targeted sequencing of selected genomic regions can be used to detect low amounts of ctDNA, but broader analysis with methods such as whole-exome sequencing (WES) and shallow whole-genome sequencing (sWGS) is only generally informative when ctDNA content is ~10% or greater (*3*–*5*). The concentration of ctDNA can exceed 10% of the total cfDNA in patients with advanced-stage cancers (*6*–*8*), but is much lower in patients with low tumor burden (*9*–*12*) and in patients with some cancer types such as gliomas and renal cancers (*6*). Current strategies to improve ctDNA detection rely on increasing depth of sequencing coupled with various error correction methods (*2*, *13*, *14*). However, approaches that focus only on genomic alterations do not take advantage of the potential differences in chromatin organization or fragment sizes of ctDNA (*15*–*17*). Results of ever-deeper sequencing are also confounded by the likelihood of false-positive results from detection of mutations from noncancerous cells, clonal expansions in normal epithelia, or clonal hematopoiesis of indeterminate potential (CHIP) (*13*, *18*, *19*).

The cell of origin and the mechanism of cfDNA release into blood can mark cfDNA with specific fragmentation signatures, potentially

providing precise information about cell type, gene expression, cell physiology or pathology, or action of treatment (*15*, *16*, *20*). cfDNA fragments commonly show a prominent mode at 167 bp, suggesting release from apoptotic caspase-dependent cleavage (Fig. 1A) (*21–24*). Circulating fetal DNA has been shown to be shorter than maternal DNA in plasma, and these size differences have been used to improve sensitivity of noninvasive prenatal diagnosis (*22*, *25–27*). The size distribution of tumor-derived cfDNA has only been investigated in a few studies, encompassing a small number of cancer types and patients, and showed conflicting results (*28–33*). A limitation of previous studies is that determining the specific sizes of tumor-derived DNA fragments requires detailed characterization of matched tumor-derived alterations (*30*, *33*), and the broader understanding and implications of potential biological differences have not previously been explored.

We hypothesized that we could improve the sensitivity for noninvasive cancer genomics by selective sequencing of ctDNA fragments and by leveraging differences in the biology that determine DNA fragmentation. To test this, we established a pan-cancer catalog of cfDNA fragmentation features in plasma samples from patients with different cancer types and healthy individuals to identify biological features enriched in tumor-derived DNA. We developed methods for selecting specific sizes of cfDNA fragments before sequencing and investigated the impact of combining cfDNA size selection with genome-wide sequencing to improve the detection of ctDNA and the identification of clinically actionable genomic alterations.

RESULTS

Surveying the fragmentation features of tumor cfDNA
We generated a catalog of cfDNA fragmentation features (Fig. 1A) in 344 plasma samples from 200 patients with 18 different cancer types and additional 65 plasma samples from healthy controls (Fig. 1B, fig. S1, and tables S1 and S2). The size distribution of cfDNA fragments in patients with cancer differed in the size ranges of 90 to 150 bp, 180 to 220 bp, and 250 to 320 bp compared to healthy individuals (Fig. 1B and fig. S2). cfDNA fragment sizes in plasma of healthy individuals and in plasma of patients with late-stage glioma, renal, pancreatic, and bladder cancers were significantly longer than in other late-stage cancer types including breast, ovarian, lung, melanoma, colorectal, and cholangiocarcinoma (Kruskal-Wallis, *P* < 0.001; Fig. 1C). Sorting the 18 cancer types according to the proportion of cfDNA fragments in the size range of 20 to 150 bp resulted in an order very similar to that obtained by Bettegowda *et al.* (*6*) based on the concentrations of ctDNA measured by individual mutation assays (Fig. 1D). In contrast to previous reports (*6*, *34*), this sorting was performed without any analysis or prior knowledge of the presence of mutations or somatic copy number alterations (SCNAs) yet allowed the investigation of ctDNA content in different cancers.

Sizing up mutant ctDNA
We determined the size profile of mutant ctDNA in plasma using two high-specificity approaches. First, we inferred the specific size profile of ctDNA and nontumor cfDNA with sWGS from the plasma of mice bearing human ovarian cancer xenografts (Fig. 2A). We observed a shift in ctDNA fragment sizes to less than 167 bp (Fig. 2B). Second, the size profile of mutant ctDNA was determined in plasma from 19 patients with cancer, using deep sequencing with patient-

specific hybrid-capture panels developed from whole-exome profiling of matched tumor samples (Fig. 2C). By sequencing hundreds of mutations at a depth of >300× in cfDNA, we obtained allele-specific reads from mutant and normal DNA. Enrichment of DNA fragments carrying tumor-mutated alleles was observed in fragments between ~20 and 40 bp shorter than nucleosomal DNA sizes (multiples of 167 bp; Fig. 2D). We determined that mutant ctDNA is generally more fragmented than nonmutant cfDNA, with a maximum enrichment of ctDNA in fragments between 90 and 150 bp (fig. S3), as well as enrichment in the size range of 250 to 320 bp. These data also indicated that mutant DNA in plasma of patients with advanced cancer (before treatment) is consistently shorter than predicted mononucleosomal and dinucleosomal DNA fragment lengths (Fig. 2D).

Selecting tumor-derived DNA fragments
We evaluated whether the shorter cfDNA fragments in plasma can be harnessed to improve ctDNA detection. We determined the feasibility of selective sequencing of shorter fragments using in vitro size selection with a bench-top microfluidic device followed by sWGS in 48 plasma samples from 35 patients with high-grade serous ovarian cancer (HGSOC; Fig. 3A and figs. S4 and S5). We assessed the accuracy and quality of the size selection with the plasma from 20 healthy individuals (Fig. 3B and fig. S6). We also explored the utility of in silico size selection of fragmented DNA using read-pair positioning from unprocessed sWGS data (Fig. 3A). In silico size selection was performed once reads were aligned to the genome reference, by selecting the paired-end reads that corresponded to the fragment lengths in a 90- to 150-bp size range. Figure 3 (C to E) shows the effect of in vitro size selection for one HGSOC case (see all five samples in figs. S7 and S8). First, we identified SCNAs in plasma cfDNA before treatment, when the concentration of ctDNA was high (Fig. 3C). Only a small number of focal SCNAs were observed in the subsequent plasma sample collected 3 weeks after initiation of chemotherapy (without size selection; Fig. 3D). In vitro size selection of the same posttreatment plasma sample showed a median increase of 6.4× in the amplitude of detectable SCNAs without size selection. Selective sequencing of shorter fragments in this sample resulted in the detection of multiple other SCNAs that were not observed without size selection (Fig. 3E) and a genome-wide copy number profile that was similar to that obtained before treatment when ctDNA concentrations were four times higher, with additional copy number alterations identified in this sample despite the lower initial concentration of ctDNA (Fig. 3C). In silico size selection also enriched ctDNA but to a lower extent than using in vitro size selection (fig. S7). We concluded that selecting short DNA fragments in plasma can enrich tumor content on a genome-wide scale.

Quantifying the impact of size selection
To quantitatively assess the enrichment after size selection on a genome-wide scale, we developed a metric from sWGS data (<0.4× coverage) called t-MAD (trimmed median absolute deviation from copy number neutrality; see Fig. 4A). All sWGS data were down sampled to 10 million sequencing reads for comparison. To define the detection threshold, we measured the t-MAD score for sWGS data from 65 plasma samples from 46 healthy individuals and took the maximal value (median, 0.01; range, 0.004 to 0.015). We compared t-MAD to the mutant allele fraction (MAF) in high ctDNA cancer types as assessed by digital polymerase chain reaction (dPCR)
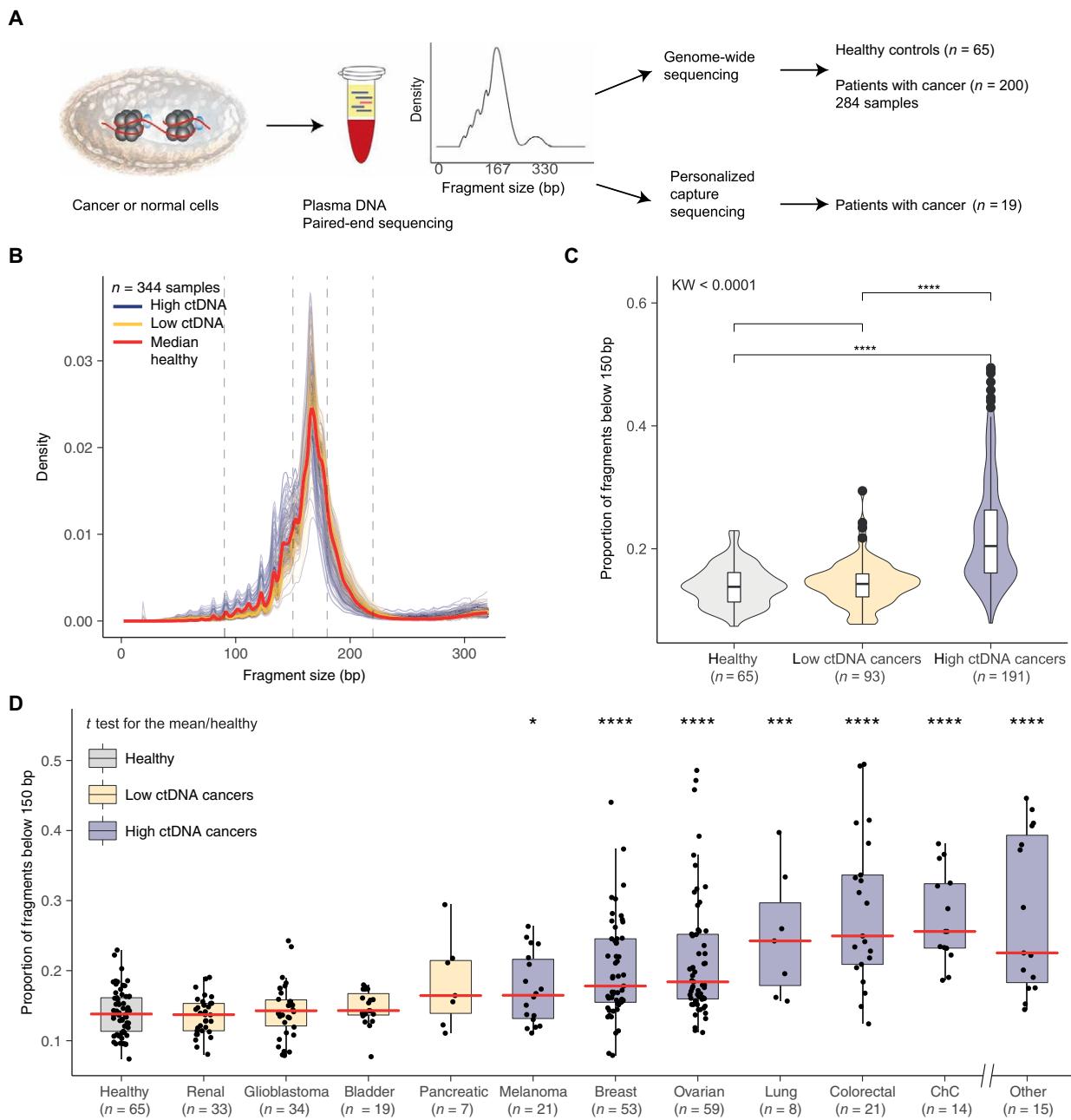
Fig. 1. Survey of plasma DNA fragmentation with genome-wide sequencing on a pan-cancer scale. (**A**) The size profile of cfDNA can be determined by paired-end sequencing of plasma samples and reflects its organization around the nucleosome. cfDNA is released into the blood circulation by various means, each of which leaves a signature on the DNA fragment sizes. We inferred the size profile of cfDNA by analyzing with sWGS ($n = 344$ plasma samples from 65 healthy controls and 200 patients with cancer) and the size profile of mutant ctDNA by personalized capture sequencing ($n = 19$ plasma samples). (**B**) Fragment size distributions of 344 plasma samples from 200 patients with cancer. Samples are split into two groups based on the previous literature (6), with orange representing samples from patients with cancer types previously observed to have low amounts of ctDNA (renal, bladder, pancreatic, and glioma) and blue representing samples from patients with cancer types previously observed to have higher amounts of ctDNA (breast, melanoma, ovarian, lung, colorectal, cholangiocarcinoma, and others; see table S1). (**C**) Proportion of cfDNA fragments below 150 bp in those samples, grouped into cancer types as defined in (B). The Kruskal-Wallis (KW) test for difference in size distributions indicated a significant difference between the group of samples from cancer types releasing high amounts of ctDNA and the group of samples from cancer types releasing low amounts, as well as the group of samples from healthy individuals. (**D**) Proportion of cfDNA fragments below 150 bp by cancer type (all samples). Cancer types represented by fewer than four individuals are grouped in the "other" category. Red lines indicate the median proportion for each cancer type. ChC, cholangiocarcinoma. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

or WES in 97 samples. We observed a high correlation (Pearson correlation, $r = 0.80$) between t-MAD and MAF (Fig. 4B) for samples with t-MAD greater than the detection threshold (0.015) or with MAF > 0.025. Figure S9 shows that the slope of t-MAD versus MAF fit lines differed between cancer types (range, 0.17 to 1.12), likely reflecting differences in the extent of SCNAs. We estimated the

Fig. 2. Determining the size profile of mutant ctDNA with animal models and personalized capture sequencing. (**A**) A mouse model with xenografted human tumor cells enabled the discrimination of DNA fragments released by cancer cells (reads aligning to the human genome) from the DNA released by healthy cells (reads aligning to the mouse genome), with the use of sWGS. (**B**) Fragment size distribution from the plasma extracted from a mouse xenografted with a human ovarian tumor, showing ctDNA originating from tumor cells (red) and cfDNA from noncancerous cells (blue). Two vertical dashed lines indicate 145 and 167 bp. The fraction of reads shorter than 150 bp is indicated. (**C**) Design of personalized hybrid-capture sequencing panels developed to specifically determine the size profiles of mutant DNA and nonmutant DNA in plasma from 19 patients with late-stage cancers. Capture panels included somatic mutations identified in tumor tissue by WES. A mean of 165 mutations per patient was then analyzed from matched plasma samples. Reads were aligned and separated into fragments carrying either the reference or the mutant sequence. Fragment sizes for paired-end reads were calculated. (**D**) Size profiles of mutant DNA and nonmutant DNA in plasma from 19 patients with late-stage cancers were determined by tumor-guided capture sequencing. The fraction of reads shorter than 150 bp is indicated.
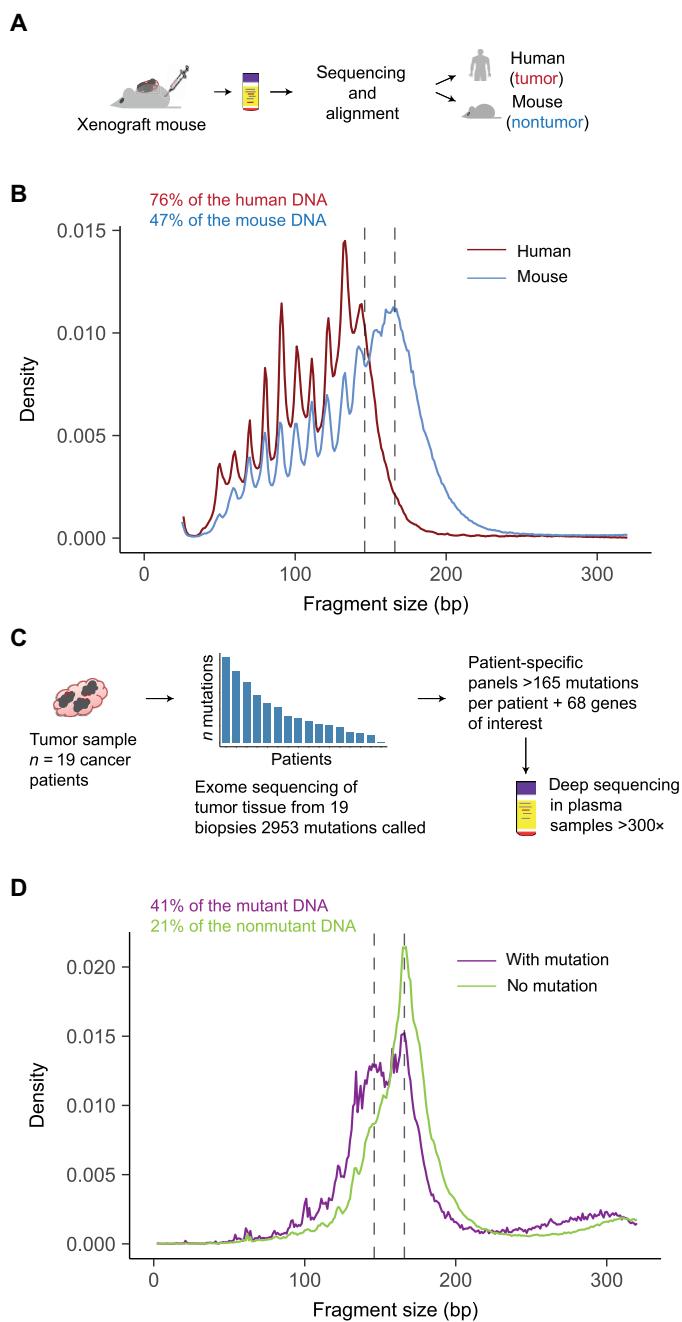
sensitivity of t-MAD for detecting low amounts of ctDNA using a spike-in dilution of DNA from a patient with a *TP53* mutation into DNA from a pool of seven healthy individuals (fig. S10), which confirmed that the t-MAD score was linear with ctDNA fraction down to MAF of ~0.01. In addition, t-MAD scores greater than the detection threshold (0.015) for samples were present even in samples with MAF as low as 0.004. t-MAD was also strongly correlated with tumor volume determined by RECIST1.1 (Pearson correlation, $r = 0.6$; $P < 0.0001$; $n = 35$; fig. S11).

Using t-MAD, we detected ctDNA from 69% (130 of 189) of the samples from cancer types where ctDNA concentrations were shown to be high (Fig. 4C). From cancer types for which ctDNA concentrations are suspected to be low (glioma, renal, bladder, and pancreatic), we detected ctDNA in 17% (10 of 57) of the cases (Fig. 4C). We used in silico size selection of the DNA fragments between 90 and 150 bp from the high ctDNA cancers ($n = 189$) and healthy controls ($n = 65$) to improve the sensitivity for detecting t-MAD (Fig. 4D). Receiver operating characteristic (ROC) analysis comparing the t-MAD score for the samples revealed an area under the curve (AUC) of 0.90 after in silico size selection, against an AUC of 0.69 without size selection (Fig. 4D).

We explored whether size-selected sequencing could improve the detection of response or disease progression. We used sWGS of longitudinal plasma samples from six patients with cancer (Fig. 4, E and F) and in silico size selection of the cfDNA fragments between 90 and 150 bp. In two patients, size-selected samples indicated tumor progression 60 and 87 days before detection by imaging or unselected t-MAD analysis (Fig. 4, E and F). Other longitudinal samples exhibited improvements in the detection of ctDNA with t-MAD and size selection (Fig. 4F).

Identifying more clinically relevant genomic alterations with size selection

We next tested whether size selection could increase the sensitivity for detecting cancer genomic alterations in cfDNA. To test effects on copy number aberrations, we studied 35 patients with HGSOC as the archetypal copy number–driven cancer (*35*). t-MAD was used to quantify the enrichment of ctDNA with in vitro size selection in 48 plasma samples, including samples collected before and after initiation of chemotherapy treatment. In vitro size selection resulted

in an increase in the calculated t-MAD score from the sWGS data for 47 of 48 of the plasma samples (98%; *t* test, $P = 0.06$) with a mean of 2.5 and median of 2.1-fold increase (Fig. 5A and table S3). We compared the t-MAD scores against those obtained by sWGS for the plasma samples from healthy individuals. Thirty-nine of the 48 size-selected HGSOC plasma samples (82%) had a t-MAD score greater than the highest t-MAD value determined in the in vitro size-selected healthy plasma samples (Fig. 5A and figs. S6 and S12), compared to 24 of 48 without size selection (50%). ROC analysis comparing the t-MAD score for the samples from patients with cancer (pre- and posttreatment initiation, $n = 48$) and healthy controls ($n = 46$) revealed an AUC of 0.97 after in vitro size selection, with maximal sensitivity and specificity of 90 and 98%, respectively.
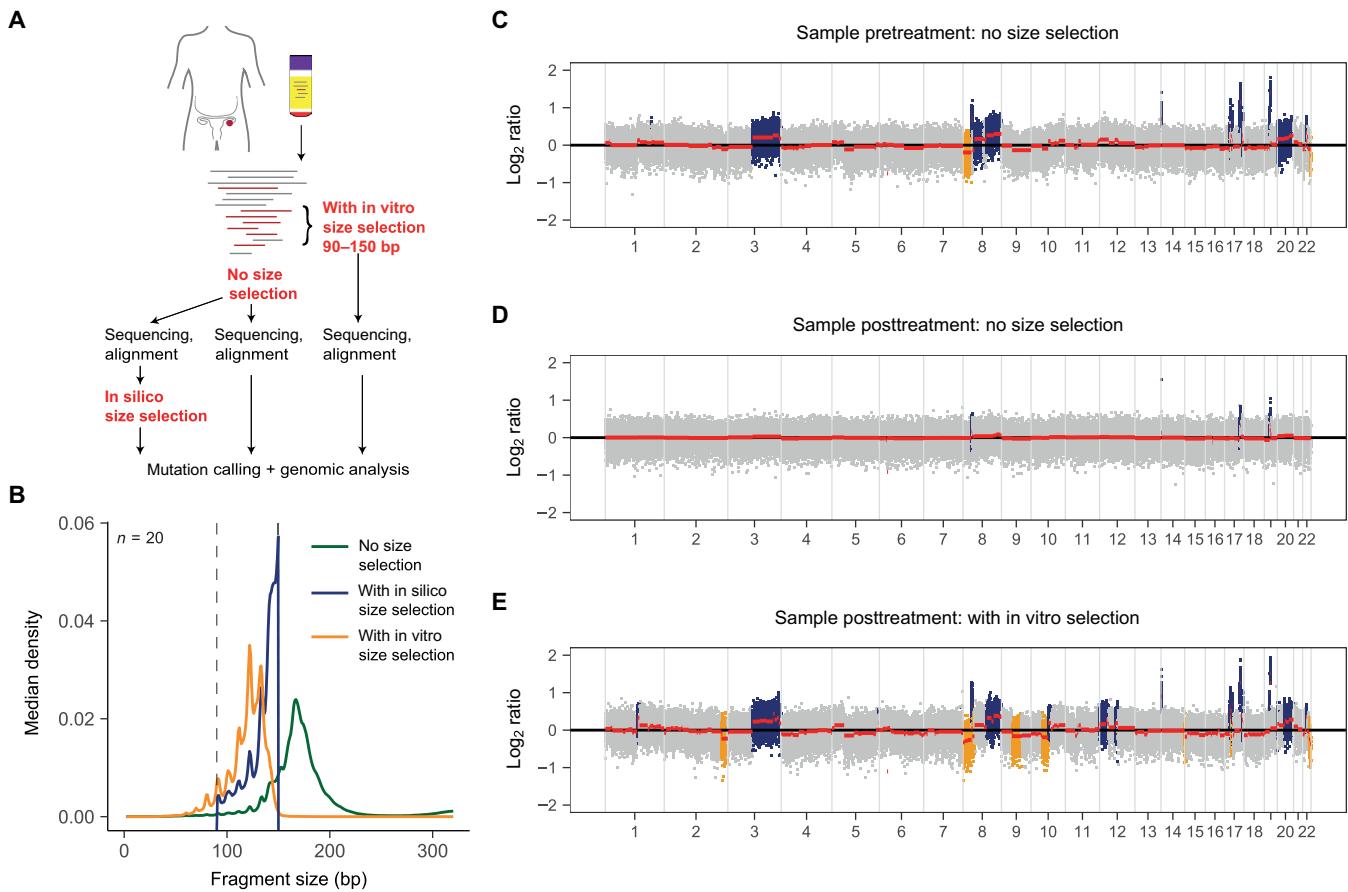
Fig. 3. Enhancing the tumor fraction from plasma sequencing with size selection. (**A**) Plasma samples collected from patients with ovarian cancer were analyzed in parallel without size selection or using either in silico or in vitro size selection. (**B**) Accuracy of the in vitro and in silico size selection determined on a cohort of 20 healthy controls. The size distribution before size selection is shown in green, after in silico size selection (with sharp cutoff at 90 and 150 bp) in blue and after in vitro size selection in orange. Vertical lines indicate 90 and 150 bp. (**C**) SCNA analysis with sWGS from plasma DNA of a patient with ovarian cancer collected before initiation of treatment, when ctDNA MAF was 0.271 for a *TP53* mutation as determined by tagged-amplicon deep sequencing (TAm-Seq). Inferred amplifications are shown in blue and deletions in orange. Copy number neutral regions are shown in gray. (**D**) SCNA analysis of a plasma sample from the same patient as in (C), collected 3 weeks after treatment start. The MAF for the *TP53* mutation at this time point was 0.068, and sWGS revealed only limited evidence of copy number alterations (before size selection). (**E**) Analysis of the same plasma sample as in (D) after in vitro size selection of fragments between 90 and 150 bp in length. The MAF for the *TP53* mutation increased to 0.402 after in vitro size selection, and SCNAs were apparent by sWGS. More SCNAs were detected in comparison to (C) and (D) (for example, in chr2, chr9, and chr10). SCNAs were also detected in this sample after in silico size selection (fig. S7).

This was superior to detection by sWGS without size selection (AUC, 0.64; Fig. 5B).

We then determined whether this improved sensitivity resulted in the detection of SCNAs with potential clinical value. Across the genome, t-MAD scores evaluating SCNAs were higher after size selection in 33 of 35 (94%) patients with HGSOC, and the magnitude of copy number ($\log_2$ ratio) values significantly increased after in vitro size selection (*t* test for the means, $P = 0.003$; Fig. 5C). We compared the relative copy number values for 15 genes frequently altered in HGSOC (table S4). Analysis of plasma cfDNA after size selection revealed a large number of SCNAs that were not observed in the same samples without size selection (Fig. 5D), including amplifications in key genes such as *NF1*, *TERT*, and *MYC* (fig. S13).

We also tested whether similar enrichment was seen for substitutions to exclude the possibility that size selection might only increase the sensitivity for sWGS analysis. We performed WES of plasma cfDNA from 23 patients with seven cancer types (fig. S1). We used the WES data to compare the size distributions of fragments carrying mutant or nonmutant alleles (Fig. 6A) and to test whether size selection could identify additional mutations. We first selected six patients with HGSOC and performed WES of plasma DNA with and without in vitro size selection in the range of 90 to 150 bp, analyzing time points before and after initiation of treatment (*36*). In addition, in silico size selection for the same range of fragment sizes was performed (Fig. 6A). Analysis of the MAF of SNVs revealed statistically significant enrichment of the tumor fraction with both in vitro size selection (mean, 4.19-fold; median, 4.27-fold increase; *t* test, $P < 0.001$) and in silico size selection (mean, 2.20-fold; median, 2.25-fold increase; *t* test, $P < 0.001$; Fig. 6A and fig. S14). Three weeks after initiation of treatment, ctDNA fractions are often lower (*36*), and therefore, we further analyzed posttreatment plasma samples using TAm-Seq (*37*). We observed enrichment of MAFs by in vitro size selection between 0.9 and 11 times (mean, 2.1 times; median, 1.5 times), with one outlier sample exhibiting a relative enrichment of 118 times compared to the same samples without size selection (fig. S15).

A Patient plasma

sWGS

>0.5 M reads per sample

Copy number

Quantify amount of genome-wide copy number aberration

t-MAD score →
1. Represents tumor fraction
2. Quantifies effect of size selection

B

$r = 0.80$, $P = 6.3 \times 10^{-14}$, slope = 0.44

t-MAD < 0.015 ∩ MAF < 0.025
t-MAD > 0.015 ∪ MAF > 0.025

t-MAD score

MAF

C

High ctDNA cancers
Low ctDNA cancers
Healthy

t-MAD score

10/57
17%

130/189
69%

97.5% specificity

Samples

D

Sensitivity

t-MAD (AUC = 0.69)
t-MAD with in silico size selection (AUC = 0.90)

1 - Specificity

E

PR PR PR SD SD PD

t-MAD score

Days

No size selection
With in silico size selection
Time point detected by size selection

F

Treatment start

PC1
CRC1
ChC2
BC3
BC10
BC30

Days posttreatment start

○ Nondetected ● Detected by in silico size selection ● Detected ▮ Progressive disease ▮ Stable disease/partial response

Fig. 4. Quantifying the ctDNA enrichment by sWGS with in silico size selection and t-MAD. (**A**) Workflow to quantify tumor fraction from SCNA as a genome-wide score named t-MAD. (**B**) Correlation between the MAF of single-nucleotide variants (SNVs) determined by dPCR or hybrid-capture sequencing and t-MAD score determined by sWGS. Data included 97 samples from patients with multiple cancer types with matched MAF measurements and t-MAD scores. Pearson correlation (coefficient r) between MAF and t-MAD scores was calculated for all cases with MAF > 0.025 and t-MAD > 0.015. Linear regression indicated a fit with a slope of 0.44 (purple solid line). (**C**) Comparison of t-MAD scores determined from sWGS between healthy samples and samples collected from patients with cancer types that exhibit low amounts of ctDNA and from patients with cancer types that exhibit high amounts of ctDNA (as in Fig. 1). All samples for which t-MAD could be calculated have been included. (**D**) ROC analysis comparing the classification of these plasma samples from high ctDNA cancer samples ($n = 189$) and plasma samples from healthy controls ($n = 65$) using t-MAD had an AUC of 0.69 without size selection (black solid curve). After applying in silico size selection to the samples from patients with cancer, we observed an AUC of 0.90 (black dashed curve). (**E**) Determination of t-MAD from longitudinal plasma samples of a patient with colorectal cancer. t-MAD was analyzed before and after in silico size selection of the DNA fragments between 90 and 150 bp and then compared to the RECIST status for this patient. PR, partial response; SD, stable disease; PD, progressive disease. (**F**) Application of in silico size selection to six patients with long-term follow-up. t-MAD score was determined before and after in silico size selection of the short DNA fragments. Dark blue circles indicate samples in which ctDNA was detected both with and without in silico size selection. Light blue circles indicate samples where ctDNA was detected only after in silico size selection. Open circles indicate samples where ctDNA was not detected by either analysis. Times when RECIST status was assessed are indicated by a red bar for progression or an orange bar for regression or stable disease. PC, prostate cancer; CRC, colorectal cancer; ChC, cholangiocarcinoma; BC, breast cancer. The numbers correspond to the patients.

Size selection with both in vitro and in silico methods increased the number of mutations detected by WES by an average of 53% compared to no size selection (Fig. 6B). We identified a total of 1023 mutations in the samples without size selection. An additional 260 mutations were detected by in vitro size selection, and an additional 310 mutations were called after in silico size selection (Fig. 6B and table S5). To exclude the possibility that the improved sensitivity for mutation detection was a result of sequencing artifacts, we validated whether new mutations were also detectable in tumor specimens. We used in silico size selection in an independent cohort of 16 patients for whom matched tumor tissue DNA was available (table S6). In silico size selection enriched the MAF for nearly all mutations (2061 of 2133, 97%), with an average increase of MAF of 1.7× (Fig. 6C). For 13 of 16 patients (81%), we identified additional mutations in plasma after in silico size selection. Of these 82 additional mutations, 23 (28%) were confirmed to be present in the matched tumor tissue DNA (Fig. 6D). This included mutations in key cancer genes including *BRAF*, *ARID1A*, and *NF1* (fig. S16).

Detecting cancer by supervised machine learning combining cfDNA fragmentation and somatic alteration analysis

Although in vitro and in silico size selection increase the sensitivity of detection, they also result in a loss of cfDNA for analysis. In analysis of ctDNA based on genomic signals, potentially informative data are lost because regions of the cancer genome that are not mutated or altered do not contribute to detection (fig. S17). We hypothesized that leveraging other biological properties of the cfDNA fragmentation profile could enhance the detection of ctDNA.

We defined other cfDNA fragmentation features from sWGS data including (i) the proportion of fragments in multiple size ranges, (ii) the ratios of proportions of fragments in different sizes, and (iii) the amplitude of oscillations in fragment size density with 10-bp periodicity (see Materials and Methods and Fig. 7A). These fragmentation features were compared between patients with cancer and healthy individuals (fig. S18),
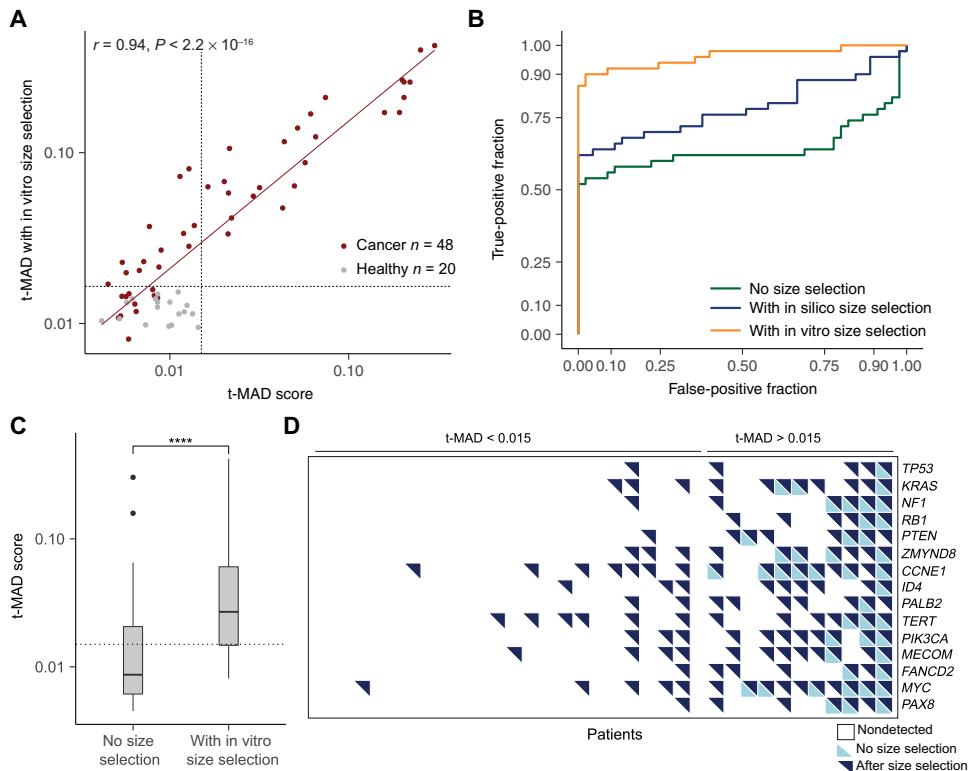
Fig. 5. Quantifying the ctDNA enrichment by sWGS with in vitro size selection. (**A**) The effect of in vitro size selection on the t-MAD score. For each of 48 plasma samples collected from 35 patients, the t-MAD score was determined from the sWGS after in vitro size selection (y axis) and without size selection (x axis). In vitro size selection increased the t-MAD score for nearly all samples, with a median increase of 2.1-fold (range from 1.1- to 6.4-fold). t-MAD scores determined from sWGS for 46 samples from healthy individuals were all <0.015 both before and after in vitro size selection. (**B**) ROC analysis comparing the classification of plasma samples from patients with cancer ($n = 48$) and plasma samples from healthy controls ($n = 46$) using t-MAD had an AUC of 0.64 without size selection (green curve). After applying in silico size selection to the samples from the patients and controls, we observed an AUC of 0.78 (blue curve), and after in vitro size selection, an AUC of 0.97 (orange curve). (**C**) Comparison of t-MAD scores determined from sWGS between matched ovarian cancer samples with and without in vitro size selection. The t test for the difference in means indicates a significant increase in tumor fraction (measured by t-MAD) with in vitro size selection (****$P < 0.0001$). (**D**) Detection of SCNAs across 15 genes frequently mutated in recurrent ovarian cancer, measured in plasma samples collected during treatment for 35 patients. Patients were ranked from left to right by increasing tumor fraction as quantified by t-MAD (before in vitro size selection). SCNAs were labeled as detected for a gene if the mean $\log_2$ ratio in that region was greater than 0.05. Empty squares represent copy number neutral regions, bottom left triangles in light blue indicate that SCNAs were detected without size selection, and top right triangles in dark blue represent SCNAs detected after in vitro size selection.

and the feature representing the proportion (P) of fragments between 20 and 150 bp exhibited the highest AUC (0.819). Principal components analysis (PCA) of the samples represented by t-MAD and fragmentation features showed a separation between healthy samples and samples from patients with cancer and identified fragment features that were aligned (in PCA) with t-MAD scores (Fig. 7B).

We next explored the potential of fragmentation features to enhance the detection of tumor DNA in plasma samples. A predictive analysis was performed using the t-MAD score and nine fragmentation features across 304 samples (239 from patients with cancer and 65 from healthy controls; Fig. 7C, fig. S19, and table S2). The nine fragmentation features determined from sWGS included five features based on the proportion (P) of fragments in defined size ranges: P(20 to 150), P(100 to 150), P(160 to 180), P(180 to 220),

and P(250 to 320); three features based on ratios of those proportions: P(20 to 150)/P(160 to 180), P(100 to 150)/P(163 to 169), and P(20 to 150)/P(180 to 220); and a further feature based on the amplitude of the oscillations having 10-bp periodicity observed below 150 bp.

Variable selection and the classification of samples as "healthy" or "cancer" were performed using logistic regression (LR) and random forest (RF) models trained on 153 samples and validated on two datasets of 94 and 83 independent samples (Fig. 7C). The best feature set for the LR model included t-MAD, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320). The same five variables were independently identified using the RF model (with some differences in their ranking). Figure S20 shows performance metrics for the different algorithms on training set data using cross-validation. Using t-MAD alone in the validation pan-cancer dataset (Fig. 7D and fig. S19), we could distinguish cancer samples from healthy individuals with an AUC of 0.764. Using the LR model improved the classification of the samples to an AUC of 0.908. The RF model (trained on the 153-sample training set) could distinguish cancer from healthy individuals even more accurately in the validation dataset ($n = 94$) with an AUC of 0.994. On the second validation dataset containing low-ctDNA cancer samples ($n = 83$; Fig. 7E), t-MAD alone or the LR performed less well, with AUC values of 0.421 and 0.532, respectively. However, the RF model was still able to distinguish low-ctDNA cancer samples from healthy controls with an AUC of 0.914. At a specificity of 95%, the RF model correctly classified as cancer in 64 of 68 (94%) of the samples from high-ctDNA

cancers (colorectal, cholangiocarcinoma, ovarian, breast, and melanoma) and 37 of 57 (65%) of the samples from low-ctDNA cancers (pancreatic, renal, and glioma; Fig. 7F). In a second iteration of model training, we omitted t-MAD using only the four fragmentation features (fig. S21). The RF model could still distinguish cancer from healthy controls, albeit with slightly reduced AUCs (0.989 for cancer types with high amounts of ctDNA and 0.891 for cancer types with low amounts of ctDNA), suggesting that the cfDNA fragmentation pattern is the most important predictive component.

DISCUSSION
Our results indicate that exploiting fundamental properties of cfDNA with fragment-specific analyses can allow more sensitive evaluation
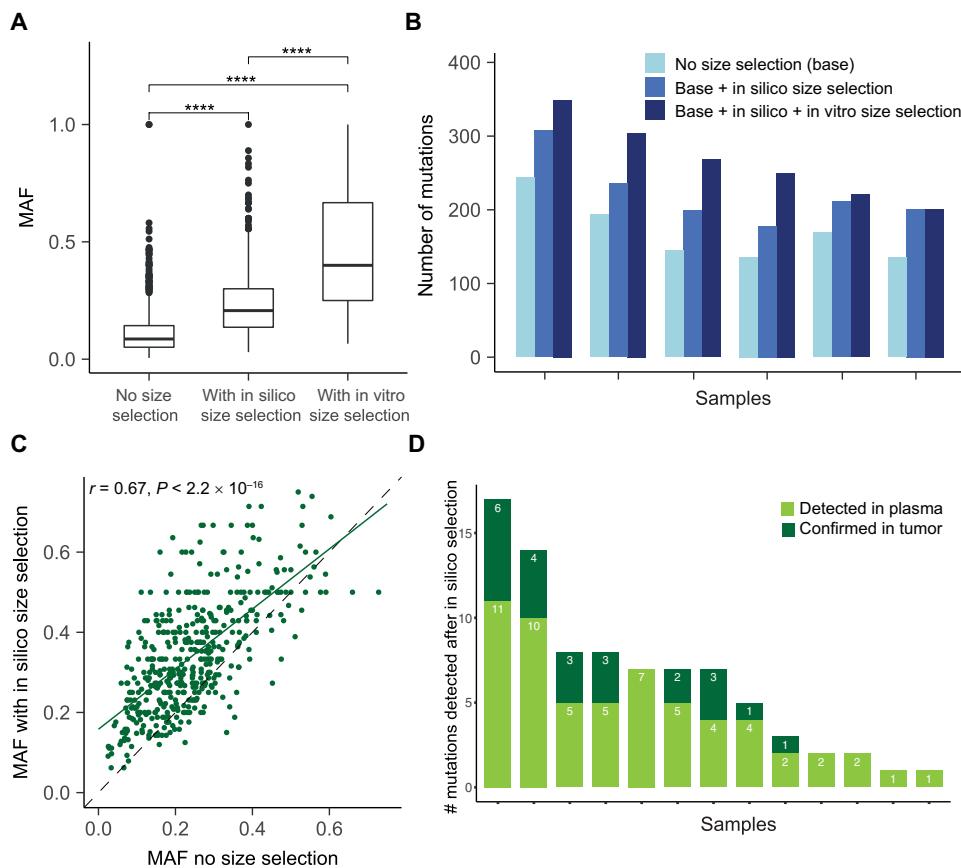
Fig. 6. Improving the detection of somatic alterations by WES in multiple cancer types with size selection. (**A**) Analysis of the MAF of mutations detected by WES in six patients with HGSOC without size selection and with either in vitro or in silico size selection. ****$P < 0.0001$. (**B**) Comparison of size-selected WES data with nonselected WES data to assess the number of mutations detected in plasma samples from six patients with HGSOC. For each patient, the first bar in light blue shows the number of mutations called without size selection, the second bar quantifies the number of mutations called after the addition of those identified with in silico size selection, and the third bar in dark blue shows the number of mutations called after addition of mutations called after in vitro size selection. (**C**) Patients ($n = 16$) were retrospectively selected from a cohort with different cancer types (colorectal, cholangiocarcinoma, pancreatic, and prostate) enrolled in early-phase clinical trials. Matched tumor tissue DNA was available for each plasma sample, and two patients also had a biopsy collected at relapse. WES was performed on tumor tissue DNA and plasma DNA samples, and in silico size selection was applied to the data. A total of 97% (2061 of 2133) of the shared mutations detected by WES showed higher MAF after in silico size selection. (**D**) Mutations detected only after in silico selection of WES data from 16 patients [as in (C)] compared to mutations called by WES of the matched tumor tissue. Three of 16 patients had no additional mutations identified after in silico size selection. Of the 82 mutations detected in plasma after in silico size selection, 23 (28%) had low signal in tumor WES data and were not identified in those samples without size selection.

of ctDNA. We based the fragment size selection criteria on a biological observation that ctDNA fragment size distribution is shifted from noncancerous cfDNA. Our work builds on a comprehensive survey of plasma cfDNA fragmentation patterns across 200 patients with multiple cancer types and 65 healthy individuals. We identified features that could determine the presence and amount of ctDNA in plasma samples, without a priori knowledge of somatic aberrations. We caution that this catalog is limited to double-stranded DNA from plasma samples and is subject to potential biases incurred by the DNA extraction and sequencing methods we used. Additional biological effects could contribute to further selective analysis of cfDNA. Other bodily fluids (urine, cerebrospinal fluid, and saliva), different nucleic acids and structures, altered mechanisms of

release into circulation, or sample processing methods could exhibit varying fragment size signatures and could offer additional exploitable biological patterns for selective sequencing.

Previous work has reported the size distributions of mutant ctDNA but only considered limited genomic loci, cancer types, or cases (*30, 32, 33*). We identified the size differences between mutant and nonmutant DNA on a genome-wide and pan-cancer scale. We developed a method to size mutant ctDNA without using high-depth WGS. By sequencing >150 mutations per patient at high depth, we obtained large numbers of reads that could be unequivocally identified as tumor derived and thus determined the size distribution of mutant ctDNA and nonmutant cfDNA in patients with cancer. A potential limitation of our approach is that capture-based sequencing is biased by probe capture efficiency and, therefore, our data may not accurately reflect ctDNA fragments of <100 or >300 bp.

Our work provides strong evidence that the modal size of ctDNA for many cancer types is less than 167 bp, which is the length of DNA wrapped around the chromatosome. In addition, our work also shows that there is enrichment of mutant DNA fragments at sizes greater than 167 bp, notably in the range of 250 to 320 bp. These longer fragments may explain previous observations that longer ctDNA can be detected in the plasma of patients with cancer (*29, 32*). The origin of these long fragments is still unknown, and their observation could be linked to technical factors. However, it is likely that mechanisms of compaction and release of cfDNA into circulation, which may differ depending on its origin, will be reflected by different fragment sizes (*38*). Improving the charac-

terization of these fragments will be important, especially for future work combining analysis of ctDNA with that of other entities in blood such as microvesicles and tumor-educated platelets (*39, 40*). Fragment-specific analyses not only increase the sensitivity for detection of rare mutations but could also be used to track modifications in the size distribution of ctDNA. Future work should address whether this approach could be used to elucidate mechanistic effects of treatment on tumor cells, for example, by distinguishing between necrosis and apoptosis based on fragment size (*41*).

Genome-wide and exome sequencing of plasma DNA at multiple time points during cancer treatment have been proposed as noninvasive means to study cancer evolution and for the identification of possible mechanisms of resistance to treatment (*3*). However, WGS and WES
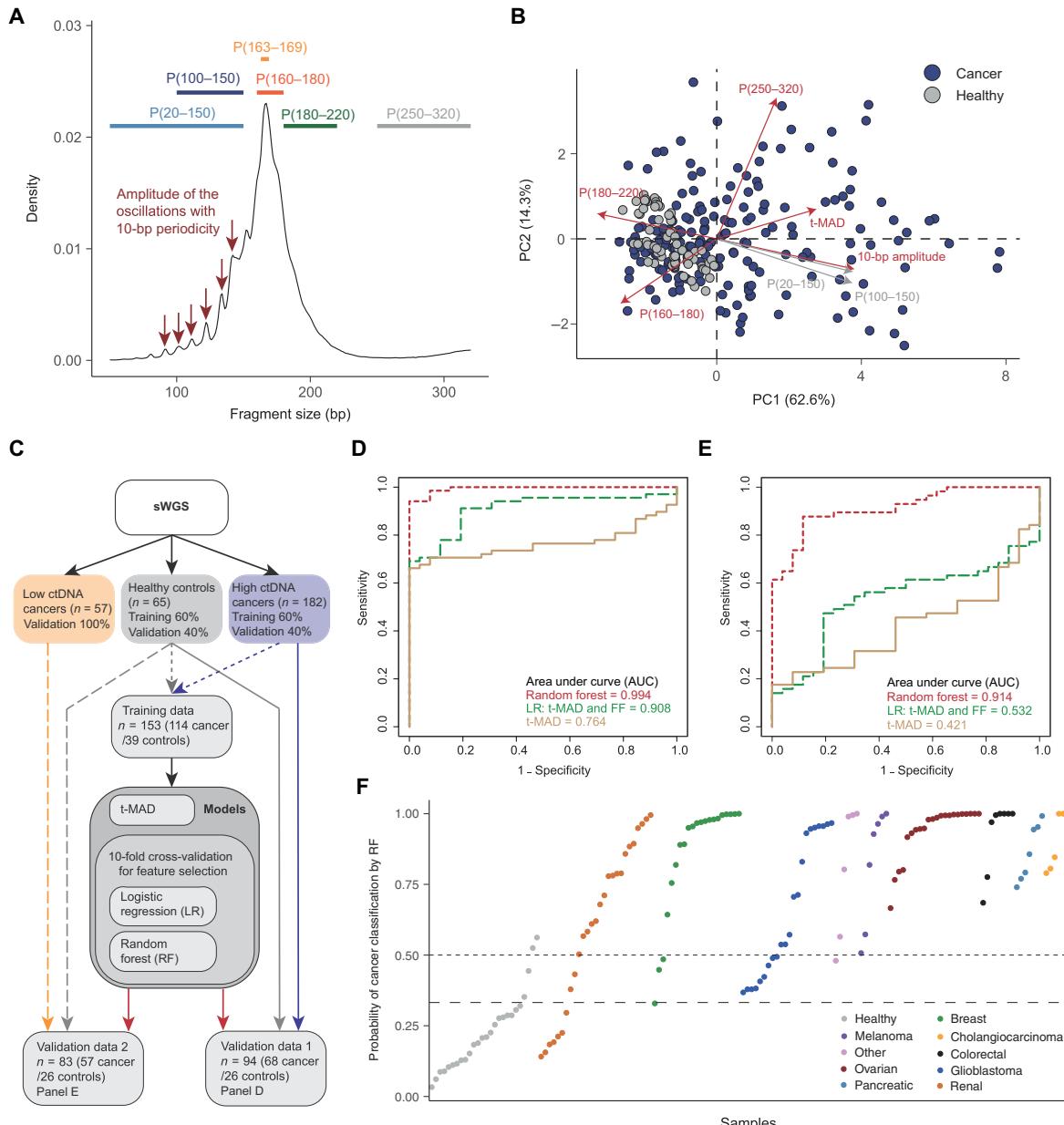
Fig. 7. Enhancing the potential for ctDNA detection by combining SCNAs and fragment size features. (**A**) Schematic illustrating the selection of different size ranges and features in the distribution of fragment sizes. For each sample, fragmentation features included the proportion (P) of fragments in specific size ranges, the ratio between certain ranges, and a quantification of the amplitude of the 10-bp oscillations in the 90- to 145-bp size range calculated from the periodic "peaks" and "valleys." (**B**) PCA comparing cancer and healthy samples using data from t-MAD scores and the fragmentation features. Red arrows indicate features that were selected as informative by the predictive analysis. (**C**) Workflow for the predictive analysis combining SCNAs and fragment size features. sWGS data from 182 plasma samples from patients with cancer types with high amounts of ctDNA (colorectal, cholangiocarcinoma, lung, ovarian, and breast) were split into a training set (60% of samples) and a validation set (validation data 1, together with the healthy individual validation set). A further dataset of sWGS from 57 samples of cancer types exhibiting low amounts of ctDNA (glioma, renal, and pancreatic) was used as validation data 2, together with the healthy individual validation set. Plasma DNA sWGS data from healthy controls were split into a training set (60% of samples) and a validation set (used in both validation data 1 and validation data 2). (**D**) ROC curves for validation data 1 (samples from patients with cancer with high ctDNA amounts, 68; healthy, 26) for three predictive models built on the pan-cancer training cohort (cancer, 114; healthy, 39). The beige curve represents the ROC curve for classification with t-MAD only, the long-dashed green line represents the LR model combining the top five features based on recursive feature elimination [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed line shows the result for a RF classifier trained on the combination of the same five features, independently chosen for the best RF predictive model. FF, fragment size features. (**E**) ROC curves for validation data 2 (samples from patients with cancer with low ctDNA amounts, 57; healthy, 26) for the same three classifiers as in (D). The beige curve represents the model using t-MAD only, the long-dashed green curve represents the LR model combining the top five features [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed curve shows the result for a RF classifier trained on the combination of same five predictive features. (**F**) Plot representing the probability of classification as cancer with the RF model for all samples in both validation datasets. Samples are separated by cancer type and sorted within each by the RF probability of classification as cancer. The horizontal dashed line indicates 50% probability (achieving specificity of 24 of 26, 92.3%), and the long-dashed line indicates 33% probability (achieving specificity of 22 of 26, 84.6%).

approaches are costly and have thus far been applicable only in samples for which the tumor DNA fraction was >5 to 10% (*3–5, 42*). We demonstrated that we could exploit the differences in fragment lengths using in vitro and in silico size selection to enrich for tumor content in plasma samples, which improved mutation and SCNA detection in sWGS and WES data. We demonstrated that size selection improved the detection of mutations that are present in plasma at low allelic fractions while maintaining low sequencing depth by sWGS and WES. Size selection can be achieved with simple means and at low cost and is compatible with a wide range of downstream genome-wide and targeted genomic analyses, greatly increasing the potential value and utility of liquid biopsies as well as the cost-effectiveness of cfDNA sequencing.

Size selection can be applied in silico, which incurs no added costs, or in vitro, which adds a simple and low-cost intermediate step that can be applied to either the extracted DNA or the libraries created from it. This approach, applied prospectively to new studies, could boost the clinical utility of ctDNA detection and analysis and creates an opportunity for reanalysis of large volumes of existing data (*4, 34, 43*). The limitation of this technique is a potential loss of material and information, because some of the informative fragments may be found in size ranges that are filtered out or deprioritized in the analysis. This may be particularly problematic if only a few copies of the fragments of interest are present in the plasma. Despite potential loss of material, we demonstrated that classification algorithms can learn from cfDNA fragmentation features and SCNA analysis and improve the detection of ctDNA with a cheap sequencing approach. Moreover, the cfDNA fragmentation features alone can be leveraged to classify cancer and healthy samples with a high accuracy [AUC, 0.989 (high ctDNA cancers) and 0.891 (low ctDNA cancers)].

Analysis of fragment sizes could provide improvements in other applications. Introducing fragment size information on each read could enhance mutation-calling algorithms from high-depth sequencing to distinguish tumor-derived mutations from other sources such as somatic variants or background sequencing noise. In addition, cfDNA from patients analyzed with CHIP is likely to be structurally different from ctDNA released during tumor cell proliferation (*18, 19*). Thus, fragmentation analysis or selective sequencing strategies could be applied to distinguish clinically relevant tumor mutations from those present in clonal expansions of normal cells. This will be critical for the development of cfDNA-based methods for identification of patients with early-stage cancer.

Size selection could also have an impact on the detection of other types of DNA in body fluids or enrichment of signals from circulating bacterial or pathogen DNA and mitochondrial DNA. These DNA fragments are not associated with nucleosomes and are often highly fragmented below 100 bp. Filtering or selection of such fragments may prove to be important in light of the recently established link between the microbiome and treatment efficiency (*17, 44*). Moreover, recent work highlights a stronger correlation of ctDNA detection with cellular proliferation than with cell death (*45*). We hypothesize that the mode of the distribution of ctDNA fragment sizes at 145 bp could reflect cfDNA released during cell proliferation, and the fragments at 167 bp may reflect cfDNA released by apoptosis or maturation/turnover of blood cells. The effect of other cancer hallmarks (*46*) on ctDNA biology, structure, concentration, and release is yet unknown.

In summary, ctDNA fragment size analysis, via size selection and machine learning approaches, boosts noninvasive genomic analysis

of tumor DNA. Size selection of shorter plasma DNA fragments enriches ctDNA and assists in the identification of a greater number of genomic alterations with both targeted and untargeted sequencing at minimal additional cost. Combining cfDNA fragment size analysis and the detection of SCNAs with a nonlinear classification algorithm improved the discrimination between samples from patients with cancer and those from healthy individuals. Because the analysis of fragment sizes is based on the structural properties of ctDNA, size selection could be used with any downstream sequencing applications. Our work could help overcome current limitations of sensitivity for liquid biopsy, supporting expanded clinical and research applications. Our results indicate that exploiting the endogenous biological properties of cfDNA provides an alternative paradigm to deeper sequencing of ctDNA.

MATERIALS AND METHODS
Study design
Three hundred forty-four plasma samples from 200 patients with multiple cancer types were collected along with plasma from 65 healthy controls. Among the patients, 172 individuals, and notably the OV04 samples, were recruited through prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference number: 07/Q0106/63; and National Research Ethics Service Committee East of England–Cambridge Central 03/018). Written informed consent was obtained from all patients, and blood samples were collected before and after initiation of treatment with surgery or chemotherapeutic agents. DNA was extracted from 2 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (QIAGEN) or QIAsymphony (QIAGEN) according to the manufacturer's instructions. In addition, 28 patients were recruited as part of the Copenhagen Prospective Personalized Oncology (CoPPO) program (PMID reference number: 25046202) at Rigshospitalet, Copenhagen, Denmark, approved by the local research ethics committee. Baseline tumor tissue biopsies were available from all 28 patients, together with rebiopsies collected at relapse from two patients, and matched plasma samples. Brain tumor patients were recruited at Addenbrooke's Hospital, Cambridge, UK as part of the BLING (bopsies of liquids in new gliomas) study (REC reference number: 15/EE/0094). Patients with bladder cancer were recruited at the Netherlands Cancer Institute, Amsterdam, The Netherlands, and approval according to national guidelines was obtained (N13KCM/CFMPB250) (*47*). Sixty-five plasma samples were obtained from healthy control individuals using a similar collection protocol (Seralab). Plasma samples have not been freeze thawed more than two times to reduce artifactual fragmentation of cfDNA. A flowchart of the study is presented in fig. S1.

SUPPLEMENTARY MATERIALS
www.sciencetranslationalmedicine.org/cgi/content/full/10/466/eaat4921/DC1
Materials and Methods
Fig. S1. Flowchart summarizing the experiments performed in this study and the sample numbers used at each step.
Fig. S2. Size distribution of cfDNA determined by sWGS for different cancer types.
Fig. S3. Insert size distribution of mutant cfDNA determined with hybrid-capture sequencing for 19 patients.
Fig. S4. DNA fragment size distribution for plasma samples from patients with ovarian cancer.
Fig. S5. Quality control assessed for in vitro size selection.
Fig. S6. Quality control assessed for in vitro and in silico size selection on healthy control samples.
Fig. S7. SCNA analysis of the segmental log$_2$ ratio determined after sWGS (<0.4× coverage) for the patient OV04-83.

REFERENCES AND NOTES

1. G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).

2. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).

3. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108–112 (2013).

4. V. A. Adalsteinsson, G. Ha, S. S. Freeman, A. D. Choudhury, D. G. Stover, H. A. Parsons, G. Gydush, S. C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. M. Francis, C.-Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M. S. Merrill, R. A. Santiago, E. P. O'Connor, S. H. Jeong, R. Leeson, R. M. Barry, J. F. Kramkowski, Z. Zhang, L. Polacek, J. G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N. M. Oliver, L. Marini, A. G. Waks, L. C. Harshman, S. M. Tolaney, E. M. Van Allen, E. P. Winer, N. U. Lin, M. Nakabayashi, M.-E. Taplin, C. M. Johannessen, L. A. Garraway, T. R. Golub, J. S. Boehm, N. Wagle, G. Getz, J. C. Love, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).

5. E. Heitzer, P. Ulz, J. Belic, S. Gutschi, F. Quehenberger, K. Fischereder, T. Benezeder, M. Auer, C. Pischler, S. Mannweiler, M. Pichler, F. Eisner, M. Haeusler, S. Riethdorf, K. Pantel, H. Samonigg, G. Hoefler, H. Augustin, J. B. Geigl, M. R. Speicher, Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **5**, 30 (2013).

6. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, l.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).

7. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz Jr., S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16368–16373 (2005).

8. S.-J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).

9. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz, Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).

10. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8**, 346ra92 (2016).

11. A. A. Chaudhuri, J. J. Chabon, A. F. Lovejoy, A. M. Newman, H. Stehr, T. D. Azad, M. S. Khodadoust, M. S. Esfahani, C. L. Liu, L. Zhou, F. Scherer, D. M. Kurtz, C. Say, J. N. Carter, D. J. Merriott, J. C. Dudley, M. S. Binkley, L. Modlin, S. K. Padda, M. F. Gensheimer, R. B. West, J. B. Shrager, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., A. A. Alizadeh, M. Diehn, Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov.* **7**, 1394–1403 (2017).

12. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, C. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).

13. I. S. Haque, O. Elemento, Challenges in using ctDNA to achieve early detection of cancer. bioRxiv 237578 [preprint]. 21 December 2017. https://doi.org/10.1101/237578

14. A. M. Newman, A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge Jr., J. B. Shrager, B. W. Loo Jr., J. W. Neal, H. A. Wakelee, M. Diehn, A. A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).

15. P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G. Pristauz, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48**, 1273–1278 (2016).

16. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).

17. P. Burnham, M. S. Kim, S. Agbor-Enoh, H. Luikart, H. A. Valantine, K. K. Khush, I. De Vlaminck, Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).

18. G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landén, M. Höglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Grönberg, C. M. Hultman, S. A. McCarroll, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

19. Y. Hu, B. Ulrich, J. Supplee, Y. Kuang, P. H. Lizotte, N. B. Feeney, N. M. Guibert, M. M. Awad, K.-K. Wong, P. A. Jänne, C. P. Paweletz, G. R. Oxnard, False-positive plasma genotyping due to clonal hematopoiesis. *Clin. Cancer Res.* **24**, 4437–4443 (2018).

20. A. J. Bronkhorst, J. F. Wentzel, J. Aucamp, E. van Dyk, L. du Plessis, P. J. Pretorius, Characterization of the cell-free DNA released by cultured cancer cells, *Biochim. Biophys. Acta* **1863**, 157–165 (2016).

21. S. Jahr, H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R.-D. Hesch, R. Knippers, DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).

22. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).

23. D. Chandrananda, N. P. Thorne, M. Bahlo, High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genomics* **8**, 29 (2015).

24. P. Jiang, Y. M. D. Lo, The long and short of circulating cell-free DNA and the ins and outs of molecular diagnostics. *Trends Genet.* **32**, 360–371 (2016).

25. S. C. Y. Yu, K. C. A. Chan, Y. W. L. Zheng, P. Jiang, G. J. W. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. J. I. Go, J. M. G. van Vugt, R. Minekawa, C. B. M. Oudejans, K. H. Nicolaides, R. W. K. Chiu, Y. M. D. Lo, Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8583–8588 (2014).

26. F. M. F. Lun, N. B. Y. Tsui, K. C. A. Chan, T. Y. Leung, T. K. Lau, P. Charoenkwan, K. C. K. Chow, W. Y. W. Lo, C. Wanapirak, T. Sanguansermsri, C. R. Cantor, R. W. K. Chiu, Y. M. D. Lo, Noninvasive prenatal diagnosis of monogenic diseases by digital size

selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19920–19925 (2008).

27. G. Minarik, G. Repiska, M. Hyblova, E. Nagyova, K. Soltys, J. Budis, F. Duris, R. Sysak, M. Gerykova Bujalkova, B. Vlkova-Izrael, O. Biro, B. Nagy, T. Szemes, Utilization of benchtop next generation sequencing platforms ion torrent PGM and MiSeq in noninvasive prenatal testing for chromosome 21 trisomy and testing of impact of in silico and physical size selection on its analytical performance. *PLOS ONE* **10**, e0144811 (2015).

28. M. B. Giacona, G. C. Ruben, K. A. Iczkowski, T. B. Roos, D. M. Porter, G. D. Sorenson, Cell-free DNA in human blood plasma: Length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* **17**, 89–97 (1998).

29. N. Umetani, A. E. Giuliano, S. H. Hiramatsu, F. Amersi, T. Nakagawa, S. Martino, D. S. B. Hoon, Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J. Clin. Oncol.* **24**, 4270–4276 (2006).

30. F. Mouliere, B. Robert, E. Arnau Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, High fragmentation characterizes tumour-derived circulating DNA. *PLOS ONE* **6**, e23418 (2011).

31. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).

32. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1317–E1325 (2015).

33. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, Fragment length of circulating tumor DNA. *PLOS Genet.* **12**, e1006162 (2016).

34. O. A. Zill, K. C. Banks, S. R. Fairclough, S. A. Mortimer, J. V. Vowles, R. Mokhtari, D. R. Gandara, P. C. Mack, J. I. Odegaard, R. J. Nagy, A. M. Baca, H. Eltoukhy, D. I. Chudova, R. B. Lanman, A. Talasaz, The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin. Cancer Res.* **24**, 3528–3538 (2018).

35. G. Macintyre, T. E. Goranova, D. De Silva, D. Ennis, A. M. Piskorz, M. Eldridge, D. Sie, L.-A. Lewsley, A. Hanif, C. Wilson, S. Dowson, R. M. Glasspool, M. Lockley, E. Brockbank, A. Montes, A. Walther, S. Sundar, R. Edmondson, G. D. Hall, A. Clamp, C. Gourley, M. Hall, C. Fotopoulou, H. Gabra, J. Paul, A. Supernat, D. Millan, A. Hoyle, G. Bryson, C. Nourse, L. Mincarelli, L. N. Sanchez, B. Ylstra, M. Jimenez-Linan, L. Moore, O. Hofmann, F. Markowetz, I. A. McNeish, J. D. Brenton, Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).

36. C. A. Parkinson, D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld, J. D. Brenton, Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: A retrospective study. *PLOS Med.* **13**, e1002198 (2016).

37. T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68 (2012).

38. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35**, 347–376 (2016).

39. M. G. Best, N. Sol, B. A. Tannous, P. Wesseling, T. Wurdinger, RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).

40. M. G. Best, N. Sol, S. G. J. G. In 't Veld, A. Vancura, M. Muller, A.-L. N. Niemeijer, A. V Fejes, L.-A. Tjon Kon Fat, A. E. Huis In 't Veld, C. Leurs, T. Y. Le Large, L. L. Meijer, I. E. Kooi, F. Rustenburg, P. Schellen, H. Verschueren, E. Post, L. E. Wedekind, J. Bracht, M. Esenkbrink, L. Wils, F. Favaro, J. D. Schoonhoven, J. Tannous, H. Meijers-Heijboer, G. Kazemier, E. Giovannetti, J. C. Reijneveld, S. Idema, J. Killestein, M. Heger, S. C. de Jager, R. T. Urbanus, I. E. Hoefer, G. Pasterkamp, C. Mannhalter, J. Gomez-Arroyo, H.-J. Bogaard, D. P. Noske, W. P. Vandertop, D. van den Broek, B. Ylstra, R. J. A. Nilsson, P. Wesseling, N. Karachaliou, R. Rosell, E. Lee-Lewandrowski, K. B. Lewandrowski, B. A. Tannous, A. J. de Langen, E. F. Smit, M. M. van den Heuvel, T. Wurdinger, Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* **32**, 238–252.e9 (2017).

41. A. L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmann-Groth, M. Schneider, T. Muley, M. Thomas, H. Sültmann, Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients. *Sci. Rep.* **6**, 33505 (2016).

42. J. Belic, M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher, E. Heitzer, Rapid identification of plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. *Clin. Chem.* **61**, 838–849 (2015).

43. D. G. Stover, H. A. Parsons, G. Ha, S. S. Freeman, W. T. Barry, H. Guo, A. D. Choudhury, G. Gydush, S. C. Reed, J. Rhoades, D. Rotem, M. E. Hughes, D. A. Dillon, A. H. Partridge, N. Wagle, I. E. Krop, G. Getz, T. R. Golub, J. C. Love, E. P. Winer, S. M. Tolaney, N. U. Lin, V. A. Adalsteinsson, Association of cell-free DNA tumor fraction and somatic copy number alterations with survival in metastatic triple-negative breast cancer. *J. Clin. Oncol.* **36**, 543–553 (2018).

44. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J. R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).

45. C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann; TRACERx consortium; PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).

46. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

47. K. M. Patel, K. E. van der Vos, C. G. Smith, F. Mouliere, D. Tsui, J. Morris, D. Chandrananda, F. Marass, D. van den Broek, D. E. Neal, V. J. Gnanapragasam, T. Forshew, B. W. van Rhijn, C. E. Massie, N. Rosenfeld, M. S. van der Heijden, Association of plasma and urinary mutant DNA with clinical outcomes in muscle invasive bladder cancer. *Sci. Rep.* **7**, 5554 (2017).

48. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

49. I. Scheinin, D. Sie, H. Bengtsson, M. A. van de Wiel, A. B. Olshen, H. F. van Thuijl, H. F. van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson, B. Ylstra, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).

Citation: F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

🔔 Get alerts about upcoming free webinars.

Sign up at: webinar.sciencemag.org/stayinformed

Transforming the **Future** of **Robotics**

As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

Learn more at: **ScienceRobotics.org**

Science Robotics

AAAS
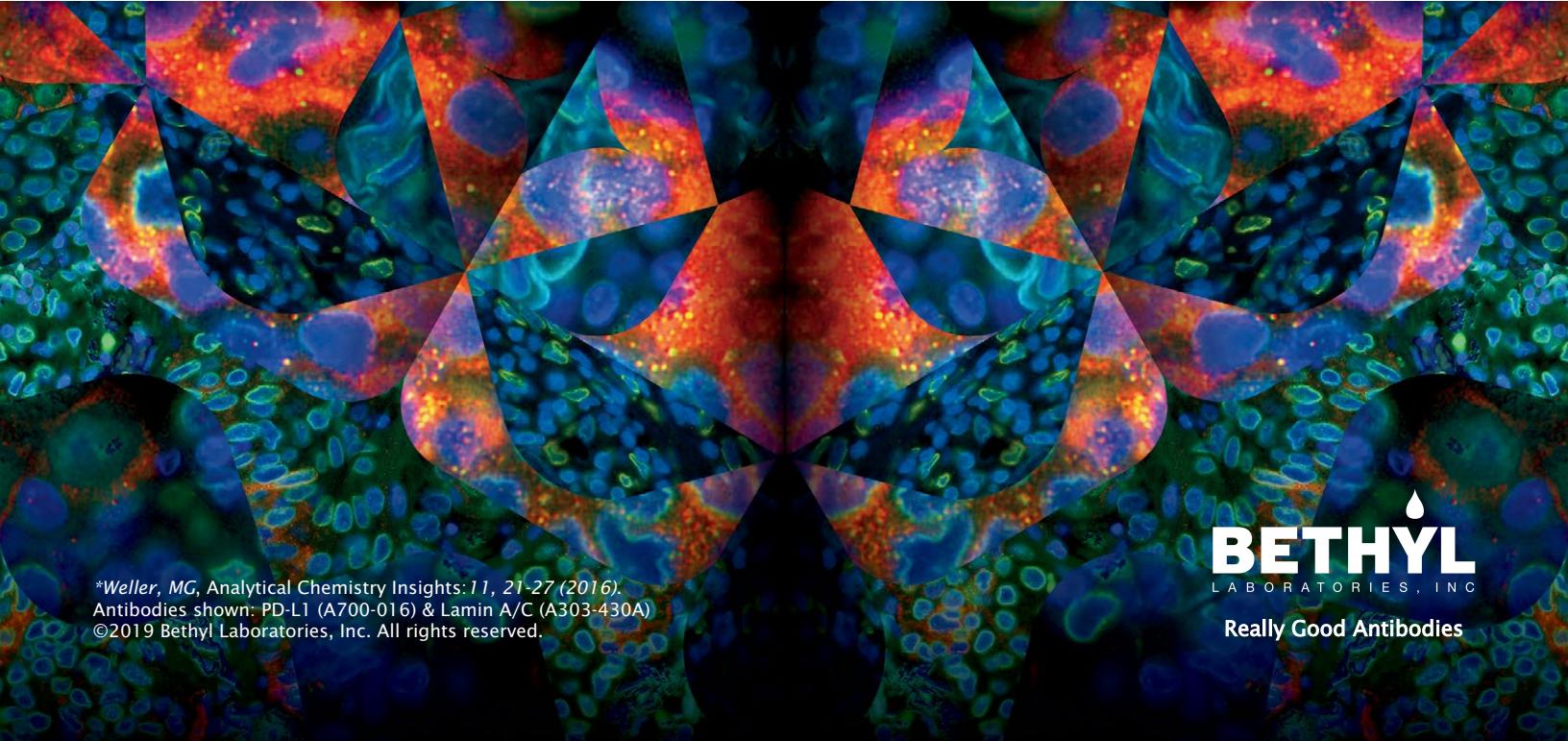
Brilliant visualization

See the difference to make big breakthroughs in your tumor microenvironment research

In the world of next generation immuno-oncology research, having confidence in your immunoassay results is vital. Unfortunately, 75% of antibodies in today's market are non-specific or simply do not work at all.* That's why at Bethyl, we manufacture and validate every antibody on-site to ensure target specificity and sensitivity. More than 10,000 independent citations over the past 15 years have proven that our antibodies will function as designed in your assay — and we offer a 100% guarantee. Work with Bethyl to bring your discovery into full focus.

See our data at bethyl.com/immuno-oncology

BETHYL
LABORATORIES, INC

Really Good Antibodies