ARTICLE

https://doi.org/10.1038/s41586-019-1103-9

Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens

Fiona M. Behan[1,2,12], Francesco Iorio[1,2,3,12], Gabriele Picco[1,12], Emanuel Gonçalves[1], Charlotte M. Beaver[1], Giorgia Migliardi[4,5], Rita Santos[6], Yanhua Rao[7], Francesco Sassi[4], Marika Pinnelli[4,5], Rizwan Ansari[1], Sarah Harper[1], David Adam Jackson[1], Rebecca McRae[1], Rachel Pooley[1], Piers Wilkinson[1], Dieudonne van der Meer[1], David Dow[2,6], Carolyn Buser-Doepner[2,7], Andrea Bertotti[4,5], Livio Trusolino[4,5], Euan A. Stronach[2,6], Julio Saez-Rodriguez[2,3,8,9,10], Kosuke Yusa[1,2,11,13]* & Mathew J. Garnett[1,2,13]*

Functional genomics approaches can overcome limitations—such as the lack of identification of robust targets and poor clinical efficacy—that hamper cancer drug development. Here we performed genome-scale CRISPR–Cas9 screens in 324 human cancer cell lines from 30 cancer types and developed a data-driven framework to prioritize candidates for cancer therapeutics. We integrated cell fitness effects with genomic biomarkers and target tractability for drug development to systematically prioritize new targets in defined tissues and genotypes. We verified one of our most promising dependencies, the Werner syndrome ATP-dependent helicase, as a synthetic lethal target in tumours from multiple cancer types with microsatellite instability. Our analysis provides a resource of cancer dependencies, generates a framework to prioritize cancer drug targets and suggests specific new targets. The principles described in this study can inform the initial stages of drug development by contributing to a new, diverse and more effective portfolio of cancer drug targets.

The molecular features of a patient's tumour influence clinical responses and can be used to guide therapy, leading to more effective treatments and reduced toxicity[1]. However, most patients do not benefit from such targeted therapies in part owing to a limited knowledge of candidate targets[2]. Lack of efficacy is a leading cause of the 90% attrition rate in the development of cancer drugs, and fewer molecular entities to new targets are being developed[3]. Unbiased strategies that effectively identify and prioritize targets in tumours could expand the range of targets, improve success rates and accelerate the development of new cancer therapies.

CRISPR–Cas9 screens that use libraries of single-guide RNAs (sgRNAs) have been used to study gene function and their role in cellular fitness[4,5]. CRISPR–Cas9-based genome editing provides high specificity and produces penetrant phenotypes as null alleles can be generated. Here we present genome-scale CRISPR–Cas9 fitness screens in 324 cancer cell lines and an integrative analysis that enables the prioritization of candidate cancer therapeutic targets (Fig. 1a), which we illustrate through the identification of Werner syndrome ATP-dependent helicase (WRN) as a target for tumours with microsatellite instability (MSI).

Genome-scale CRISPR–Cas9 screens in cancer cell lines

To comprehensively catalogue genes that are required for cancer cell fitness (defined as genes required for cell growth or viability), we performed 941 CRISPR–Cas9 fitness screens in 339 cancer cell lines, targeting 18,009 genes (Extended Data Fig. 1a, b and Supplementary Table 1). Following stringent quality control (Extended Data Fig. 1c–h), the final analysis set included 324 cell lines from 30 different cancer types, across 19 different tissues (Extended Data Fig. 1i). These cell lines are part of the collection of Cell Model Passports of highly genomically annotated cell lines[6], broadly represent the molecular features of tumours in patients[7], and include common cancers (such as lung, colon and breast cancers) and cancers of particular unmet clinical need (such as lung and pancreatic cancers). Analysis of screen data from these 324 cell lines demonstrated high sensitivity, specificity and precision in classifying essential and non-essential genes[8] (Extended Data Fig. 1g, h, j), and results were not biased by experimental factors (Extended Data Fig. 2a–e).

Defining core and context-specific fitness genes

Genes required for cell fitness in specific molecular or histological contexts are likely to encode favourable drug targets, because of a reduced likelihood of inducing toxic effects in healthy tissues. Conversely, fitness genes that are common to the majority of tested cell lines or common within a cancer type (referred to as pan-cancer or cancer-type-specific core fitness genes, respectively) may be involved in essential processes in cells and have greater toxicity. It is therefore important to distinguish context-specific fitness genes from core fitness genes.

We identified a median of 1,459 fitness genes in each cell line (Extended Data Fig. 2f–n and Supplementary Table 2). In total, 41% ($n = 7{,}470$) of all targeted genes induced a fitness effect in one or more cell lines and the majority (83%) of these genes induced a dependency in less than 50% of the tested cell lines (Fig. 1b). To identify core fitness genes, we developed a statistical method, the adaptive daisy model (ADaM; Extended Data Fig. 3a–d), to adaptively determine the minimum number of dependent cell lines that are required for a gene to be classified as a core fitness gene (Fig. 1c). Genes that were defined as core fitness in at least 12 out of 13 cancer types (also adaptively determined) were classified as pan-cancer core fitness genes (Extended Data Fig. 3e–g). This yielded a median of 866 cancer-type-specific and 553 pan-cancer core fitness genes (Fig. 1c and Supplementary Table 3).

[1]Wellcome Sanger Institute, Cambridge, UK. [2]Open Targets, Cambridge, UK. [3]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK. [4]Candiolo Cancer Institute-FPO, IRCCS, Turin, Italy. [5]Department of Oncology, University of Torino, Turin, Italy. [6]GlaxoSmithKline Research and Development, Stevenage, UK. [7]GlaxoSmithKline Research and Development, Collegeville, PA, USA. [8]Faculty of Medicine, Joint Research Centre for Computational Biomedicine, RWTH Aachen University, Aachen, Germany. [9]Institute for Computational Biomedicine, Heidelberg University, Faculty of Medicine, Bioquant, Heidelberg, Germany. [10]Heidelberg University Hospital, Heidelberg, Germany. [11]Present address: Stem Cell Genetics, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan. [12]These authors contributed equally: Fiona M. Behan, Francesco Iorio, Gabriele Picco. [13]These authors jointly supervised this work: Kosuke Yusa, Mathew J. Garnett. *e-mail: k.yusa@infront.kyoto-u.ac.jp; mathew.garnett@sanger.ac.uk

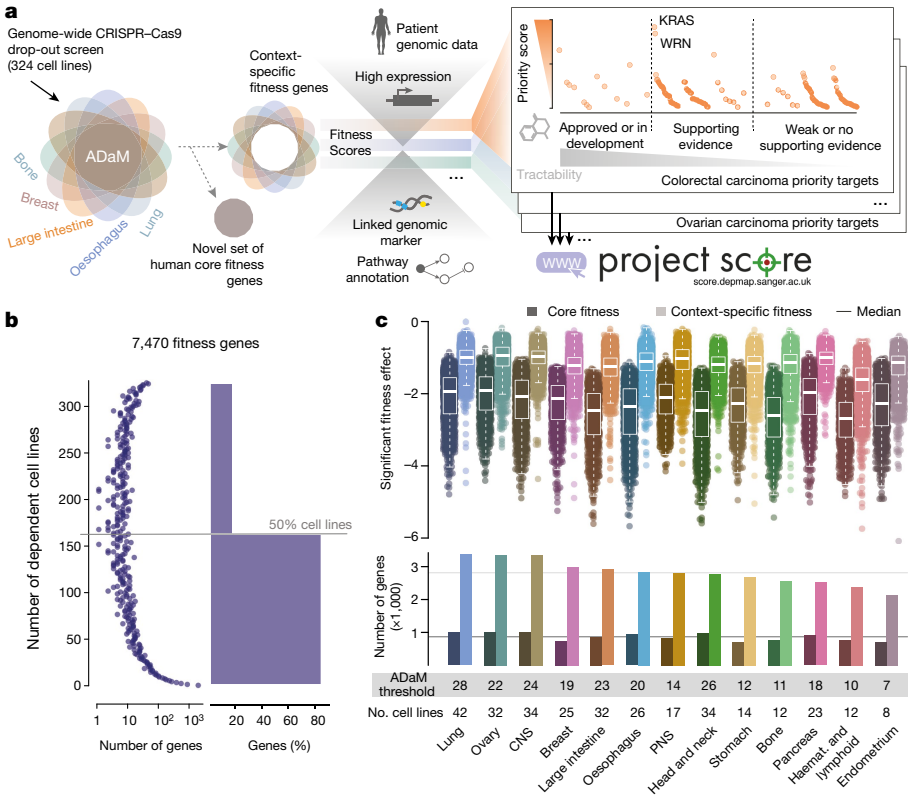


Fig. 1 | Target prioritization framework. a, Strategy to prioritize targets in multiple cancer types, incorporating CRISPR–Cas9 gene fitness effects, genomic biomarkers and target tractability for drug development. ADaM (adaptive daisy model) distinguishes context-specific and core fitness genes. Datasets are available on the project Score website (https://score.depmap.sanger.ac.uk/). **b**, Number of genes exerting a fitness defect in a given number of cell lines. The bars indicate the percentage of genes that induce a dependency in less than (bottom bar) or at least (top bar) 50% of cell lines. **c**, Bottom, number of core and context-specific fitness genes identified by ADaM for 13 cancer types (median = 866 and 2,813, respectively). The ADaM threshold is the number of cell lines for a gene to be classified as core fitness. Top, comparison of the effect size for ADaM core and context-specific fitness genes (only significant genes are shown, BAGEL FDR = 5%). CNS, central nervous system; Haemat., haematological; PNS, peripheral nervous system.

Of the pan-cancer core fitness genes identified using ADaM, 399 were previously defined as essential genes[8,9] and 125 are genes involved in essential cellular processes[10,11] (Extended Data Fig. 4a). The remaining 132 (24%) genes were newly identified and are also significantly enriched in cellular housekeeping genes and pathways (Extended Data Fig. 4b, c and Supplementary Table 4). In comparison to previously identified reference core fitness gene sets[8,9], our pan-cancer core fitness gene set showed greater recall of genes involved in essential processes (median = 67%, versus 28% and 51% in the previously published gene sets of refs. [8] and [9], respectively, Extended Data Fig. 4d), with similar false discovery rates (FDRs) for putative context-specific fitness genes (taken from a previous study[12]; Extended Data Fig. 4e). Blood cancer cell lines had the most distinctive profile of core fitness genes (31 exclusive core fitness genes; Extended Data Fig. 4f). Cancer-type-specific core fitness genes are generally highly expressed in matched healthy tissues (Extended Data Fig. 4g), consistent with their predicted role in fundamental cellular processes and suggesting that they show potential toxicity if used as targets. Notably, five genes were core fitness in a single cancer type and were lowly or not expressed at the basal level in the matched normal tissues (Extended Data Fig. 4g), suggesting that they could induce cancer-cell-specific dependencies in these tissues.

Overall, using a statistical approach, we refined and expanded our current knowledge of core fitness genes in humans and identified genes that have a high likelihood of toxicity, which thus represent less favourable therapeutic targets. Furthermore, owning to the large scale of our dataset, we could now define context-specific fitness genes (median $n = 2{,}813$ genes per cancer type), many of which had a loss-of-fitness effect that was similar to or stronger than core fitness genes (Fig. 1c).

A quantitative framework for target prioritization

To nominate promising therapeutic targets from our list of context-specific fitness genes, we developed a computational framework that integrated multiple lines of evidence to assign each gene a target priority score—which ranged from 0 to 100—and generated ranked lists of candidates for an individual cancer type or a pan-cancer candidate (Fig. 1a and Extended Data Fig. 5a). To exclude genes that are likely to be poor targets because of potential toxicity, core fitness genes were scored as '0', as were potential false-positive genes, such as genes that were not expressed or homozygously deleted. For each gene, 70% of the priority score was derived from CRISPR–Cas9 experimental evidence and averaged across dependent cell lines on the basis of the fitness effect size, the significance of fitness deficiency, target gene expression, target mutational status and evidence for other fitness genes in the same pathway. The remaining 30% of the priority score was based on evidence of a genetic biomarker that was associated with a target dependency and the frequency at which the target was somatically altered in tumours in patients[7]. For the biomarker analysis, we performed an analysis of variance (ANOVA; Fig. 2a, Extended Data Fig. 5b and supplementary data 1) to test associations between fitness genes and the presence of 484 cancer driver events (151 single-nucleotide variants and 333 copy number variants)[7] or MSI, in each cancer type with a sufficiently large sample size ($n \geq 10$ cell lines) and pan-cancer. We derived a priority score threshold (55 and 41 for pan-cancer and cancer-type-specific analyses, respectively) based on scores calculated for targets with approved or preclinical cancer compounds (Extended Data Fig. 5c and Supplementary Table 5).

In total, we identified 628 unique priority targets, including 92 pan-cancer and 617 cancer-type-specific targets (Fig. 2b and

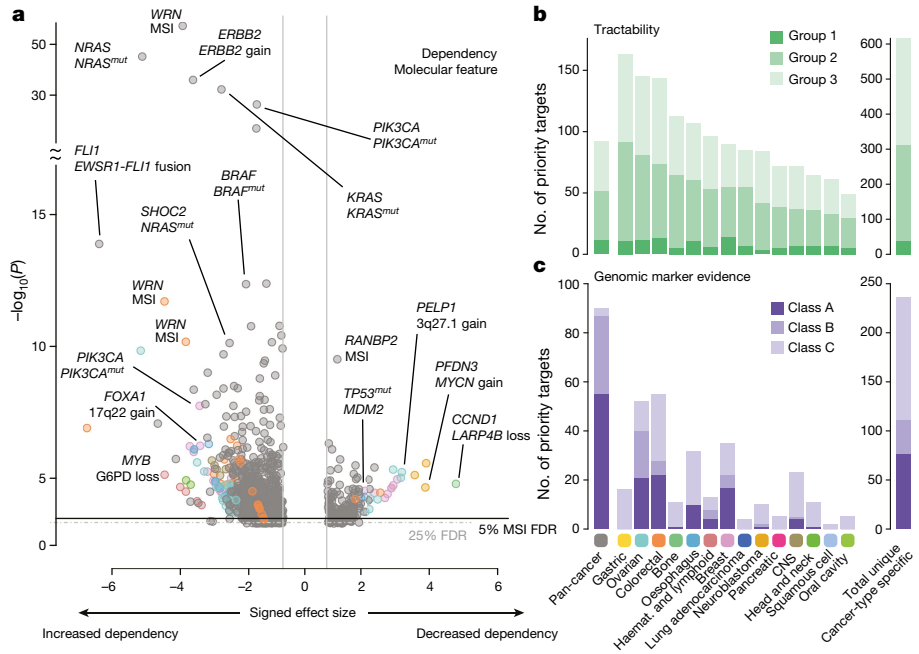


Fig. 2 | Target prioritization and biomarker discovery. a, Differential dependency biomarkers were analysed by ANOVA. Each point is an association between the fitness effect of a gene (top name) and a molecular feature or MSI (bottom name). Colours indicate results from 13 cancer-type-specific (number of cell lines indicated in Supplementary Table 1) or pan-cancer ($n = 319$) analyses. FDRs were calculated using the Storey–Tibshirani method. **b**, Cancer-type-specific and pan-cancer priority targets classified based on tractability for drug development as groups 1, 2 and 3 (strong, weak and absence of evidence, respectively). **c**, Priority targets with a genomic biomarker defined as class A, B or C (from strongest to weakest, based on statistical significance and effect size).

Supplementary Tables 6, 7). The number of priority targets varied approximately threefold across cancer types with a median of 88 targets. The majority of cancer-type priority targets ($n = 457$, 74%) were identified in only one (56%) or two (18%) cancer types, underscoring their context specificity. Most priority pan-cancer targets (88%) were also identified in the cancer-type-specific analyses (Extended Data Fig. 5d). The 11 priority targets that were identified only in the pan-cancer analysis typically included dependencies that occurred in a small subset of cell lines from multiple cancer types (for example, CREBBP and JUP) or in a cancer type for which the limited numbers of available cell lines prevented a cancer-type-specific analysis being performed (for example, SOX10 in melanoma; Extended Data Fig. 5e).

Of the 628 priority targets, 120 (19%) were associated with at least one biomarker identified using ANOVA with high significance and large effect size (defined as class A targets) and these proteins would therefore be of particular interest for drug development (Fig. 2c). For example, PIK3CA is a class A target in breast, oesophageal, colorectal and ovarian carcinoma; PI3K inhibitors are in clinical development for cancers with mutations in *PIK3CA*[13]. Using progressively less stringent significance thresholds expanded the targets with at least one biomarker association as identified by ANOVA, which were defined as class B ($n = 61$, 10%) followed by class C ($n = 117$, 19%) targets, some of which were identified in multiple cancer types (Supplementary Table 8). Taken together, these results highlight the potential of a data-driven quantitative framework to prioritize targets by combining CRISPR–Cas9 screening data from multiple cell lines and associated genomic features.

Tractability assessment of priority targets

On the basis of current drug-development strategies, targets vary in their suitability for pharmaceutical intervention and this informs target selection. Using a target tractability assessment for the development of small molecules and antibodies, we previously assigned each gene to 1 of 10 tractability buckets (with 1 indicating the highest tractability)[14]. We cross-referenced the 628 priority targets with their tractability and categorized them into three tractability groups (Fig. 2b and Supplementary Table 9).

Tractability group 1 (buckets 1–3) comprised targets of approved anticancer drugs or compounds in clinical or preclinical development, and included 40 unique priority targets, such as ERBB2, ERBB3, CDK4, AKT1, ESR1, TYMS and PIK3CB in breast carcinoma and PIK3CA, IGF1R, MTOR and ATR in colorectal carcinoma (Figs. 3a, 4 and Extended Data Fig. 6). Of these 40 priority targets, 20 have at least one drug that has been developed for the cancer type in which the target was identified as priority, whereas the remaining 20 targets have drugs that have been used or developed for treatment of other cancer types, which present opportunities for the repurposing of these drugs. A third of the priority targets in group 1 have a class A biomarker, indicating that they are highly desirable targets (Supplementary Tables 8, 9). An example is CSNK2A1, which is encoded by the highly significant fitness gene *CSNK2A1* in colorectal cancer cell lines with amplification of a chromosomal segment that contains *FLT3* and *WASF3* ($P = 6.65 \times 10^{-6}$, Glass's $\Delta > 2.9$, Fig. 3b) and targeted by silmasertib. Other priority targets in group 1 with markers show ERBB2 or ERBB3 dependency in the presence of *ERBB2* amplification, CDK2 dependency in *ASXL*-amplified oesophageal cancer cell lines, PIK3CA dependency in the presence of *PIK3CA* mutations, and PIK3CB dependency in breast cancer cell lines with *PTEN* mutations (Fig. 3b and supplementary data 1).

Tractability group 2 (buckets 4–7) contained 277 priority targets without drugs in clinical development but with evidence that support target tractability (Figs. 3a, 4, Extended Data Fig. 6 and Supplementary Table 9). Of these, 18% have a class A biomarker, including KRAS dependency in *KRAS*-mutant cell lines, USP7 dependency in *APC* wild-type colorectal cell lines, KMT2D dependency in breast cancer cell lines with amplification of a chromosomal segment that contains *PPM1D* and *CLTC*, and TRIAP1 dependency in *MYC*-amplified bone and gastric cancer cell lines (Fig. 3b and supplementary data 1). Of note, we observed a class A biomarker-type dependency on WRN in colorectal and ovarian cell lines with MSI and pan-cancer (Fig. 3b). Of the priority targets in group 2 that were not associated with a biomarker, GPX4 is a target in multiple cancer types (Fig. 4, Extended Data Fig. 6 and Supplementary Table 9). Sensitivity to GPX4 inhibition has been associated with epithelial–mesenchymal transition[15] and we

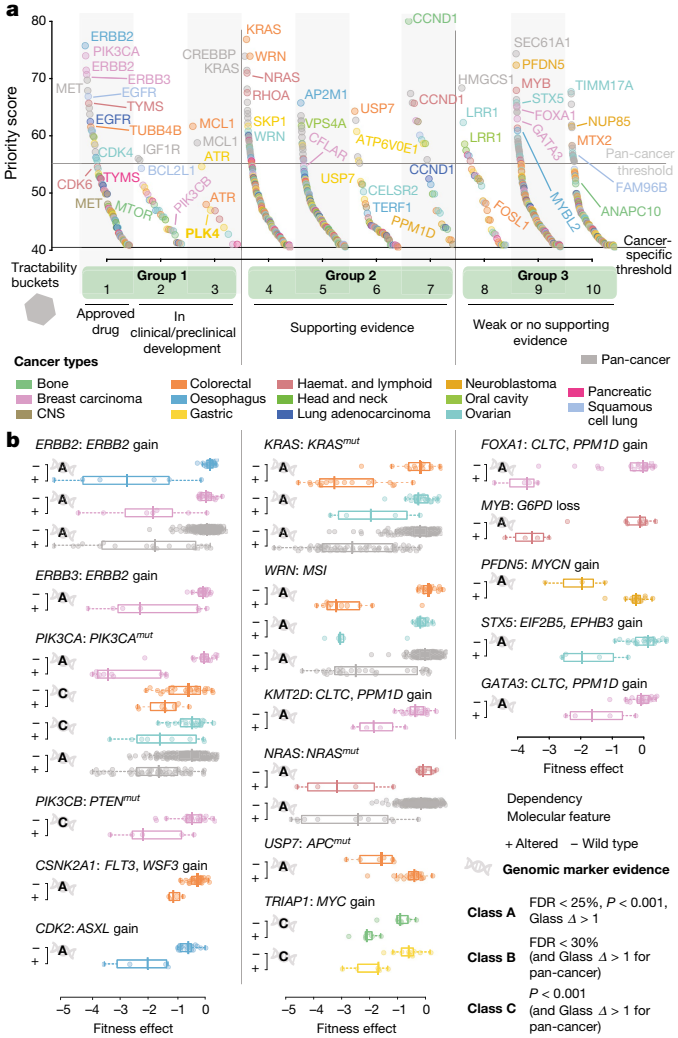


Fig. 3 | Priority targets and biomarker-linked dependencies. a, All priority targets from cancer-type and pan-cancer analyses and their tractability. Priority score thresholds are indicated and selected examples labelled. **b**, Differential fitness analysis (quantile-normalized gene depletion fold change between the average of targeting sgRNAs versus plasmid library) for selected priority targets comparing cells with (+) or without (−) a genomic marker (classes A–C as previously defined from ANOVAs). Each data point is a cell line and colours represent cancer type. Box-and-whisker plots show 1× interquartile ranges and 5–95th percentiles, centres indicate medians.

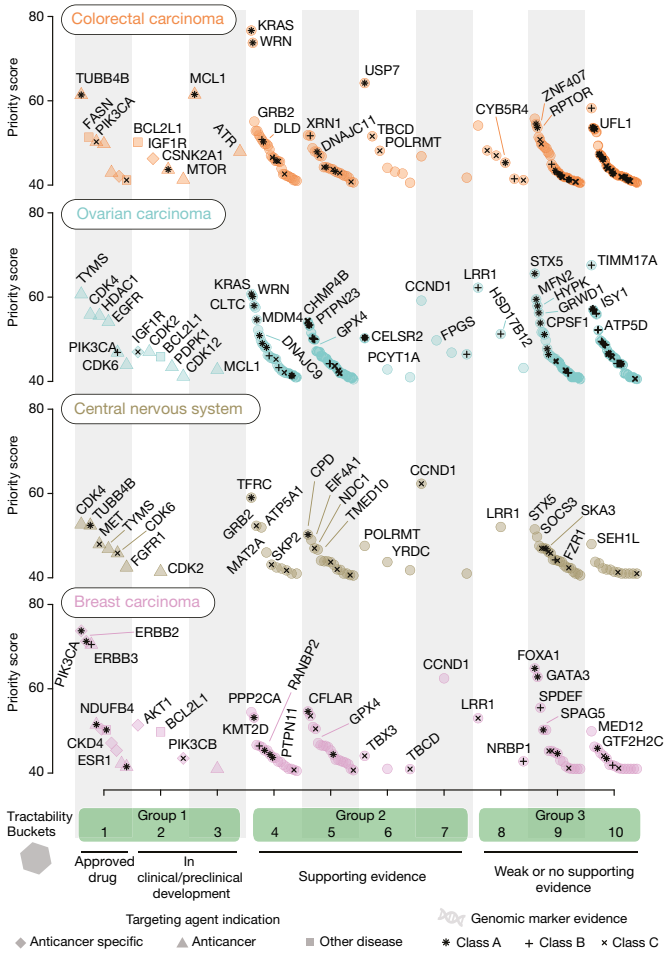


Fig. 4 | Cancer-type priority targets. Results for 4 of the 13 cancer-type-specific analyses. Points are target priority scores and the shapes indicate approved or preclinical compounds to the corresponding target (other disease (squares), anticancer targets (triangles) or those specific to the cancer type considered (rhombus)), or the absence of a compound (circles). Symbols indicate the strength of a genomic biomarker. Selected priority targets are labelled.

observed differential expression of markers associated with epithelial–mesenchymal transition in GPX4-dependent cell lines (Extended Data Fig. 7a and supplementary data 2). This is indicative of how future refinement of our target prioritization scheme can capture priority targets that are associated with an expanded set of molecular features, including gene expression, chromatin modifications and differentiation states.

Lastly, group 3 (buckets 8–10) included 311 priority targets that had no support or a lack of information that could inform tractability (Figs. 3a, 4 and Extended Data Fig. 6); this group is significantly enriched in transcription factors (Extended Data Fig. 7b and supplementary data 3). Examples of priority targets in group 3 with class A biomarkers include FOXA1 and GATA3 in breast cancer, MYB in haematological and lymphoid cancer, STX5 in ovarian cancer and PFDN5 in neuroblastoma cell lines (Fig. 3b).

Priority targets in tractability group 1 were enriched in protein kinases, highlighting a major focus of drug development against this class of targets, compared to groups 2 and 3, which included a more functionally diverse set of targets (Extended Data Fig. 7b and supplementary data 3). Targets in group 2 are most likely to be novel and tractable through conventional modalities and, therefore, represent good candidates for drug development. Newer therapeutic modalities, such as proteolysis-targeting chimaeras, may increase the range of proteins that are amenable to pharmaceutical intervention to include targets in group 3. Overall, our framework informed a data-driven list of prioritized therapeutic targets that would be strong candidates for the development of cancer drugs.

WRN is a target in cancers with MSI

To substantiate our target prioritization strategy, we investigated WRN helicase as a promising target in MSI cancers (Figs. 3, 4). WRN is one of five RecQ family DNA helicases, of which it is the only one that has both a helicase and an exonuclease domain, and has diverse roles in DNA repair, replication, transcription and telomere maintenance[16]. The MSI phenotype is caused by impaired DNA mismatch repair (MMR) due to silencing or inactivation of MMR pathway genes. MSI is associated with a high mutational load and occurs in more than 20 tumours types and is frequent in colon, ovarian, endometrial and gastric cancers (3–28%)[17].

Dependency on WRN was highly associated with MSI in the pan-cancer ANOVA, and analyses of colon and ovarian cancer cell lines (Figs. 2a, 3b and supplementary data 1). Most endometrial and gastric cancer cell lines with MSI were dependent on WRN; however, the association with MSI was not significant (for gastric) or not tested because of small sample sizes (Extended Data Fig. 7c). MSI is

rare (<1%) in many other tumour types, such as kidney, melanoma and prostate cancers[17] and most (4 out of 5 tested) MSI cell lines from these tissues were not dependent on WRN (Extended Data Fig. 7c). Other tested RecQ family members (*BLM*, *RECQL* and *RECQL5*) were not associated as fitness genes in MSI cell lines. A focused analysis of non-synonymous mutations, promoter methylation and homozygous deletions of MMR pathway genes confirmed a significant association between WRN dependency and hypermethylation of the *MLH1* promoter (Student's *t*-test, FDR = 7.72×10^{-3}) or mutations in *MSH6* (FDR = 3.85×10^{-2}); as well as mutations in the epigenetic regulator *MLL2* (also known as *KMT2D*) (FDR = 1.43×10^{-4}) (Fig. 5a).

To further validate WRN, we performed CRISPR-based co-competition assays in which the relative fitness of *WRN*-knockout versus wild-type cells was compared. *WRN* knockout using four individual sgRNAs decreased fitness of *WRN*-knockout compared to wild-type cells in six MSI cell lines from colon, ovarian, endometrial and gastric cancers (Fig. 5b, Extended Data Fig. 8a and Supplementary Table 10). By contrast, there was no difference in all microsatellite stable cell lines from these four tissues. Consistently, WRN was selectively essential for MSI cells in clonogenic assays (Extended Data Fig. 8b, c). Of note, *WRN* knockout had a potent effect on cell fitness with an effect size similar to core fitness genes (Fig. 5a, b). Furthermore, we mined data from systematic RNA interference screens and confirmed WRN dependency in MSI cancer cell lines[12] (Extended Data Fig. 8d), and confirmed that *WRN* downregulation by RNA interference robustly impaired growth in MSI HCT116 cells (Extended Data Fig. 8e, f), thus providing validation in an orthogonal experimental system. Despite the strong association between MMR deficiency and WRN dependency, knockout of *MLH1* in microsatellite-stable SW620 cell line did not induce WRN dependency; conversely complementation of HCT116 cells with chromosomes that contain *MLH1* and/or *MSH3*—to restore their expression and correct MMR deficiency[18]—did not revert the effect of *WRN* knockout (Extended Data Fig. 9).

To determine whether the loss-of-fitness effect was selective to WRN and identify a potential strategy for drug targeting, we performed functional rescue experiments using wild-type, or hypomorphic versions of mouse *Wrn* (resistant to the *WRN* sgRNAs that we used) with a mutation in the exonuclease (E78A) or helicase (R799C or T1052G) domain to impair protein function[19–21]. Expression of wild-type or exonuclease-deficient Wrn rescued knockout of *WRN* in MSI cells, whereas expression of helicase-deficient Wrn led to no (R799C) or weak (T1052G) rescue (Fig. 5c and Extended Data Fig. 10a, b). Thus, the helicase activity of WRN is required and is an important domain that can be used for therapeutic targeting.

To evaluate in vivo sensitivity of MSI cells to WRN depletion, we developed a doxycycline-inducible *WRN* sgRNA system in HCT116 cells (Extended Data Fig. 10c, d). Following subcutaneous engraftment of *WRN* sgRNA-expressing HCT116 cells in mice, treatment with doxycycline led to significant growth suppression of established tumours and a reduction in the number of proliferating cells (Fig. 5d–f and Extended Data Fig. 10e, f). These findings confirm that WRN is necessary to sustain in vivo growth of colorectal cancer cells with MSI.

Discussion

New approaches are needed to effectively prioritize candidate therapeutic targets for cancer treatments. We performed CRISPR–Cas9 screens in a diverse collection of cancer cells lines and combined this with genomic and tractability data to systematically nominate new cancer targets in an unbiased way. Confirmatory studies are necessary to further evaluate the priority targets that we identified. Even a modest improvement in drug-development success rates, and an expanded repertoire of targets, through approaches such as ours could provide benefits to patients with cancer. Our CRISPR–Cas9 screening results are also a resource with diverse applications in fundamental and evolutionary biology, genome engineering and disease genetics. Results are available through the project Score database (https://score.depmap.sanger.ac.uk/).

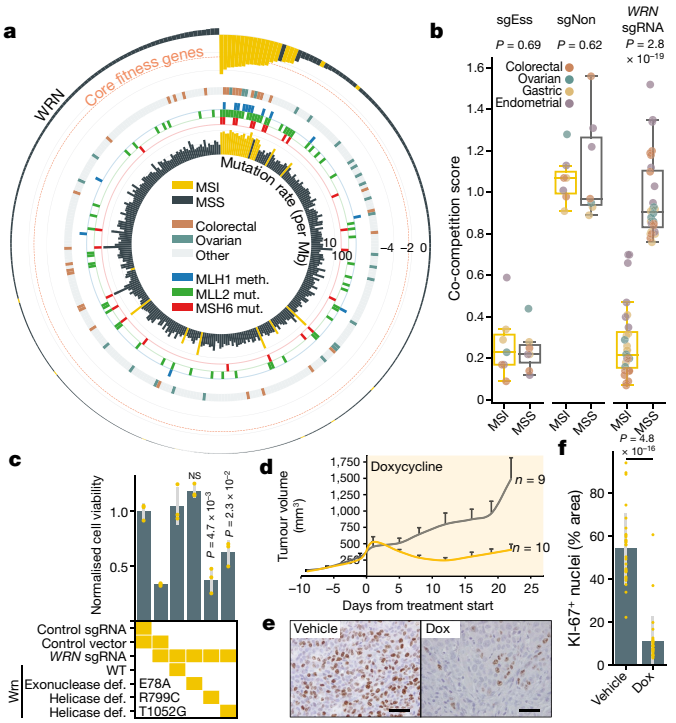


Fig. 5 | WRN is a target in MSI cancer cells. a, Circle plot of cell lines. From the outer ring to inner ring the following are shown: the fitness effect of *WRN* knockout and mean effect of core fitness genes (red dashed line); cancer-type; *MLH1* methylation (meth.) status; mutation (mut.) status of *MLL2* and *MSH6*; and the DNA mutation rate. **b**, WRN dependency in a co-competition assay. sgRNAs that target essential (sgEss) and non-essential (sgNon) genes were used as controls. Each point represents the mean co-competition score for a cell line (seven MSI and seven microsatellite stable (MSS) lines in duplicate); four WRN sgRNA guides were used. A score less than 1 denotes selective depletion of sgRNA-expressing knockout cells. Box-and-whisker plots show 1.5× the interquartile range and the median. *P* values were determined using a two-sided Welch's *t*-test. **c**, *WRN* rescue using wild-type (WT), exonuclease-deficient (def.) or helicase-deficient mouse *Wrn* in SW48 cells with MSI. Mean ± s.d. from 3 independent experiments. *P* values were calculated using a standard two-sided *t*-test assuming equal variance; comparison to wild-type *Wrn*. NS, not significant. **d**, Tumour volume of *WRN* sgRNA-expressing HCT116 (clone a) xenografts treated with doxycycline (yellow line) or vehicle (grey line). $P = 0.006$, two-way ANOVA. Data are mean ± s.e.m. Numbers of mice in each cohort are indicated. **e**, Representative KI-67 immunohistochemistry assessment of *WRN* sgRNA-expressing HCT116 (clone a) tumours explanted after one week. Scale bar, 50 μm; 40× magnification. **f**, Quantification of KI-67 staining. Data are mean ± s.d. of 10 fields from three different samples. $n = 30$; *P* value were calculated using a two-sided Welch's *t*-test.

We identified WRN as a promising new synthetic lethal target in MSI tumours. This finding is corroborated by the accompanying study by Chan et al.[22]. WRN physically interacts with MMR proteins[23], can resolve DNA recombination intermediates[24], and the yeast homologue Sgs1 has a redundant function with MMR proteins to suppress homeologous recombination in regions of nucleotide mismatch[25]. Together with our finding that modulation of MMR proteins alone is insufficient to confer WRN dependency, this suggests a model in which WRN is required to resolve the genomic structures present in MMR-deficient cells, which are possibly homeologous recombination structures, and failure to efficiently resolve these underpins the synthetic lethal dependency. Mutation of *WRN* leads to Werner syndrome, an autosomal recessive disorder characterized by premature ageing and an increased risk of cancer[16]. Thus, loss of WRN is compatible with human development; however, targeting WRN could result in damage to normal cells. Consideration should be given to maximizing therapeutic benefits through patient selection and dose scheduling. A possible route

for clinical development of WRN antagonists would be as an adjunct therapy to approved immune checkpoint inhibitors in MSI tumours[26].

In summary, we developed an unbiased and systematic framework that effectively ranks priority targets, such as WRN. Efforts such as ours, and from others[5,8,12,22,27,28], to build a compendium of fitness genes, and the identification of context-specific dependencies as part of a cancer dependency map, could be transformative to improve success rates in the development of cancer drugs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1103-9.

Received: 3 August 2018; Accepted: 8 March 2019;
Published online: 10 April 2019

1. Garraway, L. A. Genomics-driven oncology: framework for an emerging paradigm. *J. Clin. Oncol.* **31**, 1806–1814 (2013).
2. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
3. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
4. Koike-Yusa, H., Li, Y., Tan, E.-P., Del Castillo Velasco-Herrera, M. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR–guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
5. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
6. van der Meer, D. et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.* **47**, D923–D929 (2019).
7. Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).
8. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
9. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* **7**, 2719–2727 (2017).
10. Tzelepis, K. et al. A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Rep.* **17**, 1193–1205 (2016).
11. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
12. McDonald, E. R. III et al. Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell* **170**, 577–592 (2017).
13. Massacesi, C. et al. PI3K inhibitors as new cancer therapeutics: implications for clinical trial design. *OncoTargets Ther.* **9**, 203–210 (2016).
14. Brown, K. K. et al. Approaches to target tractability assessment — a practical perspective. *MedChemComm* **9**, 606–613 (2018).
15. Viswanathan, V. S. et al. Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* **547**, 453–457 (2017).
16. Chu, W. K. & Hickson, I. D. RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer* **9**, 644–654 (2009).
17. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).
18. Haugen, A. C. et al. Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer Res.* **68**, 8465–8472 (2008).
19. Perry, J. J. P. et al. WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat. Struct. Mol. Biol.* **13**, 414–422 (2006).
20. Kamath-Loeb, A. S., Welcsh, P., Waite, M., Adman, E. T. & Loeb, L. A. The enzymatic activities of the Werner syndrome protein are disabled by the amino acid polymorphism R834C. *J. Biol. Chem.* **279**, 55499–55505 (2004).
21. Ketkar, A., Voehler, M., Mukiza, T. & Eoff, R. L. Residues in the RecQ C-terminal domain of the human Werner Syndrome helicase are involved in unwinding G-quadruplex DNA. *J. Biol. Chem.* **292**, 3154–3163 (2017).
22. Chan, E. M. et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* https://doi.org/10.1038/s41586-019-1102-x (2019).
23. Saydam, N. et al. Physical and functional interactions between Werner syndrome helicase and mismatch-repair initiation factors. *Nucleic Acids Res.* **35**, 5706–5716 (2007).
24. Opresko, P. L., Sowd, G. & Wang, H. The Werner syndrome helicase/exonuclease processes mobile D-loops through branch migration and degradation. *PLoS ONE* **4**, e4825 (2009).
25. Myung, K., Datta, A., Chen, C. & Kolodner, R. D. SGS1, the *Saccharomyces cerevisiae* homologue of BLM and WRN, suppresses genome instability and homeologous recombination. *Nat. Genet.* **27**, 113–116 (2001).
26. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
27. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 (2017).
28. Wang, T. et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* **168**, 890–903 (2017).

Acknowledgements We thank D. Adams, G. Vassiliou and L. Parts for comments on the manuscript, members of the M.J.G. laboratory and Sanger Institute facilities (Wellcome Trust grant 206194). Work was funded by Open Targets (OTAR015) to M.J.G., K.Y. and J.S.-R. The K.Y. laboratory is supported by Wellcome Trust (206194). The M.J.G. laboratory is supported by SU2C (SU2C-AACR-DT1213) and Wellcome Trust (102696 and 206194). Support was also received from AIRC 20697 (A.B.) and 18532 (L.T.); 5x1000 grant 21091 (A.B. and L.T.); ERC Consolidator Grant 724748 – BEAT (A.B.); FPRC-ONLUS, 5x1000 Ministero della Salute 2011 and 2014 (L.T.); and Transcan, TACTIC (L.T.).

Author contributions M.J.G., K.Y. and C.B.-D. conceived the project. F.M.B. led CRISPR–Cas9 screening, co-developed the project Score web portal, contributed to analysis strategy, performed validation analyses and verified WRN dependency. F.I. led computational analyses and figure preparation, and contributed to the project Score web portal. G.P. performed experiments to verify WRN dependency, carried out analyses and contributed to in vivo studies. E.G. contributed to computational analysis and figure preparation. D.v.d.M. contributed to the project Score web portal. G.M., F.S., M.P., A.B. and L.T. performed in vivo studies. C.M.B., R.A., D.A.J., R.M., R.P. and P.W. performed CRISPR–Cas9 screens. R.S. performed tractability analysis. Y.R. performed WRN rescue experiments. C.M.B., S.H., A.B., L.T., E.A.S., D.D. and J.S.-R. assisted with project supervision. F.M.B., F.I., E.G., G.P., K.Y. and M.J.G. wrote the manuscript. K.Y. and M.J.G. directed the project. J.S.-R., A.B., L.T., M.J.G. and K.Y. acquired funding. All authors approved the manuscript.

Competing interests E.A.S., D.D., C.B.-D., R.S. and Y.R. are GlaxoSmithKline employees. Open Targets is a public–private initiative involving academia and industry. K.Y. and M.J.G. receive funding from AstraZeneca. M.J.G. performed consultancy for Sanofi. All other authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1103-9.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1103-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to K.Y. or M.J.G.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

CRISPR–Cas9 screening. *Plasmids.* All plasmids have previously been described[10] and are available through Addgene (Cas9 vector, 68343; gRNA vector, 67974). Plasmids were packaged using the ViraPower Lentiviral Expression System (Invitrogen, K4975-00) as per the manufacturer's instructions.

Cell culture. Cell lines used in this study (Supplementary Table 1) were selected from 1,000 cell line panel[7] of the Genomics of Drug Sensitivity in Cancer study, had been annotated in the Cell Model Passports database (https://cellmodelpassports.sanger.ac.uk/) and were maintained as previously described[7]. To control for cross-contamination and sample swaps a panel of 92 single-nucleotide polymorphisms was profiled for each cell line before and following completion of the CRISPR–Cas9 screening pipeline. This study includes commonly misidentified cell lines: Ca9-22, short tandem repeat (STR) analysis confirmed that the identity matched the Japanese Collection of Research Bioresources Cell Bank (JCRB) reference (JCRB0625) and RIKEN (RCB1976); MKN28, noted as derivative of MKN74 in Cell Model Passports and clinical information matches MKN74; KP-1N, known misidentification issue, Cell Model Passports data for both KP-1N & Panc-1 are identical; OVMIU, known misidentification issue, Cell Model Passports data for both OVMIU and OVSAYO are identical; SK-MG-1, STR profile matches JCRB profile, which internally matches Marcus, Cell Model Passport data for both SK-MG-1 and Marcus are identical. Commonly misidentified lines have been noted in Supplementary Table 1 and on the Cell Model Passport. All commonly misidentified cell lines were retained, because the misidentification does not impact tissue or cancer type of origin, and all datasets used were generated in-house from the same matched cell line.

A separate set of HCT116 cell lines was used for WRN validation experiments: HCT116 parental cells and HCT116 cells carrying Chr.3 or Chr.5, or both were a gift from M. Koi. HCT116 cells carrying Chr.2 were a gift from A. Goel. HCT116 cells carrying Chr.2 or Chr.3 were maintained in 400 $\mu g\ ml^{-1}$ G418 (Thermo Fisher Scientific, 10131027); HCT116 cells carrying Chr.5 were maintained in 6 $\mu g\ ml^{-1}$ blasticidin (Thermo Fisher Scientific, A1113903); and HCT116 cells carrying Chr.3 + Chr.5 were maintained in the presence of 400 $\mu g\ ml^{-1}$ G418 and 6 $\mu g\ ml^{-1}$ blasticidin. All cells were cultured in McCoy's 5A medium (Sigma-Aldrich, M4892) with 10% FBS.

Generation of Cas9-expressing cancer cell lines. Cells were transduced with a lentivirus containing Cas9 in T25 or T75 flasks at approximately 80% confluence in the presence of polybrene (8 $\mu g\ ml^{-1}$). Cells were incubated overnight followed by replacement of the lentivirus-containing medium with fresh complete medium. Blasticidin selection commenced 72 h after transduction at an appropriate concentration determined for each cell line using a blasticidin dose–response assay (blasticidin range, 10–75 $\mu g\ ml^{-1}$) and cell viability was assessed using the CellTiter-Glo 2.0 Assay (Promega, G9241). Cas9 activity was assessed as described previously[10]. Cell lines with Cas9 activity over 75% were used for sgRNA library transduction.

Genome-wide sgRNA library and screen. Two genome-wide sgRNA libraries were used in this study: the Human CRISPR Library v.1.0 and v.1.1. The Human CRISPR Library v.1.0 was described previously and targets 18,009 genes with 90,709 sgRNAs (Addgene, 67989)[10]. The Human CRISPR Library v.1.1 contains all sgRNAs from v.1.0 plus 1,004 non-targeting sgRNAs and 5 additional sgRNAs against 1,876 selected genes that encode kinases, epigenetic-related proteins and pre-defined fitness genes. An oligo pool of Library v.1.1 was synthesized using high-throughput silicon platform technology (Twist Bioscience) and cloned as described previously[10]. For consistency, all computational analyses were performed considering only the overlapping sgRNAs between the two libraries (90,709 sgRNAs). Data for the additional sgRNAs in Library v.1.1 can be found in the raw read count files for cell lines screened with this library version (available at available at https://cog.sanger.ac.uk/cmp/download/raw_sgrnas_counts.zip), but have been removed before quality control analysis. The HT-29 cell line was screened with both libraries and resulting datasets were kept separated for comparative analyses (results are summarized in Extended Data Fig. 2j).

A total of 3.3×10^7 cells were transduced with an appropriate volume of the lentiviral-packaged whole-genome sgRNA library to achieve 30% transduction efficiency (100× library coverage). The volume was determined for each cell line using a titration of the packaged library and assessing the percentage of blue fluorescent protein (BFP)-positive cells by flow cytometry. Transductions were performed in technical triplicate (or duplicate for cell lines with a large cell size such as glioblastoma). Owing to the large number of screens performed, multiple batches of packaged library virus were prepared. Each batch was tested in HT-29 cells to ensure consistency between batch preparations. In addition, the HT-29 cell line was screened every 3 months to ensure the quality of data generated by the pipeline was consistent. Transduction efficiency was assessed 72 h after transduction. Samples with a transduction efficiency between 15 and 60% were used for puromycin selection. The appropriate concentration of puromycin for each individual cell line was determined from a dose–response curve (puromycin range, 1–5 $\mu g\ ml^{-1}$) and cell viability was assessed using a CellTiter-Glo 2.0 Assay (Promega, G9241). The percentage BFP-positive cells was reassessed after a minimum of 96 h of puromycin selection. For samples with less than 80% BFP-positive cells, puromycin selection was extended for an additional 3 days and the percentage of BFP-positive cells was assessed again. Cells were maintained until day 14 after transduction with a minimum of 5.0×10^7 cells reseeded at each passage (500× library coverage). Approximately 2.5×10^7 cells were collected, pelleted and stored at −80 °C for DNA extraction.

DNA extraction, sgRNA PCR amplification, Illumina sequencing and sgRNA counting. Genomic DNA was extracted from cell pellets using either the QIAsymphony automated extraction platform (Qiagen, QIAsymphony DSP DNA Midi Kit, 937255) or by manual extraction (Qiagen, Blood & Cell Culture DNA Maxi Kit, 13362) as per the manufacturer's instructions. PCR amplification, Illumina sequencing (19-bp single-end sequencing with custom primers on the HiSeq2000 v.4 platform) and sgRNA counting were performed as described previously[10].

CRISPR screen data analyses. *Low-level quality control assessment and filtering.* To perform initial low-level quality control, the Pearson's correlation of treatment counts between replicates was assessed for each cell line (Extended Data Fig. 1c). The resulting correlation scores were generally high (median = 0.8), but not sufficiently distinguishable from expectation (median correlation between replicates of any pair of randomly selected cell lines). Thus, to define a reproducibility threshold, we developed an approach based on a previously published study[29]. Specifically, we selected a set of the 838 most-informative sgRNAs, defined as those with an average pairwise Pearson's correlation greater than 0.6 between corresponding patterns of the count fold changes 14 days after transfection versus plasmid library across all screened cell lines. We next computed average gene-level profiles for 308 genes targeted by these informative sgRNAs for each individual technical replicate, and then computed all possible pairwise Pearson's correlation scores between the resulting profiles. This enabled the estimation of a null distribution of replicate correlations (plotted in grey in Extended Data Fig. 1d). We then defined a reproducibility threshold R value of 0.68, for which the estimated probability mass function of the correlation scores that was computed between replicates of the same cell line (considering the identified 308 genes only) was at least twice that of the null mass probability function (Extended Data Fig. 1d). Of the 332 screened cell lines with at least two technical replicates, 305 had an average replicate correlation higher than this threshold, and therefore passed the reproducibility assessment; for 7 cell lines there were no replicates. Excluding the least reproducible replicate for the 14 cell lines that did not pass the first reproducibility assessment allowed their average replicate correlation to exceed the threshold defined above, thus resulting in a set of 326 cell lines that passed the low-level quality control assessment (Supplementary Table 1).

Screening performance assessment. We considered the genome-wide profiles of gene-level sgRNA fold change values (averaged across targeting sgRNAs and replicates) of each cell line to be a classifier of predefined sets of essential and non-essential genes[30] by means of receiver operating characteristic (ROC) indicators (Extended Data Fig. 1g and Supplementary Table 1). In addition, we measured the magnitude of the depletion signal observed in each screened cell line by evaluating the median log(change in sgRNA count), and the discriminative distance between their distributions (as measured by the Glass's Δ) for predefined essential and non-essential genes[30] and ribosomal protein genes[31]. In total, 2 out of the 326 cell lines were manually removed, because they had area under the ROC curve, area under the precision/recall curve and both Glass's Δ values that were 3 s.d. lower than the average. On the basis of our low-level quality control and screening performance, the final analysis set was composed of 324 cell lines (Supplementary Table 1). Further details on these analyses are included in the Supplementary Information.

sgRNA count preprocessing and CRISPR-bias correction. The analysis set of 324 cell lines was further processed using CRISPRcleanR[32] (https://github.com/francescojm/CRISPRcleanR). sgRNAs with less than 30 reads in the plasmid counts and sgRNAs belonging to only the Library v1.1 were first removed. The remaining sgRNAs were assembled into one file per cell line, including the read counts from the matching library plasmid and all replicates and then normalized using a median–ratio method to adjust for the effect of library sizes and read count distributions[33]. Depletion/enrichment fold changes for individual sgRNAs were quantified between post library-transduction read counts and library plasmid read counts at the individual replicate level. This was performed using the ccr.NormfoldChanges function of CRISPRcleanR. Next we performed a correction of gene-independent responses to CRISPR–Cas9 targeting[34] using the ccr.GWclean function of CRISPRcleanR with default parameters.

Calling CRISPR–Cas9 gene knockout fitness effects. The CRISPRcleanR-corrected sgRNAs-level values (corrected fold change values) were used as input into an in-house-generated R implementation of the BAGEL method[30] to call significantly depleted genes (code publicly available at https://github.com/francescojm/BAGELR). Our BAGEL implementation computes gene-level Bayesian factors by the sgRNAs on a targeted-gene basis, by averaging instead of summing them.

Additionally, it uses reference sets of predefined essential and non-essential genes[30]. However, in order to avoid their status (essential or non-essential) being defined a priori, we removed any high-confidence cancer driver genes as defined previously[7] from these sets. The resulting curated reference gene sets are available as built-in data objects in the R implementation of BAGEL (curated_BAGEL_essential.rdata and curated_BAGEL_nonEssential.rdata, both available at https://github.com/francescojm/BAGELR/tree/master/data). A statistical significance threshold for gene-level Bayesian factors was determined for each cell line as described previously[8]. Each gene was assigned a scaled Bayesian factor computed by subtracting the Bayesian factor at the 5% FDR threshold defined for each cell line from the original Bayesian factor, and a binary fitness score equal to 1 if the resulting scaled Bayesian factor was greater than 0. Further details on these analyses are included in the Supplementary Information.

In addition, CRISPRcleanR-corrected sgRNA treatment counts were derived from the corrected sgRNA-level count fold changes (using the ccr.correctCounts function of CRISPRcleanR) and used as input into MAGeCK[35] to compute the depletion significance using mean–variance modelling. This was performed using the MAGeCK Python package (version 0.5.3), specifying in the command line call that no normalization was required (as this was already performed by CRISPRcleanR). At the end of this stage, the following gene-level depletion score matrices were produced for each cell line: raw count fold changes, copy number bias-corrected count fold changes, Bayesian factors, scaled Bayesian factors, binary fitness scores and MAGeCK depletion FDRs. All scores are summarized for each cell line and available at https://cog.sanger.ac.uk/cmp/download/essentiality_matrices.zip, together with all the sgRNAs raw count files (available at https://cog.sanger.ac.uk/cmp/download/raw_sgrnas_counts.zip).

High-level CRISPR screen data analyses. *Adaptive daisy model (ADaM) to identify core fitness genes.* We designed the adaptive daisy model (ADaM), an heuristic algorithm for the identification of core fitness genes, implemented it in an R package and made it publicly available at https://github.com/francescojm/ADaM. ADaM is based on the daisy model[8], but it adaptively determines the minimal number of cell lines m from a given cancer type in which a gene should exert a significant fitness effect for that gene to be considered a core fitness gene for that cancer type. ADaM is described further in the Supplementary Information. In order to identify pan-cancer core fitness genes, we applied the same method to determine the minimal number k of cancer types for which a gene should be predicted as a pan-cancer core fitness gene.

Characterization of ADaM pan-cancer core fitness genes. Reference sets of essential and non-essential genes were extracted from a previously published study[30]. Other reference gene sets (used while characterizing the ADaM pan-cancer core fitness genes, described below) were derived from the Molecular Signature Database (MSigDB[36]) and post-processed as described previously[32]. A more recent set of a priori known essential genes was derived from a previously published study[9]. The pan-cancer core fitness genes that did not belong to any of the aforementioned gene sets were tested for gene family enrichments (using a hypergeometric test) by deriving gene annotations using the BioMart R package[37] and biological pathway enrichments using a comprehensive collection of pathways gene sets from Pathway Commons[38] (post-processed to reduce redundancies across different sets as described previously[39]). All enrichment P values were corrected using the Benjamini–Hochberg method. Results are shown in Supplementary Table 4.

Comparison between the ADaM pan-cancer core fitness genes and other reference sets of essential genes. We compared the pan-cancer core fitness genes identified by ADaM with the BAGEL reference set of essential genes[30], and a more recently proposed larger set of essential genes[9] in terms of size, estimated precision (number of included true positive genes/number of included genes) and recall (number of included true positive genes/total number of true positive genes). In these comparisons, we used gold-standard essential genes involved in cell essential processes (downloaded from the MSigDB[36] and post-processed as described previously[32]). In addition, we estimated FDRs for the three gene sets (number of included false positive genes/total number of false positive genes) considering genes predicted to be strongly context-specific essential (thus not core-fitness essential) to be false-positive genes according to a previous publication[12], and using three different confidence levels, as further described in the Supplementary Information.

Basal expression of cancer-type specific core fitness genes in normal tissues. Basal gene median reads per kilobase of transcript per million mapped reads in normal human tissues were downloaded from the GTEx Portal[40], log-transformed and quantile-normalized on a tissue-type basis.

Statistical and computational analyses. *ANOVA to identify genomic correlates with gene fitness.* We performed a systematic ANOVA to test associations between gene-level fitness effects and the presence of 484 cancer driver events (CDEs; 151 single-nucleotide variants and 333 copy number variants)[7] or MSI status at the pan-cancer as well as individual cancer-type levels. In total, 10 cancer types with at least 10 screened cell lines were analysed (breast carcinoma, colorectal carcinoma, gastric carcinoma, head and neck carcinoma, lung adenocarcinoma, neuroblastoma, oral cavity carcinoma, ovarian carcinoma, pancreatic carcinoma and squamous cell lung carcinoma). The remaining cancer types were collapsed on a tissue basis (annotation in Supplementary Table 1) and the resulting tissues with at least 10 cell lines were included in the analysis (bone, central nervous system, oesophagus, haematopoietic and lymphoid). A total of 14 analyses (referred for simplicity as cancer-type-specific ANOVAs in the main text and below) plus a pan-cancer analysis including all screened cell lines were performed. Each ANOVA was performed using the analytical framework described previously[7] and implemented in a Python package[41] (https://github.com/CancerRxGene/gdsctools). Only genes that did not belong to any set of prior known essential genes (defined in the previous sections) and not predicted by ADaM to be core fitness genes were included in the analyses. For all tested gene fitness–CDE associations, effect size estimations versus pooled s.d. (quantified using Cohen's d), effect sizes versus individual s.d. (quantified using two different Glass's Δ metrics, for the CDE-positive and the CDE-negative populations separately), CDE P values and all other statistical scores were obtained from the fitted models. An association was tested only if at least three cell lines were contained in the two sets resulting from the dichotomy induced by CDE status (that is, at least three CDE-positive and three CDE-negative cell lines). The P values from all ANOVAs were corrected together using the Tibshirani–Storey method[42]. Subsequently, MSI status was also tested for statistical associations with differential gene fitness effects for pan-cancer and cancer types with at least three MSI cell lines. We used the following statistical significance and effect size thresholds for category associations between gene fitness effects and genomic markers:

Class A marker: a P-value threshold of 10^{-3} with a FDR threshold equal to 25% (or 5% for MSI) and with Glass's $\Delta > 1$. Different FDR thresholds were used for associations with CDEs or MSI because the number of tests performed in the former was six orders of magnitude larger than the latter.

Class B marker: a FDR threshold of 30% with at least one Glass's $\Delta > 1$ for pan-cancer associations.

Class C marker or weaker: an ANOVA P-value threshold of 10^{-3} and for pan-cancer associations at least one Glass's $\Delta > 1$; for weaker, a simple Student's t-test (for difference assessment of the mean depletion fold change between CDE-positive/CDE-negative cell lines) P-value threshold of 0.05 and for pan-cancer associations, at least one Glass's $\Delta > 1$.

The additional constraint of Glass's Δ values (quantifying the effect size with respect to the s.d. of the two involved sub-populations of samples) was considered for the pan-cancer markers in order to account for the significantly larger number of samples analysed in the pan-cancer setting, which might result in highly significant P values even for small effect size associations. Further details on this analysis are reported in the Supplementary Information.

Target priority scores and target tractability. Computation of the target priority scores and their significance is described in the Supplementary Information. To estimate the likelihood of a target to bind a small molecule or the likelihood of a target to be accessible to an antibody, we made use of a genome-wide target tractability assessment pipeline[14]. The in silico pipeline integrates data from public sources, and assigns human protein-coding genes into hierarchical qualitative buckets. Predicted tractability and confidence in the data increased from bucket 10 to bucket 1; targets in bucket 1 were considered to be the most tractable. Of note, targets in lower buckets (that is, buckets 10 to 8) were considered to have an uncertain tractability, and should not be ruled out as 'intractable' without a deep tractability assessment. Further details are provided in the Supplementary Information.

Characterization of target protein families and enrichment analysis. To characterize protein families and compute statistical enrichment, we made use of the Panther online tool[43].

GPX4 differential expression analysis. RNA-sequencing gene expression measurements transformed using voom[44] were obtained from a previously published study[45]. For *GPX4* analysis, cell lines were divided into two groups according to their loss-of-fitness response to *GPX4* knockout (using BAGEL FDR < 5% as significance threshold for gene depletion) and gene expression fold changes were calculated between the *GPX4* non-dependent and dependent cell lines ($\log_2$ values of the mean difference). Differential gene expression was statistically assessed using the R package Limma[46]. Gene set enrichment analysis was performed with ssGSEA[36] and cancer hallmark gene sets were used to identify significant enrichment among the top differentially expressed genes. Then, 10,000 random permutations were performed for each signature to calculate empirical P values and a Benjamini–Hochberg FDR correction was applied.

WRN dependency in MSI cell lines. *Co-competition assay.* The sequences of sgRNAs that target *WRN* and cell lines used in validation experiments are described in Supplementary Table 10. This included two sgRNA from the original screen and two independent sgRNAs. The sgRNAs were cloned into pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (Addgene, 67974). Cell lines were transduced at around 50% efficiency as described above in six-well plates. A co-competition

score was determined as the ratio of the percentage BFP-positive cells (that is, sgRNA-positive cells) on day 14 compared to day 4, as measured by flow cytometry. A co-competition score less than 1 indicates a relative reduction in BFP-positive cells, resulting from targeting of a loss-of-fitness gene.

Clonogenic assay. Cell lines were transduced with lentivirus that encodes *WRN* sgRNA at around 100% efficiency as described above in six-well plates (2,000 cells per well), typically for 15–21 days. Cells were fixed using 100% ice-cold ethanol for 30 min followed by Giemsa staining overnight at room temperature.

Western blot analysis. Cells were transduced at around 100% as described above in 10-cm dishes. Day 5 after transduction, cells were lysed with 200 μl RIPA buffer supplemented with protease and phosphatase inhibitors and lysates were used for SDS–PAGE and immunoblot analysis. Antibodies used were: WRN (Cell Signaling Technologies, 4666; dilution 1:2,000), WRN for domain rescue experiment (Thermo Fisher Scientific, PA5-27319); MLH1 (Cell Signaling Technologies, 3515; dilution 1:1,000); MSH3 (Santa Cruz Biotechnology, sc-271080; dilution 1:1,000); anti-Flag M2 (Sigma-Aldrich, F3165); β-actin (Cell Signaling Technologies, 4970); and anti-β-tubulin (Sigma-Aldrich, T4026: dilution 1:5,000). Secondary antibodies included: IRDye 800CW donkey anti-mouse antibody (LI-COR, 926-32212); IRDye 680LT donkey anti-rabbit IgG (H+L) (LI-COR, 925-68023); anti-mouse IgG HRP-linked secondary antibody (GE Healthcare, NA931). Molecular weight markers included: SeeBlue Plus2 Pre-stained Protein Standard (Thermo Fisher Scientific, 5925) and Precision Plus Protein Standards (BioRad, 161-0373).

WRN rescue experiment. SW620 and SW48 cells (2×10^5 cells) were transfected by nucleofection (Lonza 4D Nucleofector Unit X) with Cas9–sgRNA ribonucleoproteins (RNP) targeting human *MAVS* (used as a non-essential knockout control) or *WRN*, together with overexpression of 200 ng pmGFP control or 200 ng mouse *Wrn* cDNA (Origene, MR226496). From each sample after nucleofection, 5,000 cells were seeded in a 96-well plate and allowed to grow for 5 days, after which cells were collected for either CellTiter-Glo assay (Promega, G9241) or western blot analysis. CellTiter-Glo data were read on an Envision Multiplate Reader and data analysis was performed using GraphPad Prism 7 software. Student's *t*-test was performed using the multiple *t*-test module in Prism 7. The sgRNA sequences that were used are listed in Supplementary Table 10.

RNA interference. A pool of four siRNAs that target *WRN* were used (Dharmacon, L-010378-00-0005). HCT116 cells were grown and transfected with siRNA using the RNAiMAX (Invitrogen) transfection reagent following the manufacturer's instructions. Each experiment included: mock control (transfection lipid only), ON-TARGETplus Non-targeting Control Pool (Dharmacon, D-001810-10-05) as a negative control, and polo-like kinase 1 (*PLK1*) (Dharmacon, L-003290-00-0010), which served as a positive control. siRNA sequences are listed in Supplementary Table 10.

Rescue of WRN dependency in HCT116 isogenic lines. HCT116 parental cells and derivatives carrying Chr.2, Chr.3, Chr.5 or Chr.3 + Chr.5 were transduced to express Cas9. After transduction, all lines displayed Cas9 activity >80%. To assess WRN dependency, cells were seeded at 1.5×10^3 cells per well in 100 μl complete growth medium in 96-well plastic cell culture plates. At day 0, cells were transduced with viral particles containing sgRNAs targeting essential or non-essential genes, or *WRN* sgRNA 1 and *WRN* sgRNA 4 in order to achieve a >90% transduction efficiency. The following day, the medium was replaced and 48 h after transduction puromycin was added at final concentration of 2 μg ml^{-1}. Plates were incubated at 37 °C in 5% CO_2 for 7 days, after which the cell viability was assessed using CellTiter-Glo (Promega) by measuring luminescence on an Envision multiplate reader. Clonogenic assays were performed as described in the 'WRN dependency in MSI cell lines' section; and 48 h after transduction puromycin was added at a final concentration of 2 μg ml^{-1}.

In vivo validation. *WRN knockout using an inducible CRISPR–Cas9 system.* To generate inducible *WRN* sgRNA-expressing HCT116 cells, we cloned *WRN* sgRNA 4 into the pRSGT16H-U6Tet-(sg)-CMV-TetRep-TagRFP-2A-Hygro vector (Cellecta). Cas9-expressing HCT116 cells were transduced and selected with 500 μg ml^{-1} of hygromycin (Thermo Fisher Scientific). To obtain cell populations that both uniformly express Cas9 and contain the inducible *WRN*-targeting sgRNA, we generated single-cell clones by serial dilution. To measure the growth rate of *WRN* sgRNA-expressing HCT116 cells after conditional induction of *WRN* knockout, cells were grown in flasks in the presence or absence of 2 μg ml^{-1} doxycycline for 24 h and then seeded in 96-well plates, with or without the same concentration of doxycycline. Cell growth was monitored every 6 h using an automated IncuCyte-FLR 4X phase-contrast microscope (Essen Instruments). The average object-summed intensity was calculated using the IncuCyte software (Essen Instruments).

Mouse xenograft studies. Female non-obese diabetic/severe combined immunodeficiency (NOD/SCID) mice (Charles River Laboratories) were used in all in vivo studies. All animal procedures were approved by the Ethical Committee of the Institute and by the Italian Ministry of Health (authorization 806/2016-PR). The methods were carried out in accordance with the approved guidelines. Mice were purchased from Charles River Laboratories, maintained in hyperventilated cages and manipulated under pathogen-free conditions. In particular, mice were housed in individually sterilized cages; each cage contained a maximum of seven mice and optimal amounts of sterilized food, water and bedding. HCT116 xenografts were established by subcutaneous inoculation of 2×10^6 cells into the right posterior flank of 5- to 6-week-old mice. Tumour size was evaluated by calliper measurements, and the approximate volume of the mass was calculated using the formula $4/3\pi \times (d/2)^2 \times (D/2)$, where d is the minor tumour axis and D is the major tumour axis. When tumours reached an average size of approximately 250–300 mm^3, animals with the most homogeneous size were selected and randomized by tumour size. Doxycycline (Sigma-Aldrich, D9891) was dissolved in water and administered daily at a 50 mg kg^{-1} concentration by oral gavage. For each experimental group, 8–10 mice were used to enable reliable estimation of within-group variability. Operators allocated mice to the different treatment groups during randomization but were blinded during measurements. The maximal tumour volume permitted in our in vivo experiments was 3,500 mm^3 and this limit was never exceeded. In vivo procedures and related biobank data were managed using the Laboratory Assistant Suite, a web-based proprietary data management system for automated data tracking[47].

Immunohistochemistry. Formalin-fixed, paraffin-embedded tissues explanted from cell xenografts were partially sectioned (10-μm thick) using a microtome. Then, 4-μm paraffin tissue sections were dried in a 37 °C oven overnight. Slides were deparaffinized in xylene and rehydrated through graded alcohol to water. Endogenous peroxidase was blocked in 3% hydrogen peroxide for 30 min. Microwave antigen retrieval was carried out using a microwave oven (750 W for 10 min) in 10 mmol l^{-1} citrate buffer, pH 6.0. Slides were incubated with monoclonal mouse anti-human KI-67 (1:100; DAKO) overnight at 4 °C inside a moist chamber. After washings in TBS, anti-mouse secondary antibody (DAKO Envision+System horseradish peroxidase-labelled polymer, DAKO) was added. Incubations were carried out for 1 h at room temperature. Immunoreactivities were revealed by incubation in DAB chromogen (DakoCytomation Liquid DAB Substrate Chromogen System, DAKO) for 10 min. Slides were counterstained in Mayer's haematoxylin, dehydrated in graded alcohol, cleared in xylene and a coverslip was applied using DPX (Sigma-Aldrich). A negative control slide was processed with only the secondary antibody, omitting the primary antibody incubation. Immunohistochemically stained slides for KI-67 were scanned with a 40× objective. Ten representative images selected from three cases were then analysed using ImageJ (NIH), which segmented cells with positive and negative nuclei. The percentage of the area containing positive cells was calculated as the brown area (positively stained cells) divided by the sum of brown and blue areas (negatively stained cells). The software interpretation was manually verified by visual inspection of the digital images to ensure accuracy.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

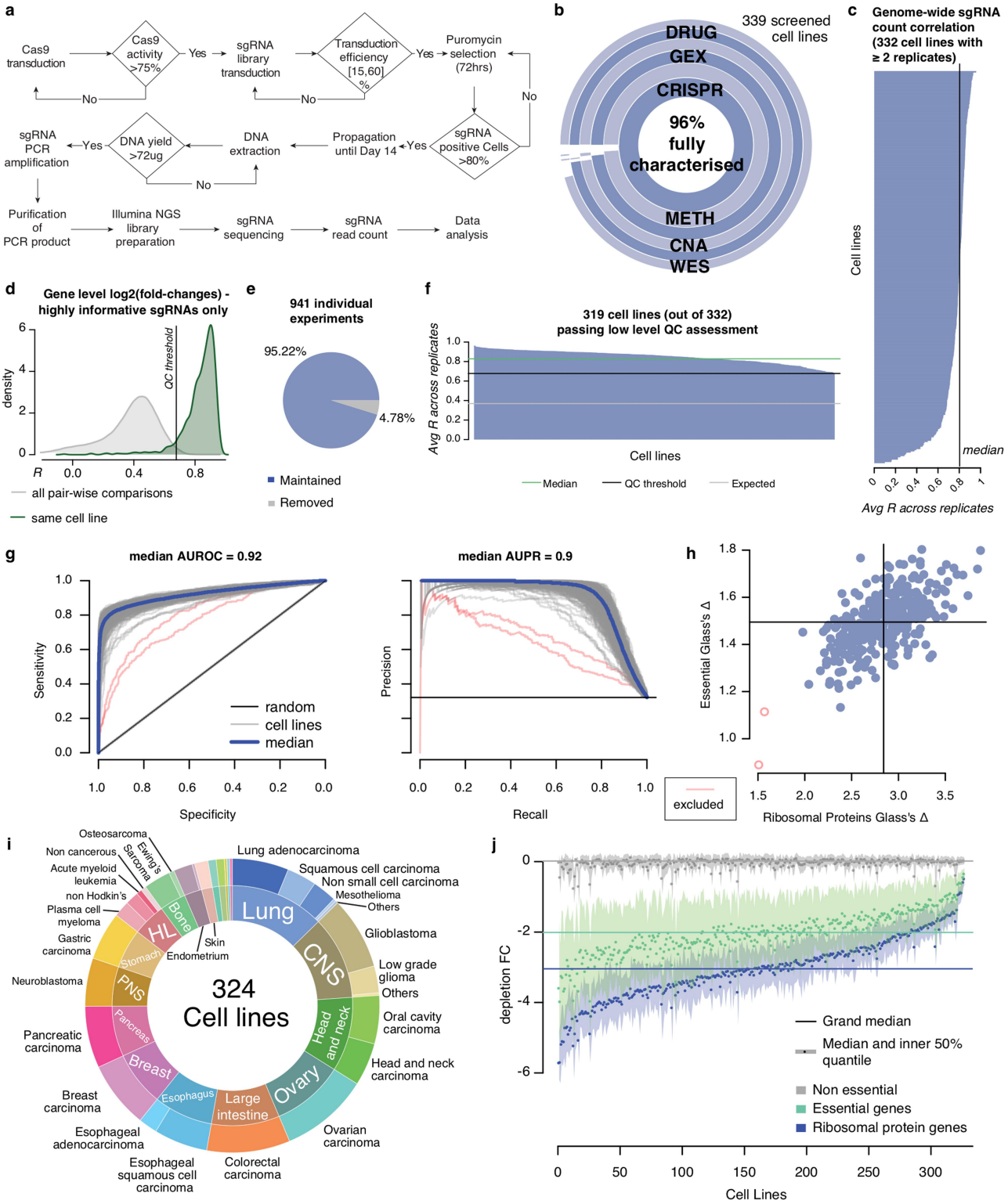
Data and analyses are included in the published article and supplementary data 1, 2 and 3 are available from FigShare (https://figshare.com/projects/CRISPRtargetID/60146). The gene fitness scores of the cell lines, raw counts of the sgRNA data, and processed data and results are available from the project Score web portal: https://score.depmap.sanger.ac.uk.

Code availability

Software code are available through GitHub at https://github.com/francescojm/CRISPRcleanR, https://github.com/francescojm/ADAM and https://github.com/francescojm/BAGELR.

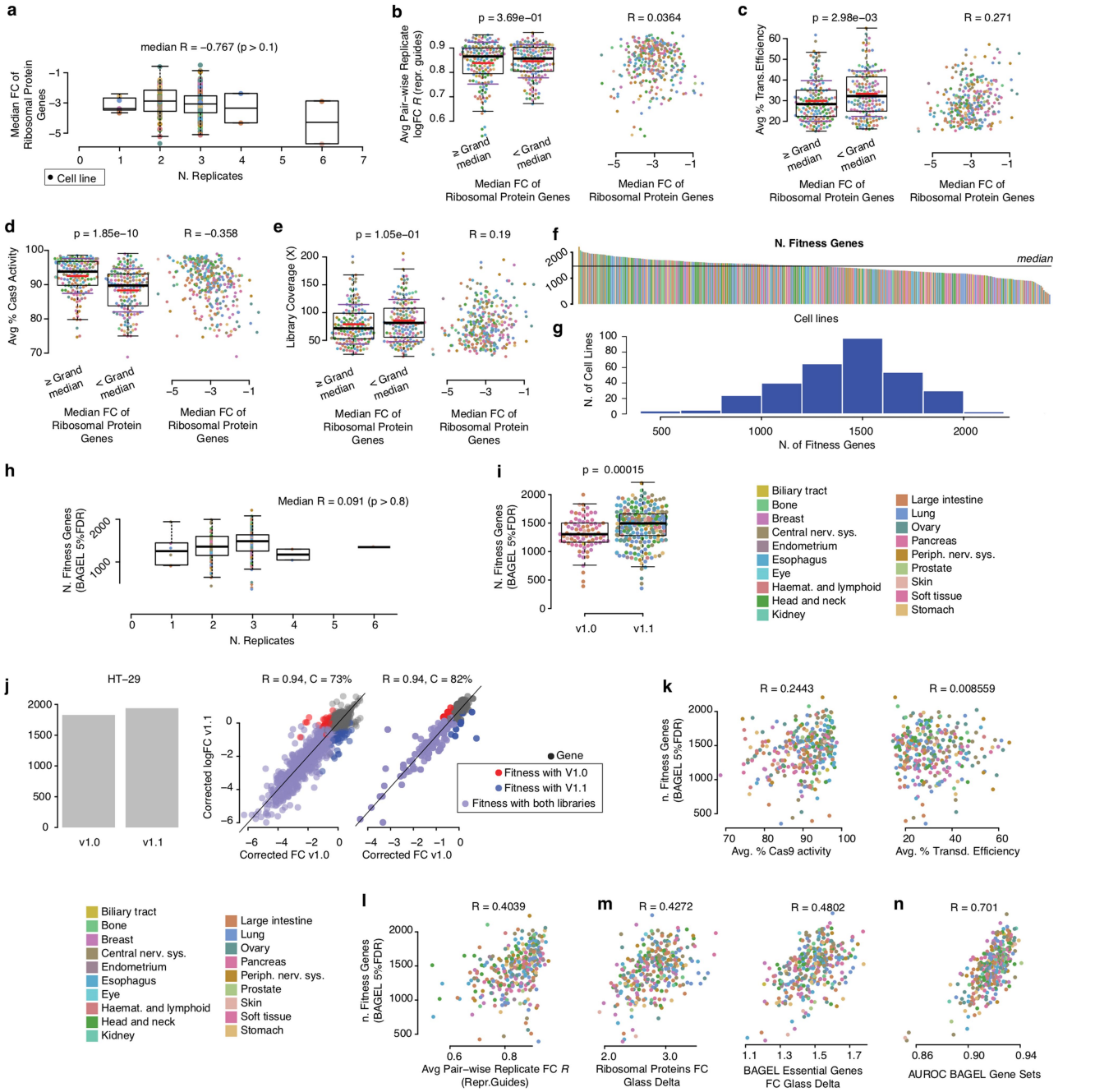
29. Ballouz, S. & Gillis, J. AuPairWise: a method to estimate RNA-seq replicability through co-expression. *PLOS Comput. Biol.* **12**, e1004868 (2016). Home (25 Doggett St)
30. Hart, T. & Moffat, J. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164 (2016).
31. Yoshihama, M. et al. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* **12**, 379–390 (2002).
32. Iorio, F. et al. Unsupervised correction of gene-independent cell responses to CRISPR–Cas9 targeting. *BMC Genomics* **19**, 604 (2018).
33. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
34. Aguirre, A. J. et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
35. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
36. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

37. Durinck, S. et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
38. Cerami, E. G. et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–D690 (2011).
39. Iorio, F. et al. Pathway-based dissection of the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich. *Sci. Rep.* **8**, 6713 (2018).
40. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
41. Cokelaer, T. et al. GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics* **34**, 1226–1228 (2018).
42. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
43. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* **41**, D377–D386 (2013).
44. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
45. Garcia-Alonso, L. et al. Transcription factor activities enhance markers of drug sensitivity in cancer. *Cancer Res.* **78**, 769–780 (2018).
46. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
47. Baralis, E., Bertotti, A., Fiori, A. & Grand, A. LAS: a software platform to support oncological data management. *J. Med. Syst.* **36**, 81–90 (2012).

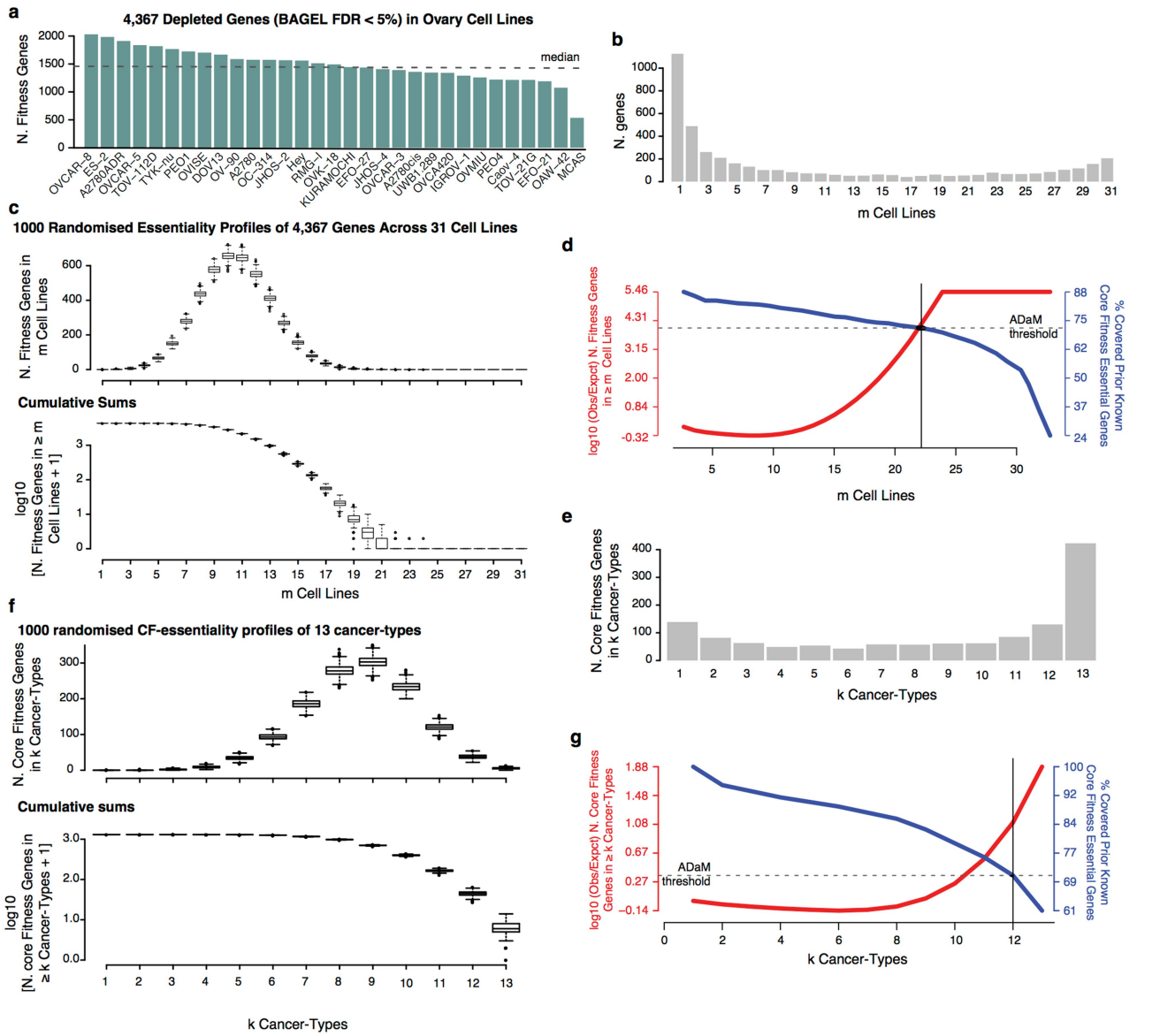


Extended Data Fig. 1 | See next page for caption.

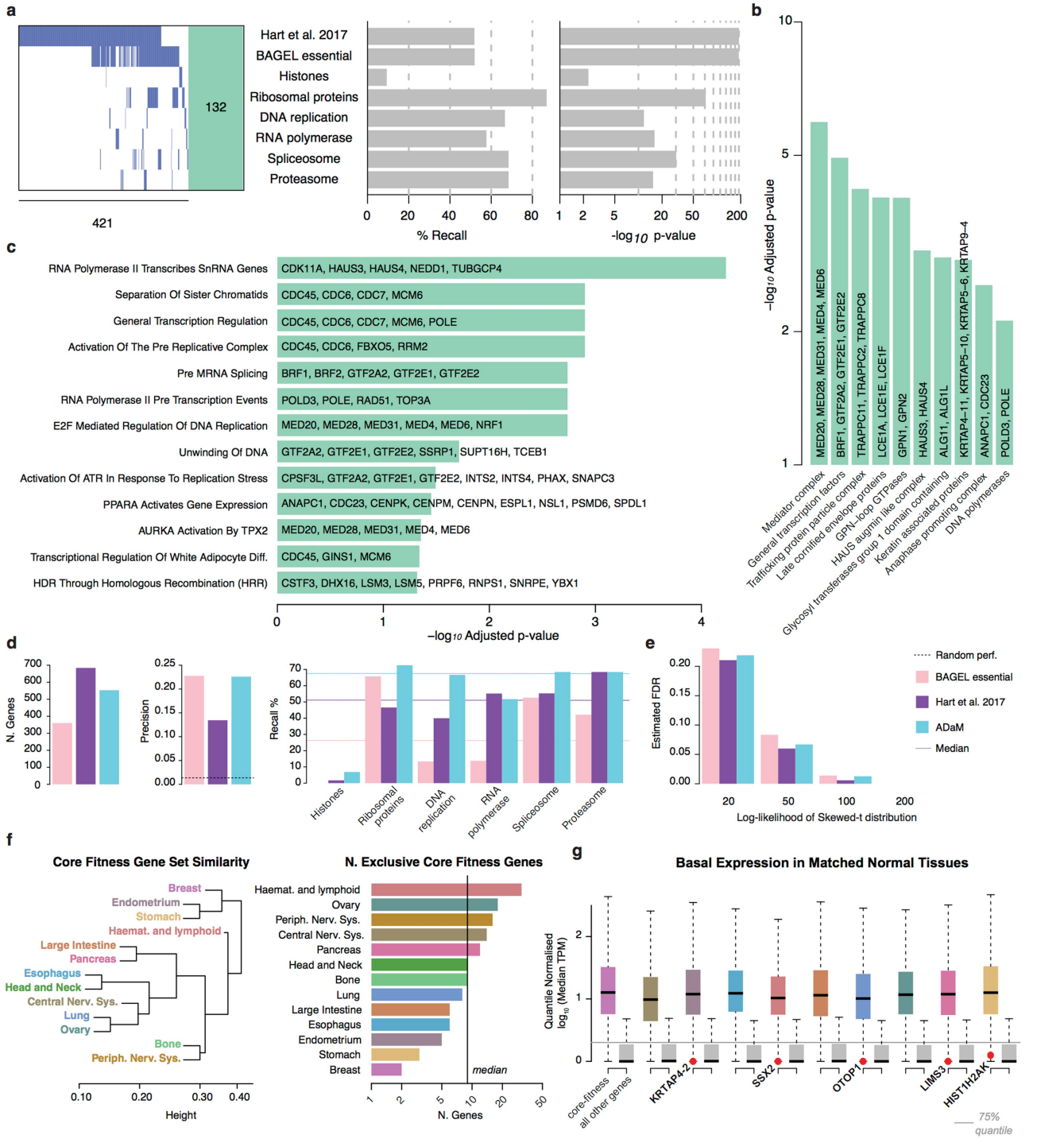
Extended Data Fig. 1 | Project Score CRISPR–Cas9 screening pipeline, data quality control and analysis set. a, CRISPR–Cas9 screening pipeline workflow, including quality control steps and go/no-go decisions. **b**, Genomic characterization of the CRISPR–Cas9-screened cell lines. **c**, Average Pearson's correlation of replicate sgRNA counts ($n = 86{,}875$) for individual cell lines. **d**, Data quality control threshold based on the distributions of Pearson's correlation values of sgRNA fold change values between replicates of the same cell line (in green) and all possible pairwise comparisons (in grey), considering the 838 highly informative sgRNAs (described in the Methods). **e**, Percentage of experiments passing the quality control filter defined in **d**. **f**, Pearson's correlation values as described in **d** for the cell lines in the final analysis set. **g**, ROC and precision/recall curves were obtained after classifying predefined essential ($n = 354$) and non-essential ($n = 747$) genes based on gene-level rank positions calculated using depletion fold changes. The median areas under the curve across all cell lines are reported. **h**, Glass's Δ scores quantifying the depletion effect size for genes that encode ribosomal proteins ($n = 61$) and a priori known essential ($n = 354$) genes for all cell lines. **i**, Cell lines in the final analysis set grouped by tissue (inner ring) and cancer-type (outer ring). **j**, Median gene-level depletion fold change (FC) values and interquartiles for reference gene sets defined in **g** and **h** for the 324 cell lines included in the analysis set. GEX, gene expression; METH, methylation; CNA, copy number alteration; WES, whole-exome DNA sequencing; AUROC, area under receiver operating characteristic; AUPR; area under precision/recall curve.



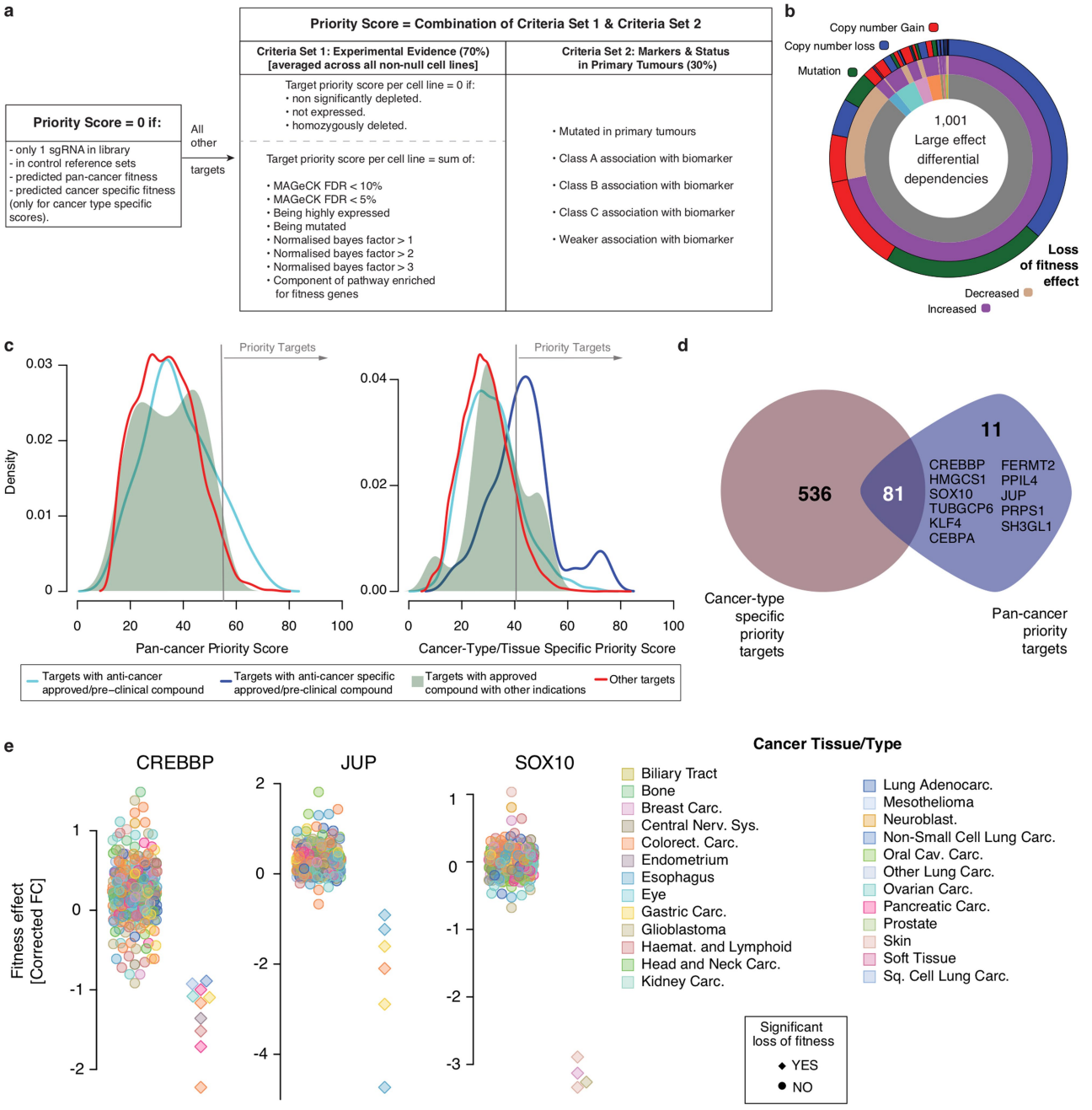
Extended Data Fig. 2 | Assessment of technical confounders in CRISPR–Cas9 screening data and summary of fitness genes. **a**, Absence of association between screening data quality and the number of replicates (as quantified by a Pearson's correlation with respect to the number of replicates, $n = 5$ distinct values). Data quality was assessed using the fitness effect (the median fold change) of genes that encode ribosomal protein ($n = 61$) in each cell line as a reference. **b**, Absence of an association between data quality (quantified as in **a**) and average Pearson's correlation between replicates of individual screened cell lines ($n = 324$). The *P* value refers to a two-sample Student's *t*-test, the score on the right plot is a Pearson's correlation. **c**, Weak correlation and significant association between sgRNA library transduction efficiency in cell lines (averaged for replicates) and data quality. **d**, Weak correlation and significant association between the Cas9 activity of a cell line (averaged for replicates) and data quality. **e**, Absence of an association between library coverage and data quality. In **c**–**e**, *P* values, *R* and sample sizes (n) are defined as for **b**. **f**, Number of fitness genes in each cell line (BAGEL FDR < 5%; median = 1,459). **g**, Number of cell lines with fixed intervals of numbers of fitness genes. **h**, Absence of correlation between number of significant fitness genes per cell line and number of replicates, *R* defined as for **a**. **i**, The effect of the version of the sgRNA screening library on the number of fitness genes identified. A new version of the library (v.1.1) with additional guides for a subset of genes yields moderately larger numbers of fitness genes; however, this is equally variable in both groups and confounded by the tissue of origin of the cell lines. *P* value is from a two-sample Student's *t*-test. **j**, Reproducible calling of fitness genes in HT-29 across sgRNA libraries. Left, the number of fitness genes detected in each library. Right, scatter plots of depletion scores at the genome-wide level or considering only highly informative sgRNAs for each library. In both cases, *P* values from a Fisher's exact test are below machine precision ($<10^{-16}$). *R* indicates Pearson's correlation; *C* indicates the percentage of genes called as significantly depleted with both libraries over those detected as significantly depleted with one library only. **k**, Pearson's correlation between the number of fitness genes per cell line and Cas9 activity level and library transduction efficiency. **l**, Pearson's correlation between the number of fitness genes per cell line and the average Pearson's correlation of cell line replicates. **m**, **n**, Pearson's correlation between the number of fitness genes per cell line and the ability to detect a defined essential genes. For all panels, each data point is a cell line coloured by cancer type (except **g** and **j**). Box-and-whisker plots show the median, interquartile range and 95 percentiles.



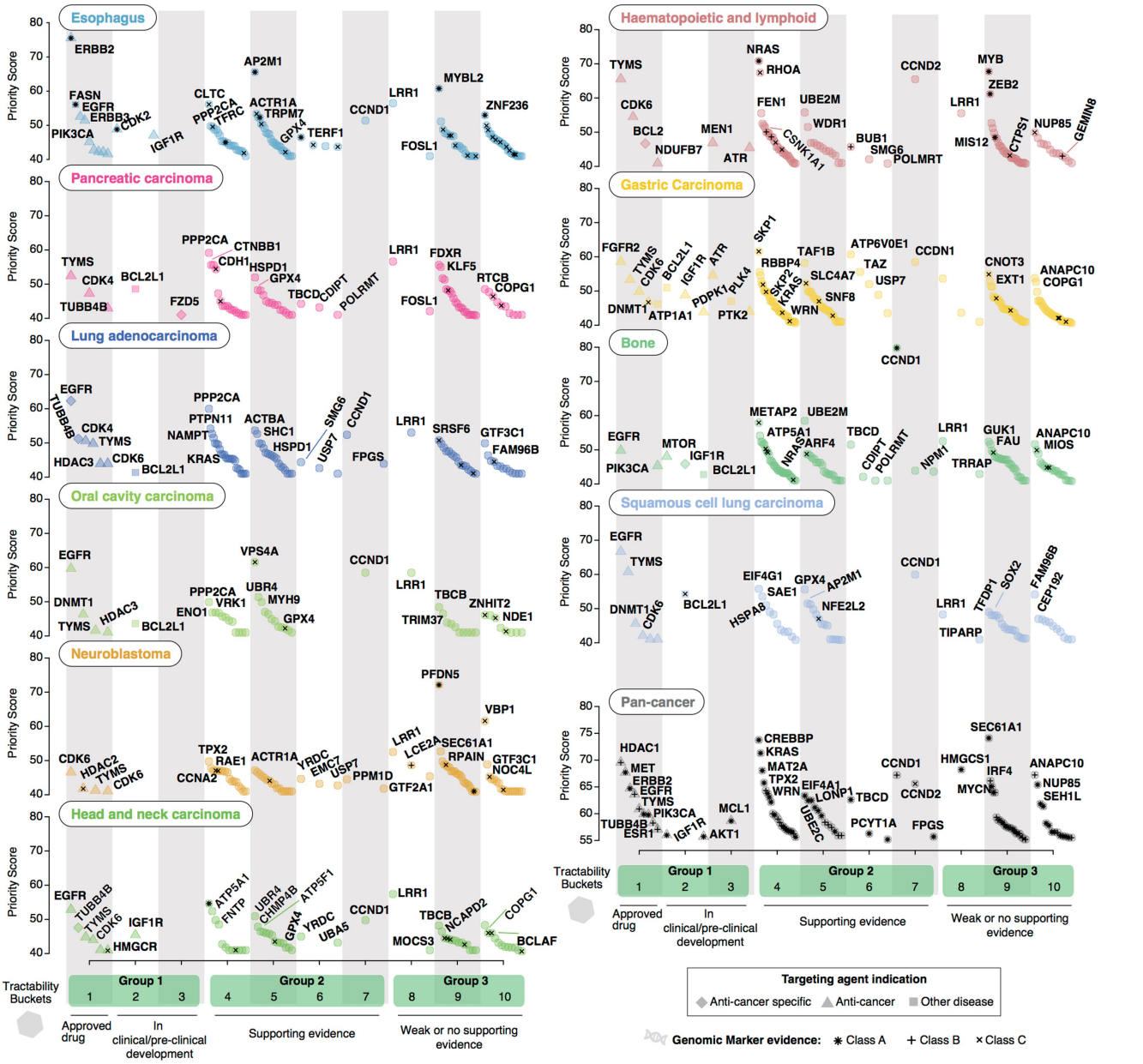
Extended Data Fig. 3 | Computation of ovary-specific and pan-cancer core fitness genes with ADaM, and a summary of context-specific and core fitness genes. **a**, Number of fitness genes in each cell line. **b**, Number of fitness genes in a fixed number (m) of cell lines. **c**, Distributions and cumulative distributions of number of fitness genes observed in m cell lines across 1,000 randomized versions of the depletion scores for ovary cell lines. **d**, True-positive rates (for which a priori known essential genes are counted as positive) when considering the genes that are depleted (fitness genes) in at least m cell lines (blue curve) as predictions and the deviance of the number of these genes from expectations (computed using the randomized data shown in **c**) for all possible values of m (red curve). The x coordinate (rounded by excess) of the intersection of these two curves estimates the minimal number of cell lines $m*$ in which a gene should be significantly depleted in order to be predicted as a core fitness gene for a cancer type. **e**, Number of genes predicted to be cancer-type-specific core fitness genes for a fixed number (k) of cancer types. **f**, Distributions (top) and cumulative distributions (bottom) of the number of core fitness genes predicted for a fixed number of tissue types for 1,000 randomized versions of the cancer-type-specific core fitness profiles. **g**, True-positive rates (for which a priori known essential genes are counted as positive) when considering the genes that are core fitness genes for at least k cancer types (blue curve) as predictions and the deviance of the number of these genes from expectation (computed using the randomized data shown in **f**; red curve). The x coordinate estimates the minimal number of cancer types $k*$ for which a gene should have been predicted as a cancer-type-specific core fitness gene in order to be classified as a pan-cancer core fitness gene. All box-and-whisker plots show the interquartile ranges and 95th percentiles, with centres indicating medians.



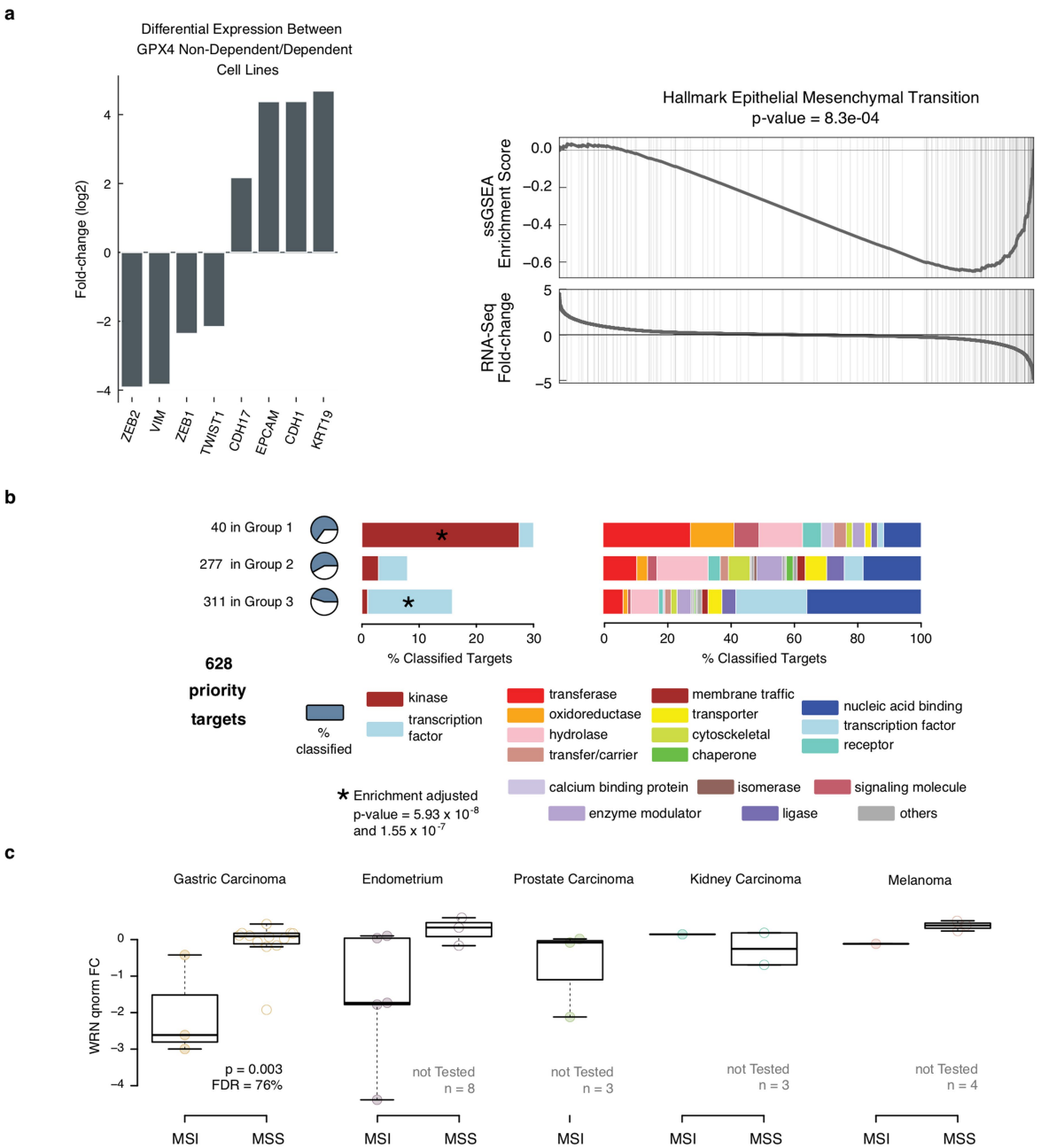
Extended Data Fig. 4 | Characterization of ADaM pan-cancer core fitness genes. a, The 553 pan-cancer core fitness genes in reference essential gene sets are shown[9,10]. Respective recall and enrichment significance *P* values from a hypergeometric test when considering the whole set of genes targeted in the CRISPR–Cas9 screen as the background population ($n = 17{,}995$). The 132 newly identified core fitness genes fall outside of these reference gene sets. **b**, **c**, Pathways (**b**) and gene families (**c**) enriched in the 132 newly identified pan-cancer core fitness genes (Benjamini–Hochberg-adjusted hypergeometric test $P < 0.05$). **d**, Comparison of the ADaM core fitness genes with two previously reported reference sets[9,10] of essential genes in terms of number of genes, estimated precision and recall (the genes included in reference gene sets corresponding to cellular essential process were considered to be true-positive genes). **e**, FDRs of putative context-specific fitness genes at different thresholds of reliability ($n = 7{,}393$, 2,233, 426 and 82 putative context-specific fitness genes, respectively, for thresholds equal to 20, 50, 100 and 200 of log-likelihood of skewed *t*-distributions). **f**, Clustering of cancer types based on core fitness gene similarity (left) and numbers of cancer-type core-specific fitness genes exclusive to each cancer type (right). **g**, Basal expression of cancer-type specific core fitness genes (*n*, across tissues indicated in Fig. 1c) in matched normal tissues compared with all the other genes in the genome, across cancer types (as indicated by the different colours). Five genes were identified as core fitness genes in a single cancer type and are not expressed at the basal level (<5% quantile) in matched normal tissue (red points). Cancer types are coloured as shown in **f**. Box-and-whisker plots show interquartile ranges and 95th percentiles, with sample sizes indicated in **f** (right), centres indicate median values.



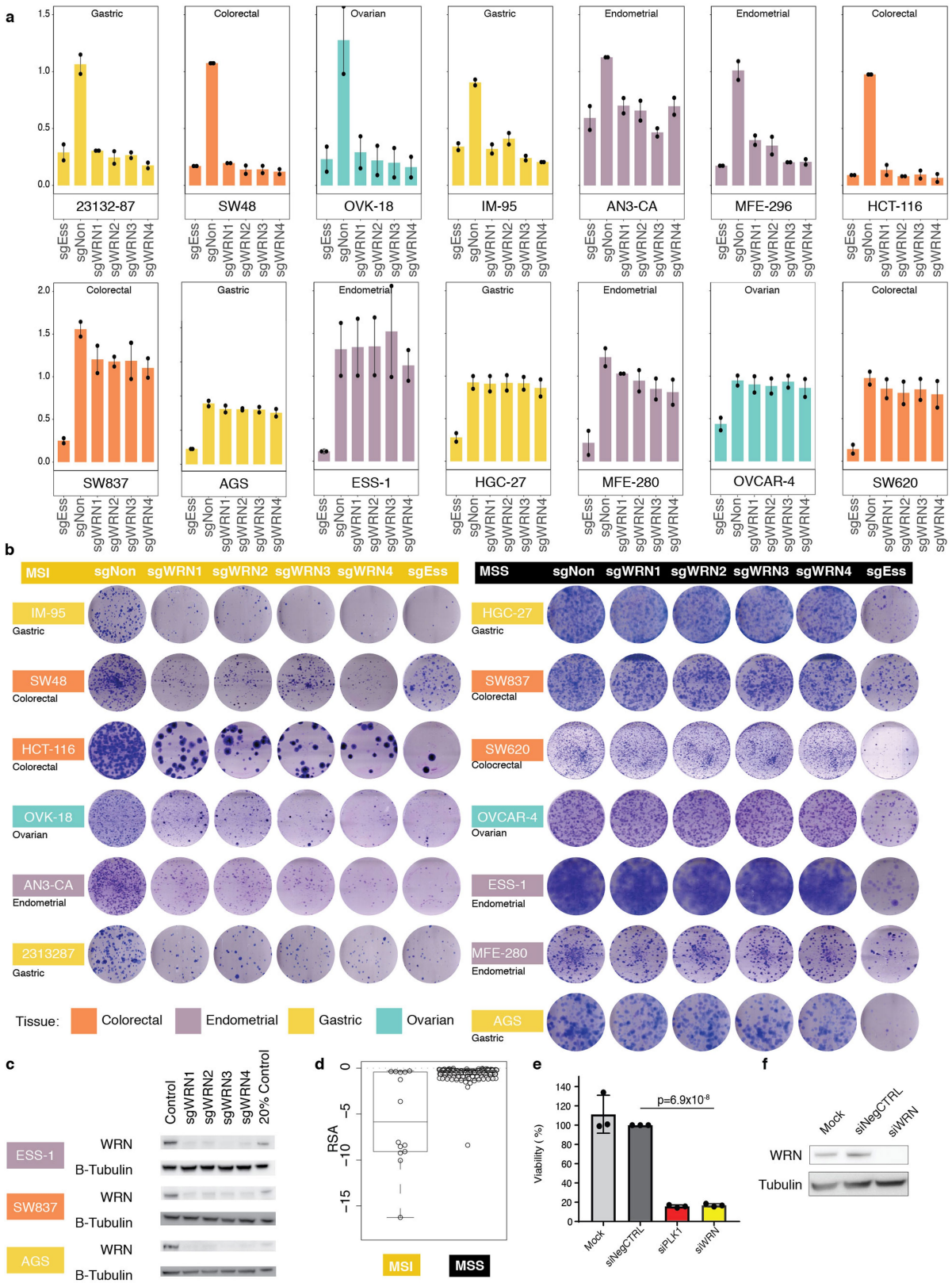
Extended Data Fig. 5 | Pan-cancer and cancer-type-specific priority scores. a, Criteria for the target prioritization scoring system. **b**, ANOVA results from differential dependency biomarker analyses with all 1,001 significant associations classified as pan-cancer or cancer-type-specific associations (inner circle), loss- or gain-of-fitness marker (middle circle) and whether the marker is a mutation, copy number gain or loss (outer circle). **c**, Distributions of pan-cancer (left) and cancer-type-specific (right) non-null target priority scores based on the therapeutic indication of approved or preclinical compounds. The significance threshold was based on the distribution of scores for targets with approved anticancer compounds (specific anticancer compounds for the cancer-type-specific priority score) versus scores for targets with no available anticancer compounds. **d**, Overlap between cancer-type-specific priority targets (for at least one cancer type) and pan-cancer priority targets. **e**, Example priority targets identified only in the pan-cancer context. Each symbol is an individual cell line coloured by cancer type and symbol shapes indicate a significant dependency ($n = 324$ cell lines).



Extended Data Fig. 6 | Priority therapeutic targets in 10 cancer types and pan-cancer. Each data point is a target with a priority score classified into tractability buckets and groups. The shapes represent the indication of the approved and/or preclinical compound to the corresponding target (other disease (square), anticancer (triangle) or specific to the cancer type considered (rhombus)); circles indicate the absence of a compound. Symbols within each data point indicate the strength of the genomic marker associated with differential dependency on the target (class A to C indicate strong to weak associations).

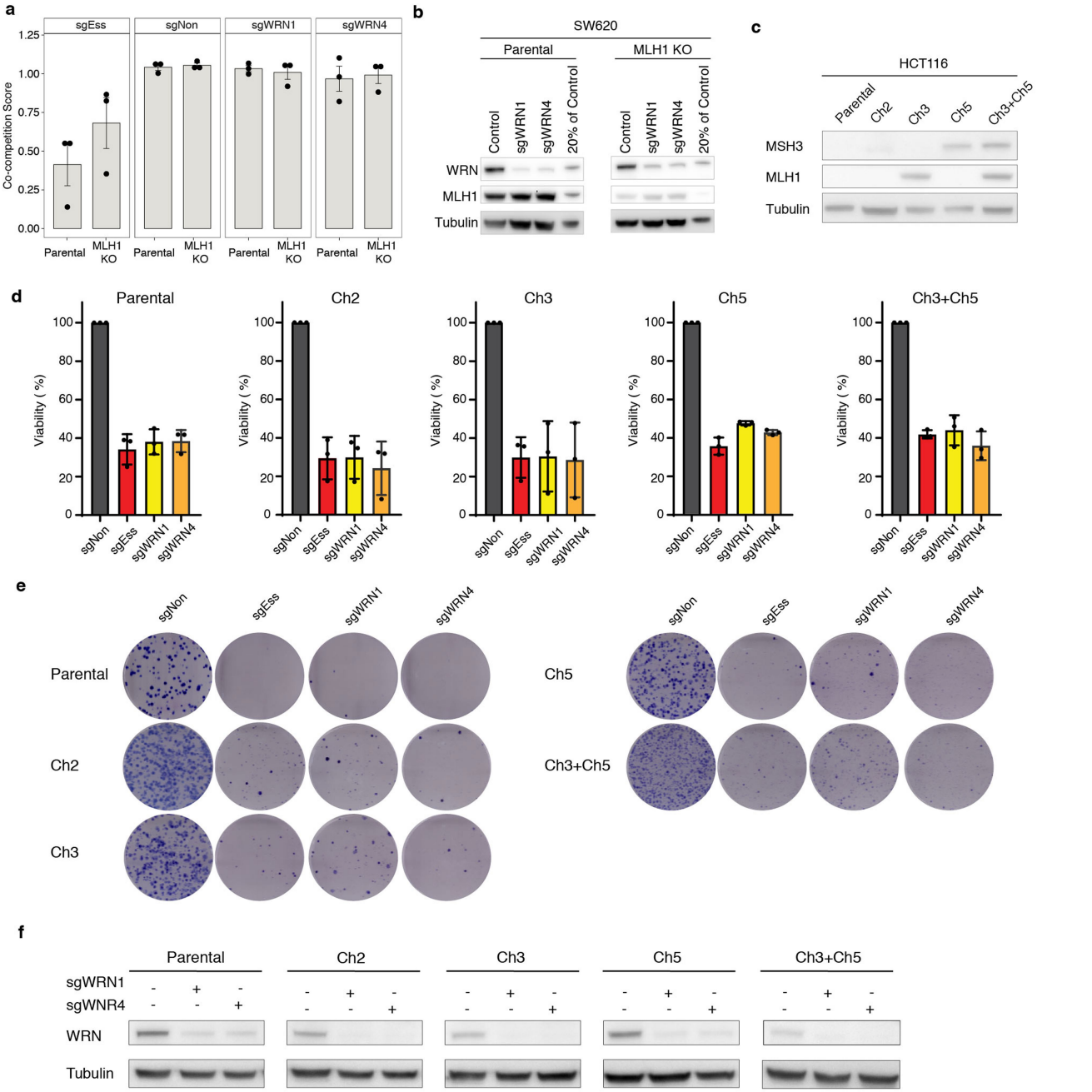


Extended Data Fig. 7 | GPX4 fitness selectivity for cells undergoing epithelial–mesenchymal transitions, functional classification of priority targets and WRN differential fitness in other cancer types. **a**, Differentially expressed genes in cell lines that are dependent on GPX4 (left) ($n = 113$, non-dependent versus dependent, moderated *t*-statistic FDR estimates). Epithelial–mesenchymal transition is the top differentially enriched cancer hallmark gene signature in *GPX4*-dependent cell lines (right). *P* values from single-sample gene set enrichment analyses were obtained by randomly permuting gene signatures 10,000 times and adjusted for multiple testing using the Benjamini–Hochberg FDR correction. **b**, Functional classification of priority targets in each tractability group using the PANTHER database. For clarity, kinases (a subset of transferases) and transcription factors are shown separately. Protein classes are indicated by colour. Statistical enrichment was calculated using a systematic hypergeometric test across protein families, following correction for multiple testing with the Benjamini–Hochberg method. Pie charts indicate the percentage of targets in each group classified according to protein families. **c**, WRN dependency in multiple cancer types. Each data point is a cell line showing the quantile-normalized WRN sgRNA fold change value stratified by MSI status. Box-and-whisker plots show interquartile ranges and 95th percentiles and centres indicate median values. Individual values are shown as dots. Statistical significance was calculated from the systematic ANOVA analysis for each cancer type for which the number of cell lines was greater than 10 ($n = 14$ for gastric carcinoma).

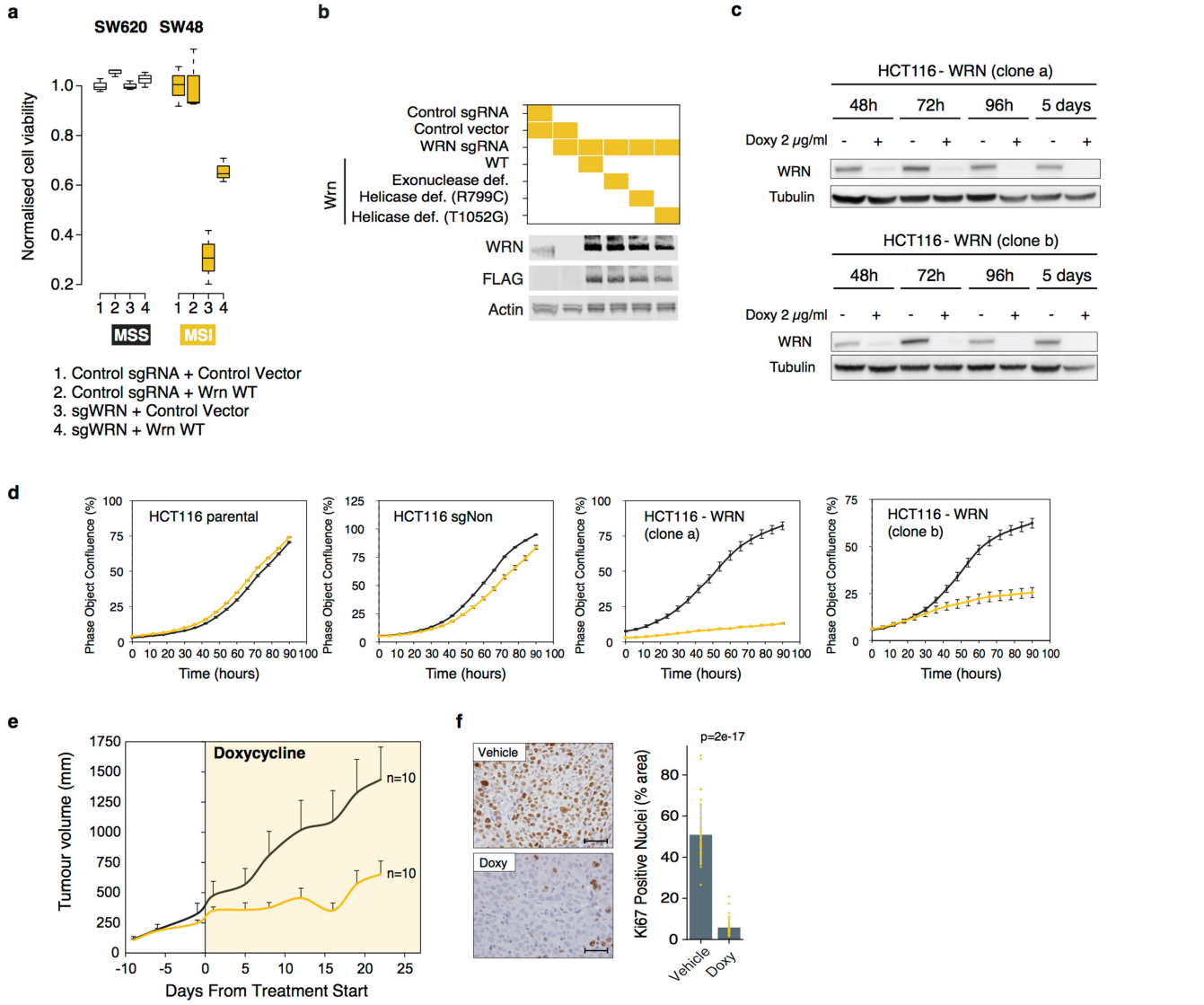


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Verification of WRN as a target in MSI cancers. **a**, *WRN* dependency using a co-competition assay in MSI (top row, $n = 7$) and MSS (bottom row, $n = 7$) cell lines from four cancer types. sgRNAs targeting essential (sgEss) and non-essential (sgNon) genes were used as controls. Bars represent mean co-competition score; lines represent maximum and minimum values; individual data points overlaid. **b**, Selective WRN dependency in MSI versus MSS cell lines was confirmed using clonogenic assays in four cancer types (images are representative of two independent experiments). **c**, A reduction in WRN protein levels with all *WRN* sgRNAs was confirmed by western blot (images are representative of two independent experiments). **d**, An association between WRN dependency and MSI status was confirmed by mining data from an independent study that used RNA interference, project DRIVE[12] (Student's *t*-test, $P = 0.004$; $n = 214$). Each circle represents the WRN RNA-interference dependency score in a cancer cell line. Box-and-whisker plots represent median and 1.5× interquartile range. **e**, siRNA depletion of *WRN* inhibited proliferation of HCT116 cells. Data are mean ± s.d. of three independent experiments. The *P* value was determined using a non-parametric Student's *t*-test. **f**, siRNA-mediated depletion of *WRN* was verified by western blot (images are representative of two independent experiments). For western blot source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | *MLH1* knockout, MMR rescue experiments and modulation of WRN dependency. a, A WRN co-competition assay in MSS SW620 cells with stable *MLH1* knockout. Cells were cultured for 3 months before assessing WRN dependency. Data are mean ± s.e.m. of three independent experiments. **b**, Western blotting confirmed *MLH1* and *WRN* knockout (images are representative of two independent experiments). **c**, MLH1 and MSH3 expression by western blot in HCT116 parental and isogenic cell lines complemented with chromosome 2 (Ch.2; negative control), Ch.3 (which contains *MLH1*), Ch.5 (which contains *MSH3*) and Ch.3 + Ch.5 (which contains both *MLH1* and *MSH3*). Data are representative of two independent experiments. **d**, Effect of *WRN* knockout (*WRN* sgRNAs 1 and 4 (sgWRN1 and sgWRN4, respectively)) on viability after 7 days in HCT116 parental and isogenic cell lines. Data are mean ± s.d. of three independent experiments. **e**, Clonogenic assays (14 days) after *WRN* knockout in HCT116 parental and isogenic cell lines. Data are representative of three independent experiments. **f**, Reduction in WRN levels was confirmed by western blot. Data are representative of two independent experiments. Source data for all western blots are shown in Supplementary Fig. 1.



Extended Data Fig. 10 | Functional rescue experiments and in vivo validation of WRN dependency in a MSI colorectal cancer cell line. **a**, Expression of wild-type mouse *Wrn* rescued the viability effect of *WRN* knockout in MSI cell line SW48. MSS cell line SW620 was used a negative control. Box-and-whisker plots represent the median and 1.5× interquartile range. Data represent two independent biological replicates completed in technical triplicate. **b**, Western blots confirmed expression of Flag-tagged protein using all variants of the *Wrn* vector. Images are representative of experiments performed in triplicate. **c**, *WRN* knockout induced by doxycycline treatment in *WRN* sgRNA-expressing HCT116 (HCT116-WRN) cells measured by western blot for two separate clonal lines. Data are representative of two independent experiments. **d**, Growth curves of HCT116 parental, HCT116 sgNon (non-essential sgRNA) and *WRN* sgRNA-expressing HCT116 cells grown in the absence (black line) or presence of doxycycline (2 μg ml^{-1}; yellow line). Data are mean ± s.d. of 10 technical replicate wells for each condition (1 image per well) and representative of two independent experiments. **e**, Growth curves of *WRN* sgRNA-expressing HCT116 (clone b) subcutaneous tumours from mice treated with doxycycline (50 mg kg^{-1}; yellow line) or vehicle (grey line). Tumour growth suppression was observed ($P = 0.03$, two-way ANOVA comparing doxycycline versus vehicle). The number of mice in each cohort is indicated. Data are mean ± s.e.m. **f**, Representative KI-67 immunohistochemistry assessment of *WRN* sgRNA-expressing HCT116 (clone b) tumours explanted after one week of doxycycline treatment (left). Scale bar, 50 μm; 40× magnification. Quantification of KI-67 staining (right). Data are mean ± s.d. of 10 fields from three different samples ($n = 30$) and means were compared using a two-sided Welch's *t*-test. Source data for all western blots are shown in Supplementary Fig. 1.

nature research

Corresponding author(s): Mathew J Garnett

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	no software used
Data analysis	Github URLs to the repositories containing source code and documentation of all the software used in this manuscript are specified in the supplementary material. In addition all the software is publicly available together with detailed description and further documented at https://depmap.sanger.ac.uk/programmes#analytics.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Due to their large size, the raw and processed datasets presented in our manuscript are available for the editorial office and the reviewers as encrypted compressed

files at the following URLs:
Gene essentiality matrices
Link: https://cog.sanger.ac.uk/cmp/download/essentiality_matrices.zip

Raw sgRNA counts
Link: https://cog.sanger.ac.uk/cmp/download/raw_sgrnas_counts.zip
[Passwords required to decrypt these zipped folder are included in the cover letter we enclosed to the Revised version of our submission].

Upon paper publication not encrypted files will replace their encrypted versions.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	not applicable
Data exclusions	not applicable
Replication	not applicable
Randomization	not applicable
Blinding	not applicable

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

WRN (8H3) Mouse mAb #4666; Cell Signalling Technologies
MLH1 mouse mAb #3515; Cell Signalling Technologies
MSH3 mouse mAb #sc-271080; Santa Cruz Biotechnology
WRN rabbit polyclonal #PA5-27319; Thermo Fisher
Anti-flag M2 rat mAb #F3165; Sigma Aldrich
B-actin rabbit mAb #4970; Cell Signalling Technologies
IRDye® 800CW Donkey-anti-Mouse Antibody #926-32212; LI-COR
IRDye® 680LT Donkey anti-Rabbit IgG (H + L) #925-68023; Li-COR
Anti-Mouse IGG HRP linked secondary antibody #NA931; GE Healthcare

Validation

Manufacturers Statement:

WRN (8H3) Mouse mAb detects endogenous levels of total WRN protein. Species Reactivity: Human, Mouse. Product Citations: Pal, D., Pertot, A., et al. (2017), 'TGF-β reduces DNA ds-break repair mechanisms to heighten genetic diversity and adaptability of CD44+/CD24- cancer cells.', Elife; Li, K., Wang, R., et al. (2010), 'Acetylation of WRN protein regulates its stability by inhibiting

ubiquitination.', PLoS One, 5 (4), pp. e10341; Shiratori, M., Sakamoto, S., et al. (1999), 'Detection by epitope-defined monoclonal antibodies of Werner DNA helicases in the nucleoplasm and their upregulation by cell transformation and immortalization.', J Cell Biol, 144 (1), pp. 1-9

MLH1 (4C9C7) Mouse mAb detects endogenous levels of total MLH1 protein. Species Reactivity: Human, Monkey. Product Citations: Yan, J., Shun, M. C., et al. (2018), 'HIV-1 Vpr Reprograms CLR4DCAF1 E3 Ubiquitin Ligase to Antagonize Exonuclease 1-Mediated Restriction of HIV-1 Infection.', MBio, 9 (5); Kashyap, T., Argueta, C., et al. (2018), 'Selinexor reduces the expression of DNA damage repair proteins and sensitizes cancer cells to DNA damaging agents.', Oncotarget, 9 (56), pp. 30773-30786

MSH3 (B-4) is a mouse monoclonal antibody raised against amino acids 61-360 of MSH3 of human origin. Species Reactivity: Human, Mouse and rat. Product citations: Germini, D.E., et al. 2016. Detection of DNA repair protein in colorectal cancer of patients up to 50 years old can increase the identification of Lynch syndrome? Tumour Biol. 37: 2757-2764.

β-Actin (13E5) Rabbit mAb #4970 detects endogenous levels of total β-actin protein. This antibody may cross-react with the γ-actin (cytoplasmic isoform). It does not cross-react with α-skeletal, α-cardiac, α-vascular smooth, or γ-enteric smooth muscle isoforms. Species Reactivity: Human, Mouse, Rat, Monkey, Bovine, Pig. CPT1A-mediated succinylation of S100A10 increases human gastric cancer invasion. In Journal of Cellular and Molecular Medicine on 1 January 2019 by Wang, C., Zhang, C., et al. Cytotoxic phenanthroline derivatives alter metallostasis and redox homeostasis in neuroblastoma cells. In Oncotarget on 20 November 2018 by Naletova, I., Satriano, C., et al. We did not perform additional validation of this antibody.

Monoclonal ANTI-FLAG® M2 antibody. Anti Flag M2 antibody is used for the detection of Flag fusion proteins.This monoclonal antibody is produced in mouse and recognizes the FLAG sequence at the N-terminus, Met N-terminus, and C-terminus. The antibody is also able to recognize FLAG at an internal site. The GOLD domain-containing protein TMED7 inhibits TLR4 signalling from the endosome upon LPS stimulation. Sarah L Doyle et. al Nature communications, 3, undefined (2012-3-20). Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors. Johannes Popow et. al. Nature, 511(7507), undefined (2014-5-30). We used untransfected control cell lysate to demonstrate antibody specificity against FLAG tagged proteins.

Monoclonal Anti-α-Tubulin (mouse IgG1 isotype) is derived from the B-5-1-2 hybridoma produced by the fusion of mouse myeloma cells and splenocytes from an immunized mouse. Tubulin is the major building block of microtubules. Species reactivity human, Chlamydomonas, African green monkey, chicken, kangaroo rat, bovine, mouse, rat, sea urchin. Tubulin acetyltransferase aTAT1 destabilizes microtubules independently of its acetylation activity. Kalebic N, et.al. Molecular and Cellular Biology 33(6), 1114-1123, (2013). Increased expression of αTubulin is associated with poor prognosis in patients with pancreatic cancer after surgical resection Lin C, et al. Oncotarget 7(37), 60657-60657, (2016). We did not perform additional validation of the antibody.

IRDye® 800CW Donkey-anti-Mouse Antibody. The antibody was isolated by affinity chromatography using antigens coupled to agarose beads. Based on ELISA, this antibody reacts with the heavy and light chains of mouse IgG and with the light chains of mouse IgM and IgA. This antibody was tested by ELISA and/or solid-phase absorbed to ensure minimal cross-reaction with bovine, chicken, goat, guinea pig, horse, human, rabbit, and sheep serum proteins, but the antibody may cross-react with immune-globulins from other species. The conjugates has been specifically tested and qualified for western blot and in-cell western assay applications. We did not perform additional validations of this secondary antibody.

IRDye® 680LT Donkey anti-Rabbit IgG (H + L). The antibody was isolated from antisera by immunoaffinity chromatography using antigens coupled to agarose beads. Based on immunoelectrophoresis, the antibody reacts with the heavy chains on rabbit IgG and with the light chains common to most rabbit immunoglobulins. No reactivity was detected against non-immunoglobulin serum proteins. This antibody has been tested by ELISA and/or solid-phase adsorbed to ensure minimal cross-reaction with bovine, chicken, goat, guinea pig, Syrian hamster, horse, human, mouse, rat and sheep serum proteins, but the antibody may cross-react with immunoglobulins from other species. The conjugates has been specifically tested and qualified for western blot applications. We did not perform additional validations on this secondary antibody.

Anti-Mouse IGG HRP linked secondary antibody #NA931; GE Healthcare
Highly species-specific; optimized for use with our range of ECL™ Detection Reagents; recommended dilutions to minimize background.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

All cell lines (324) are part of the Cell Model Passport Collection (https://cellmodelpassports.sanger.ac.uk/).

ATCC: hTERT-RPE-1, MDA-MB-468, MDA-MB-436, MDA-MB-415, NCI-H1355, NCI-H1299, NCI-H1155, M059J, LS-411N, Hs-746T, Hs-683, HPAF-II, HCC70, HCC38, LS-513, LS-180, LS-1034, LNCaP-Clone-FGC, LN-229, KLE, KATOIII, JHU-029, NCI-H2087, NCI-H2170, NCI-H358, NCI-H2405, NCI-H1755, NCI-H1650, NCI-H1568, NCI-H2023, NCI-H1993, NCI-H1975, NCI-H1944, NCI-H1915, NCI-H1869, LS-123, T84, A375, A253, A2058, A172, 769-P, UWB1.289, Detroit562, SW837, SW48, SW620, COLO-320-HSR, AGS, FADU, ES-2, HCC1143, DBTRG-05MG, HCC1954, HCC1937, HCC1806, HCC1395, HCC1187, AU565, AsPC-1, ARH-77, CHP-212, Caov-4, SCC-9, SCC-4, RL95-2, RKO, SK-PN-DW, SK-N-SH, SK-N-FI, SK-N-DZ, SK-N-AS, SK-MES-1, SJSA-1, NCI-H650, NCI-H520, PANC-10-05, PANC-08-13, PANC-04-03, PANC-03-27, PANC-02-03, OV-90, SNU-16, SNU-1, NCI-N87, SW1088, SW1573, SW1990, SW626, TOV-21G, TOV-112D, T98G, UACC-893, SU8686, U-87-MG, SNU-C1, LNZTA3WT4, BE2-M17, C2BBe1, MC-IXC, MM1S.

Cancer Science Institute of Singapore: OVCA420, Hey, DOV13.

CLS: RCC-FG2.

DSMZ: L-363, MFM-223, MFE-296, MFE-280, LXF-289, LP-1, LOU-NH91, LCLC-97TM1, MHH-ES-1, SU-DHL-10, SCC90, SU-DHL-5, ROS-50, TC-71, SU-DHL-8, OACM5-1, OCI-AML2, PA-TU-8988T, PA-TU-8902, OPM-2, OCI-LY-19, OCI-AML3, 22RV1, CAL-72, CAL-51, CAPAN-1, CAL-27, CAL-33, CL-11, DK-MG, DAN-G, COLO-824, COLO-680N, COLO-678, 23132-87, 42-MG-BA, BHY, 8-MG-BA, KYSE-70, KYSE-520, KYSE-510, KYSE-450, KYSE-410, KYSE-270, KYSE-150, KYSE-140, JIMT-1, FLO-1, EVSA-T, ESS-1, EPLC-272H, EGI-1, EFO-27, EFO-21, GAMG, HCC-78, HCC-15.

Duke University Medical Center: D-542MG, D-502MG, D-423MG, D-247MG.

ECACC: BICR10, BxPC-3, COLO-684, PSN1, PEO4, PEO1, PE-CA-PJ15, OV-56, OE33, OE21, OAW-42, SK-GT-4, BICR22, BICR78, MOG-G-UVW, MIA-PaCa-2, COR-L23, MDST8, MDA-MB-361, KYAE-1, HuP-T4, HuP-T3, GP5d, ESO51, ESO26, DOK, HT55, COLO-205, A2780ADR, A2780cis.

IARC: EW-22, EW-16, EW-7, EW-1.

Unknown source: DiFi, HSC-39, PCI-30, PCI-4B, TMK-1,PL4.

ICLC: A2780, CAS-1, OC-314, IST-MEL1, GI-ME-N.

JCRB: MCAS, YH-13, VMRC-LCD, TYK-nu, T-T, TGW, TE-9, TE-8, TE-5, TE-10, SAS, RMG-I, RKN, RERF-LC-Sq1, RERF-GC-1B, RCM-1, SUIT-2, SNG-M, SKN-3, SK-MG-1, SF126, SCH, OVMIU, EBC-1, Ca9-22, Becker, AM-38, LU-65, LK-2, KYSE-220, KURAMOCHI, KS-1, KP-N-YN, OVISE, OSC-20, OSC-19, KP-3, NUGC-3, no-11, no-10, NMC-G1, NH-12, GB-1, KP-1N, HSC-4, HSC-3, HO-1-u-1, HEC-1, HARA, IM-95, KON, KNS-62, KNS-42, KMS-11, KINGS-1, IM-9, LoVo, LU-99A, KCLB, SNU-81, SNU-61, SNU-C5.

Kyoto Prefectural University of Medicine: KP-N-YS

Ludwig Institute for Cancer Research, Brussels branch: BB30-HNC, LB771-HNC, LB1047-RCC.

Massachusetts General Hospital: JHU-022, JHU-011.

NCI: M14, SK-MEL-2, SF539, SF295, SF268, RPMI-8226, OVCAR-8, OVCAR-5, OVCAR-4, SNB75, T47D, OVCAR-3, U251, NCI-H322M, HCC2998, DU-145, HOP-62, Hs-578-T, HT-29, NCI-H3122, NCI-H23, NCI-H226, MDA-MB-231, MCF7, KM12, IGROV-1, HCT-116, HCT-15, A549.

NCI-Navy Medical Oncology Branch: NCI-H3118.

RIKEN: PC-14, OVK-18, OCUB-M, NB69, TE-15, TE-4, TGBC11TKB, MKN28, MKN1, MDA-MB-453, GI-1, GCIY, EC-GI-10, HGC-27, LC-1-sq, KP-4, JHOS-4, JHOS-2.

St Jude Children's Research Hospital: NB7, NB6, NB5, NB17, NB13, NB10, ES8, ES5, ES4.

The University of Hong Kong: PCI-6A, PCI-38, PCI-15A.

University of Pennsylvania Health System: HCE-4.

University of Michigan: HCT116 Parental, HCT116-Ch3, HCT116-Ch5 and HCT116-Ch3+Ch5.

Baylor Charles A. Sammons Cancer Center: HCT116-Ch2.

Authentication

Each of the cell lines have been tested using a panel of 16 STRs (AmpFLSTR Identifiler KIT, ABI), which includes the 9 currently used by most of the cell line repositories (ATCC, Riken, JCRB and DSMZ).

Mycoplasma contamination

All cell lines were tested for mycoplasma contamination and only cell lines that were negative were included in the study.

Commonly misidentified lines
(See ICLAC register)

This study includes the following commonly misidentified lines:

Ca9-22 - STR matches JCRB reference (JCRB0625) & RIKEN (RCB1976).

MKN28 - noted as derivative of MKN74 in Cell Model Passports (https://cellmodelpassports.sanger.ac.uk/passports/SIDM00260); Clinical information matches MKN74.

KP-1N - known misidentification issue; Cell Model Passports data for both KP-1N & Panc-1 identical (https://cellmodelpassports.sanger.ac.uk/passports/SIDM00583).

OVMIU - known misidentification issue; Cell Model Passports data for both OVMIU and OVSAYO are identical (https://cellmodelpassports.sanger.ac.uk/passports/SIDM00465).

SK-MG-1 - STR profile matches JCRB profile which internally matches to Marcus; Cell Model Passports data for both SK-MG-1 and Marcus are identical; removed from core collection set.

Misidentified lines have been noted in Supplementary Table 1 and on the Cell Model Passport (https://cellmodelpassports.sanger.ac.uk). Mis-identification does not impact tissue of origin or genomic data used for analyses.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mus musculus, non-obese diabetic/severe combined immunodeficient (NOD-SCID), female, 5- to 6- week-old.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the fields.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cell lines were fixed with 4% formaldehyde & resuspended in PBS before running through analyzers.
Instrument	Becton Dickinson LSRFortessa flow analyser
Software	FlowJo
Cell population abundance	Cells were not sorted; only analyzed.
Gating strategy	Gates were set for each cell line based on a negative control sample that had no treatment.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.