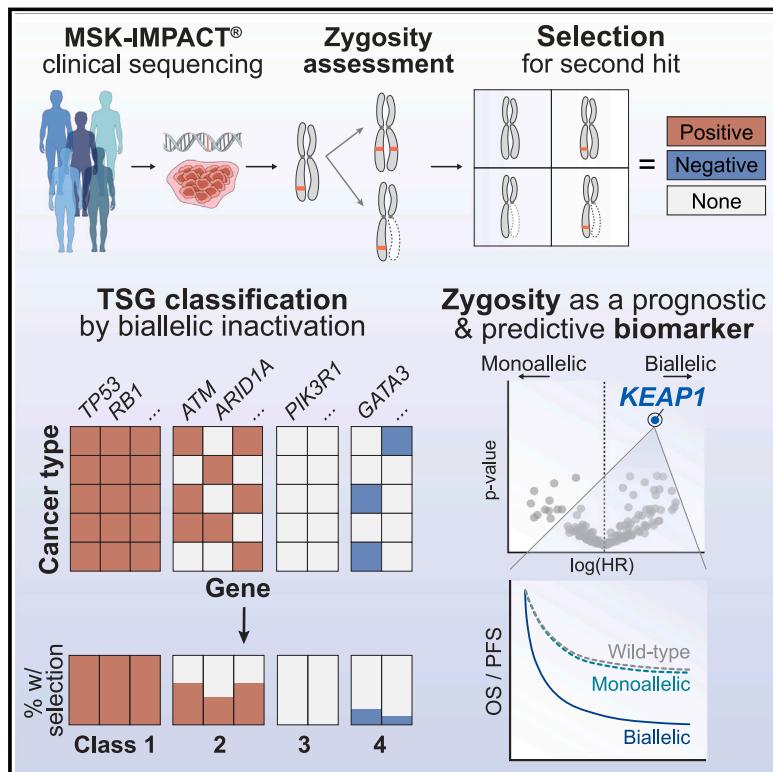


Pan-cancer analysis of biallelic inactivation in tumor suppressor genes identifies *KEAP1* zygosity as a predictive biomarker in lung cancer

Graphical abstract



Authors

Mark Zucker, Maria A. Perry,
Samuel I. Gould, ...,
Francisco J. Sánchez-Rivera, Ed Reznik,
Chaitanya Bandlamudi

Correspondence

reznike@mskcc.org (E.R.),
bandlamc@mskcc.org (C.B.)

In brief

Selective pressure for biallelic inactivation varies widely across tumor suppressor genes and is a biomarker for therapeutic response, providing new insights into Knudson's two-hit hypothesis.

Highlights

- Selection for biallelic inactivation varies widely across mutated genes and lineages
- TSGs can be classified by the frequency of selection for a second hit across lineages
- Selection for second hit reclassifies VUSs and highlights rarely mutated TSGs
- KEAP1* zygosity is a predictive biomarker for standard-of-care therapies in lung cancers



Resource

Pan-cancer analysis of biallelic inactivation in tumor suppressor genes identifies KEAP1 zygosity as a predictive biomarker in lung cancer

Mark Zucker,^{1,15} Maria A. Perry,^{2,3,15} Samuel I. Gould,^{4,5} Arielle Elkrief,^{3,6,7} Anton Safonov,⁶ Rohit Thummalaapalli,⁶ Miika Mehine,³ Debyani Chakravarty,^{2,3} A. Rose Brannon,³ Marc Ladanyi,³ Pedram Razavi,⁶ Mark T.A. Donoghue,² Yonina R. Murciano-Goroff,⁶ Kristiana Grigoriadis,^{8,9,10} Nicholas McGranahan,^{9,10} Mariam Jamal-Hanjani,^{9,11,12} Charles Swanton,^{8,9,12} Yuan Chen,¹³ Ronglai Shen,¹³ Sarat Chandrarapthy,^{6,7} David B. Solit,^{2,6} Nikolaus Schultz,^{1,2,7,14} Michael F. Berger,^{1,2,3,7,14} Jason Chang,³ Adam J. Schoenfeld,⁶ Francisco J. Sánchez-Rivera,^{4,5} Ed Reznik,^{1,2,*} and Chaitanya Bandlamudi^{2,3,16,*}

¹Computational Oncology Service, Memorial Sloan Kettering Cancer Center, New York, NY, USA

²Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

³Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

⁵David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

⁶Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁷Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁸Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK

⁹Cancer Research UK Lung Cancer Centre of Excellence, University College London, London, UK

¹⁰Cancer Genome Evolution Research Group, University College London Cancer Institute, London, UK

¹¹Cancer Metastasis Laboratory, University College London Cancer Institute, London, UK

¹²Department of Medical Oncology, University College London Hospitals, London, UK

¹³Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

¹⁴Weill Cornell Medical College, New York, NY, USA

¹⁵These authors contributed equally

¹⁶Lead contact

*Correspondence: reznike@mskcc.org (E.R.), bandlamc@mskcc.org (C.B.)

<https://doi.org/10.1016/j.cell.2024.11.010>

SUMMARY

The canonical model of tumor suppressor gene (TSG)-mediated oncogenesis posits that loss of both alleles is necessary for inactivation. Here, through allele-specific analysis of sequencing data from 48,179 cancer patients, we define the prevalence, selective pressure for, and functional consequences of biallelic inactivation across TSGs. TSGs largely assort into distinct classes associated with either pan-cancer (Class 1) or lineage-specific (Class 2) patterns of selection for biallelic loss, although some TSGs are predominantly mono-allelically inactivated (Class 3/4). We demonstrate that selection for biallelic inactivation can be utilized to identify driver genes in non-canonical contexts, including among variants of unknown significance (VUSs) of several TSGs such as *KEAP1*. Genomic, functional, and clinical data collectively indicate that *KEAP1* VUSs phenocopy established *KEAP1* oncogenic alleles and that zygosity, rather than variant classification, is predictive of therapeutic response. TSG zygosity is therefore a fundamental determinant of disease etiology and therapeutic sensitivity.

INTRODUCTION

Tumorigenesis is characterized by sequential acquisition of somatic mutations and copy-number alterations to one or both alleles of oncogenes and tumor suppressor genes (TSGs).¹ The classical, “two-hit” model of TSG inactivation posits that loss of both alleles (biallelic loss) is necessary for inactivation and subsequent tumor initiation.² Many autosomal recessive tumor suppressors have been discovered that exhibit near-ubiquitous biallelic losses in specific cancer types (e.g., *RB1* in retino-

blastoma,³ *APC* in colorectal cancer,⁴ and *VHL* in clear cell renal cell carcinoma⁵). More broadly, somatic loss of the wild-type (WT) allele has been shown to be a hallmark of inherited pathogenic variants in high- and moderate-penetrance cancer predisposition TSGs.^{6,7} In contrast, somatic loss of a single copy of a TSG (monoallelic loss) is also sufficient to attenuate tumor suppression and drive oncogenesis in specific contexts, such as haploinsufficiency in *PTEN* and dominant negativity in *TP53*.^{8,9} More broadly, a continuum model of tumor suppression, which postulates that partial loss of TSG activity can promote



oncogenic effects, has been suggested to be applicable to many TSGs.^{10,11} However, because the majority of large-scale cancer genomics studies have utilized non-allele-specific copy-number analysis methods, neither the extent of biallelic inactivation across TSGs (beyond specific cases such as *TP53*¹²) nor the functional and translational consequences of biallelic inactivation are well understood.

We reasoned that a comprehensive allele-specific analysis of TSG alterations in a large cohort of prospectively sequenced cancer patients could lead to insights into disease etiology and the role TSGs play in mediating therapeutic responses. We therefore explored the frequency of biallelic inactivation across 224 TSGs in 48,179 cancer patients. Matched tumor and normal sequencing together with deep sequencing coverage enabled robust inference of allele-specific copy number and its co-occurrence with somatic mutations. We show that the frequency of biallelic inactivation for most major TSGs varies dramatically across cancer types, with most TSGs demonstrating lineage-specific patterns of enrichment for biallelic inactivation. By modeling the selective pressure for the most common form of biallelic inactivation (mutation plus loss of WT allele via copy-number loss of heterozygosity [LOH], “MutLOH”), we identified rare but highly selected loss of *APC* in lung and prostate adenocarcinomas. Similarly, by investigating the selective pressure for MutLOH in variants of unknown significance (VUSs), we discovered that *KEAP1* VUSs in lung adenocarcinoma (LUAD) are strongly enriched for biallelic alterations and phenocopy well-established *KEAP1* oncogenic alleles. Consequently, we observe that *KEAP1* zygosity, rather than annotated oncogenic status, is correlated to overall survival and predictive of response to multiple standard-of-care therapies.

RESULTS

Characterizing somatic biallelic inactivation in TSGs

To study the incidence and selection of biallelic alterations across cancer-associated genes, we analyzed tumor profiles of 48,179 patients across 67 major cancer subtypes using the FDA-authorized MSK-IMPACT targeted clinical sequencing assay. MSK-IMPACT profiles up to 505 cancer-associated genes, including 224 genes annotated as tumor suppressors by the precision oncology knowledgebase, OncoKB¹³ (see STAR Methods). To maximize sensitivity and specificity to infer allele-specific copy-number alterations, we restricted our analysis to 23,713 tumors with at least 30% purity and excluded tumors with high tumor mutation or copy-number alteration burden (see STAR Methods; Table S1A). Somatic alterations comprising substitutions, indels, gene-level copy-number amplifications and homozygous deletions, and fusions were identified using a clinically validated pipeline¹⁴ and annotated for oncogenicity by OncoKB¹³ (STAR Methods; Tables S1B–S1D). Hereafter, all somatic alterations, unless otherwise specified, refer to OncoKB-annotated oncogenic alterations. For each somatic alteration, we inferred the local zygosity state and ascribed biallelic status for those resulting in complete loss of WT, which arise from (1) homozygous deletions, (2) oncogenic mutation with a concomitant copy-number loss on the complementary allele that resulted in a complete loss of WT (MutLOH), (3) oncogenic substitutions, insertions, or

deletions together with a genomic rearrangement event, and (4) multiple oncogenic mutations in the same gene, referred to as “composite mutations”¹⁵ (Figure 1A; see STAR Methods for further details).

The majority of oncogenic somatic alterations affecting tumor suppressors were associated with biallelic inactivation (72%), but this was not the case for oncogenes (9%) (Figures 1B and S1; see STAR Methods). MutLOH (i.e., mutations with a concomitant copy-number loss of heterozygosity) constituted the dominant form of biallelic event in both TSGs (54%) and oncogenes (83%). As expected, oncogenic homozygous deletions were exclusive to TSGs and comprised 30% of all biallelic events in these genes. The mechanism of loss of WT varied substantially by gene (Figure 1C; Table S2A): for example, the majority of biallelic events in *CDKN2A* (83%), *TGFB2* (65%), *FAT1* (62%), and *B2M* (55%) were the result of homozygous deletions, whereas nearly all biallelic events in *TP53* and *VHL* were the consequence of a mutation with a concomitant loss of heterozygosity.

Preference of mechanism for biallelic inactivation for each gene also varied by tumor lineage. We found 19 instances ($q < 0.05$) where the gene’s observed mechanism of biallelic inactivation in a given tumor lineage was significantly different from its pan-cancer pattern of biallelic inactivation (Figure 1C; Table S2A). For example, at pan-cancer resolution, *PTEN* biallelic losses most commonly arose via MutLOH (54% of cases), but homozygous deletions were the dominant mechanism among prostate adenocarcinomas (75%) and high-grade serous ovarian tumors (91%).^{16,17} In contrast, composite mutations comprised 49% of biallelic *PTEN* events in uterine endometrioid cancers despite excluding tumors with high mutational burden (see STAR Methods). Similarly, biallelic loss via composite mutations was also common among *PIK3R1* mutated uterine endometrioid tumors (70%) compared with 23% pan-cancer. This suggests that cancer-type-specific differences in mutational and copy-number burden drives the adoption of specific genetic mechanisms for second hits.

Gene-specific variation in rates of biallelic inactivation

Inactivation of both alleles is often considered a near-obligate event for most recessive TSGs. However, among the 174 TSGs that were mutated in 25 or more tumors, only 30% (52/174) demonstrated a biallelic rate (defined as the fraction of oncogenic alterations with evidence of biallelic alteration) of 80% or higher (Figure 2; Table S2B). This included many well-established and highly mutated tumor suppressors in many cancers, including *TP53* (biallelic rate 92%), *CDKN2A* (96%), *APC* (90%), *PTEN* (91%), and *RB1* (93%). Similarly, the rate of biallelic loss was near absolute among TSGs that were infrequently mutated pan-cancer but highly mutated in specific lineages in which their roles in tumorigenesis are well established. These include *BAP1* (94%) and *VHL* (95%) in clear cell renal cell carcinoma, *CDH1* (91%) in breast invasive lobular carcinoma, and *MEN1* (86%) in pancreatic neuroendocrine tumors. Somatic loss of both alleles was also common among TSGs that were very rarely mutated in any disease. For example, *CBFB* and *PRDM1* were commonly altered in hematological malignancies but rarely mutated in solid tumors (0.8% and 0.4%, respectively),

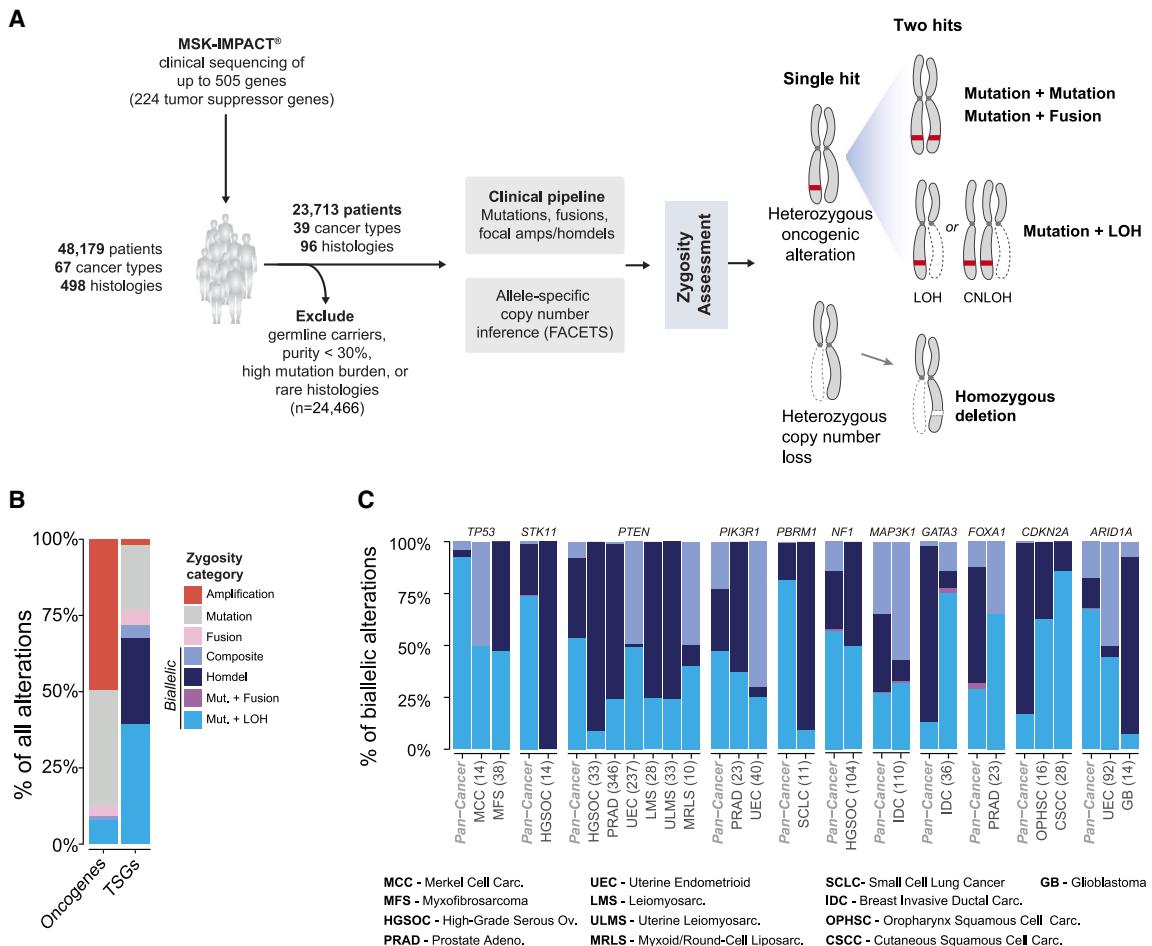


Figure 1. Somatic biallelic inactivation of tumor suppressors in prospectively sequenced patients

(A) Study schema describing the MSK-IMPACT cohort. LOH, loss of heterozygosity.

(B) Zygosity changes associated with mutations in oncogenes ($n = 186$ genes) are significantly different from those observed in tumor suppressors ($n = 224$ genes).

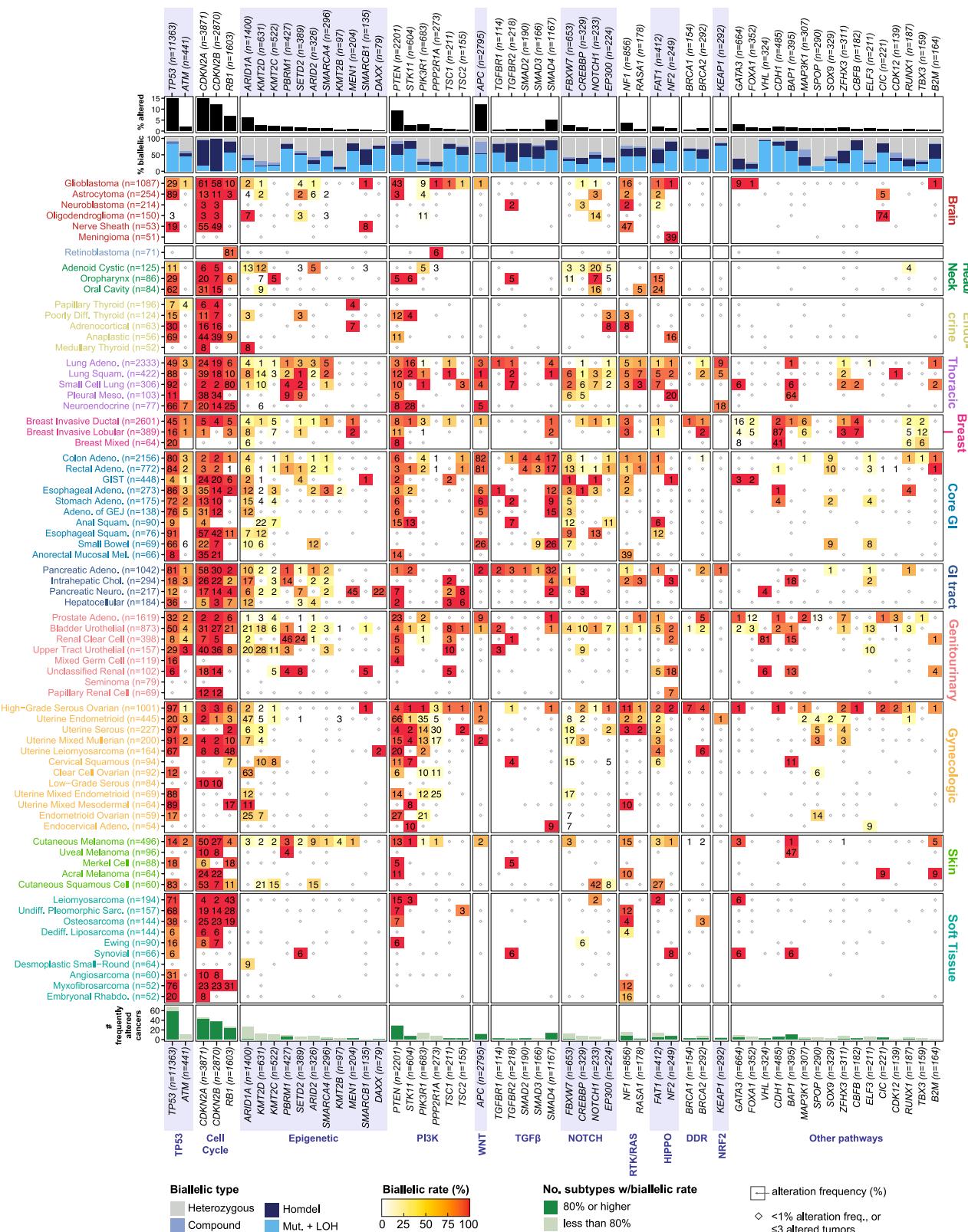
(C) Types of zygosity changes adopted for biallelic inactivation varied by cancer type for some genes when compared with their pan-cancer patterns. See also Figure S1.

but loss of WT was near universal (95% and 90%, respectively) in tumors with oncogenic mutations in these genes.

In contrast, for the remaining 70% (122/174) of TSGs, greater than 20% of oncogenic mutations arose without concomitant disruption of the WT allele. These genes clustered in certain pathways, including epigenetic regulators (22/24 genes in pathway, median biallelic alteration rate 47%), mediators of DNA damage repair (26/29, median biallelic alteration rate 48%) and NOTCH signaling (9/9, median biallelic alteration rate 42%) (Figures 2 and S2A; Table S2C). Functional monoallelic inactivation has been described previously for several of these genes. Consistent with prior evidence suggesting that heterozygous mutations in *ARID1A* are sufficient to drive tumor progression phenotypes in some diseases, nearly half of all *ARID1A* mutated tumors in our study retained the WT allele.^{18–20} There was no difference in the rate of alterations in other SWI/SNF complex genes between tumors with *ARID1A* heterozygous (11%) and *ARID1A* biallelic (9%) alterations ($p = 0.66$).

Selective pressure for MutLOH in TSGs

We also observed substantial lineage-specific variation in biallelic inactivation (Figure 2) suggesting that biallelic alterations to TSGs may be under evolutionary selection in disease-dependent contexts. To quantitatively define whether these patterns of recurrent biallelic inactivation represented signals of evolutionary selection, we focused on the most common form of biallelic inactivation (representing 55% of all cases of biallelic inactivation in TSGs; Figure 1B): mutation plus loss of heterozygosity (MutLOH), comparing the rate of LOH at the gene locus among tumors with mutation with those without the mutation (see STAR Methods). At the pan-cancer level, 47 of 60 (78%) tumor suppressors with mutations in 100 or more affected cases showed significant evidence of selection for MutLOH ($q < 0.05$; Table S3A) with 45 genes enriched for loss of WT and two genes significantly favoring the retention of the WT allele (Figure 3).



(legend on next page)

Given the heterogeneity of selection for MutLOH across genes and diseases, we developed a schema (see **STAR Methods**) to categorize all TSGs into four classes based on their preferential loss or retention of the WT allele among diseases in which they were commonly mutated (that is, with 20 or more mutated tumors). As detailed below, our classification schema relied on informed but subjective thresholds; in **Figure S3**, we demonstrate that varying these thresholds produces nearly identical classification patterns across genes (see also **STAR Methods** and **Methods S1**).

We defined Class 1 TSGs as those that showed preferential loss of WT allele in more than 80% of eligible cancer types (**Figure S1**). In all, 32 of 66 genes were identified as Class 1 and included widely mutated TSGs such as *TP53* (significant for biallelic loss in 38 of 38 cancer types with 20 or more mutations), *RB1* (11/11), and *PTEN* (10/10), as well as *NF1* (6/6), *SMAD4* (6/6), *BAP1* (4/4), *APC* (4/4) and *CDH1* (3/3) all of which were observed to ubiquitously prefer loss of the WT allele in every disease in which they were recurrently mutated (**Figure 3**).

In contrast, Class 2 TSGs ($n = 13$, including *ATM*, *ARID1A*, *FBXW7*, and *ARID2*) showed highly lineage-specific patterns of selection for biallelic inactivation and were identified as those that preferentially lost the WT allele in at least one cancer type (but fewer than 80% of all evaluable cancer types) in which they were commonly mutated. For example, although clear selection for MutLOH was apparent among *ATM*-mutant lung (odds ratio [OR] = 6.7, $q = 1.1 \times 10^{-9}$), prostate (OR = 5.2, $q = 1 \times 10^{-3}$), and colon (OR = 5.1, $q = 4.5 \times 10^{-7}$) adenocarcinomas, no such signal was evident among bladder urothelial carcinoma (OR = 2, $q = 0.3$), despite a sufficient prevalence of oncogenic *ATM* mutations (2.9% of bladder urothelial tumors with *ATM* mutations). The most archetypal example of this class of genes was *ARID1A* (mutated in 5% of all cancers). In total, 9/14 cancer types with frequent *ARID1A* mutations had no evidence of selection for MutLOH, including upper tract urothelial carcinoma (17% mutation rate, OR = 1.1, $q = 0.94$), colon adenocarcinoma (3.6%, OR = 1.1, $q = 0.86$), LUAD (3.3%, OR = 1.48, $q = 0.25$), and rectal adenocarcinoma (5.6%, OR = 0.93, $q = 0.94$). Among other Class 2 genes, enrichment for MutLOH was an exception rather than the norm. For example, 87% of *FBXW7* mutated tumors in lung squamous cell carcinomas had loss of heterozygosity (OR = 4.1, $q = 4.99 \times 10^{-2}$), whereas in six other lineages, including colon (OR = 1, $q = 0.95$) and rectal (OR = 1.4, $q = 0.39$) adenocarcinomas, no evidence for selection for MutLOH was observed.

Next, we classified 27% (18/66; **Figure 3**) of TSGs that did not show selection for MutLOH in any lineage as Class 3. These included PI3K pathway members *PIK3R1* and *PPP2R1A*, which act as negative regulators of *PIK3CA* and *AKT1/2/3*, respec-

tively.²¹ None of the seven lineages in which *PIK3R1* is commonly mutated showed either significant enrichment or depletion of second hits. Consistent with prior evidence that hot-spots in *PPP2R1A* are potentially dominant-negative,²² we found that despite the high mutation rates of *PPP2R1A* in numerous uterine cancer subtypes (serous [30%), mixed endometrial [25%], and carcinosarcoma [17%]), only 19% of *PPP2R1A* mutated tumors in these diseases harbored concomitant LOH ($q > 0.05$). Additionally, many Class 3 TSGs were chromatin modifying and remodeling genes in the epigenetic pathway, such as *ARID1B*, *CREBBP*, *KMT2A*, *KMT2C*, and *ASXL1* (**Figure S2B**; **Table S3B**).

Finally, genes that, when mutated, preferentially retained the WT allele at the pan-cancer level were designated as Class 4 ($n = 3$). This class included genes for which both oncogenic and tumor-suppressive roles have been proposed, such as *FOXA1*.^{23–26} Here, at the pan-cancer level, we find significant depletion for MutLOH in *FOXA1*-mutated cancers (OR = 0.43, $q = 7.6^{-5}$) implying a dependence on the retention of the WT that remains to be understood. Similar depletion was also observed for *GATA3* at the pan-cancer level (OR = 0.29, $q = 3.2^{-15}$) with suggestive evidence for depletion for MutLOH in breast invasive ductal cancer (OR = 0.62, $q = 0.063$).

Biallelic enrichment identifies rare driver TSG events in APC in non-canonical contexts

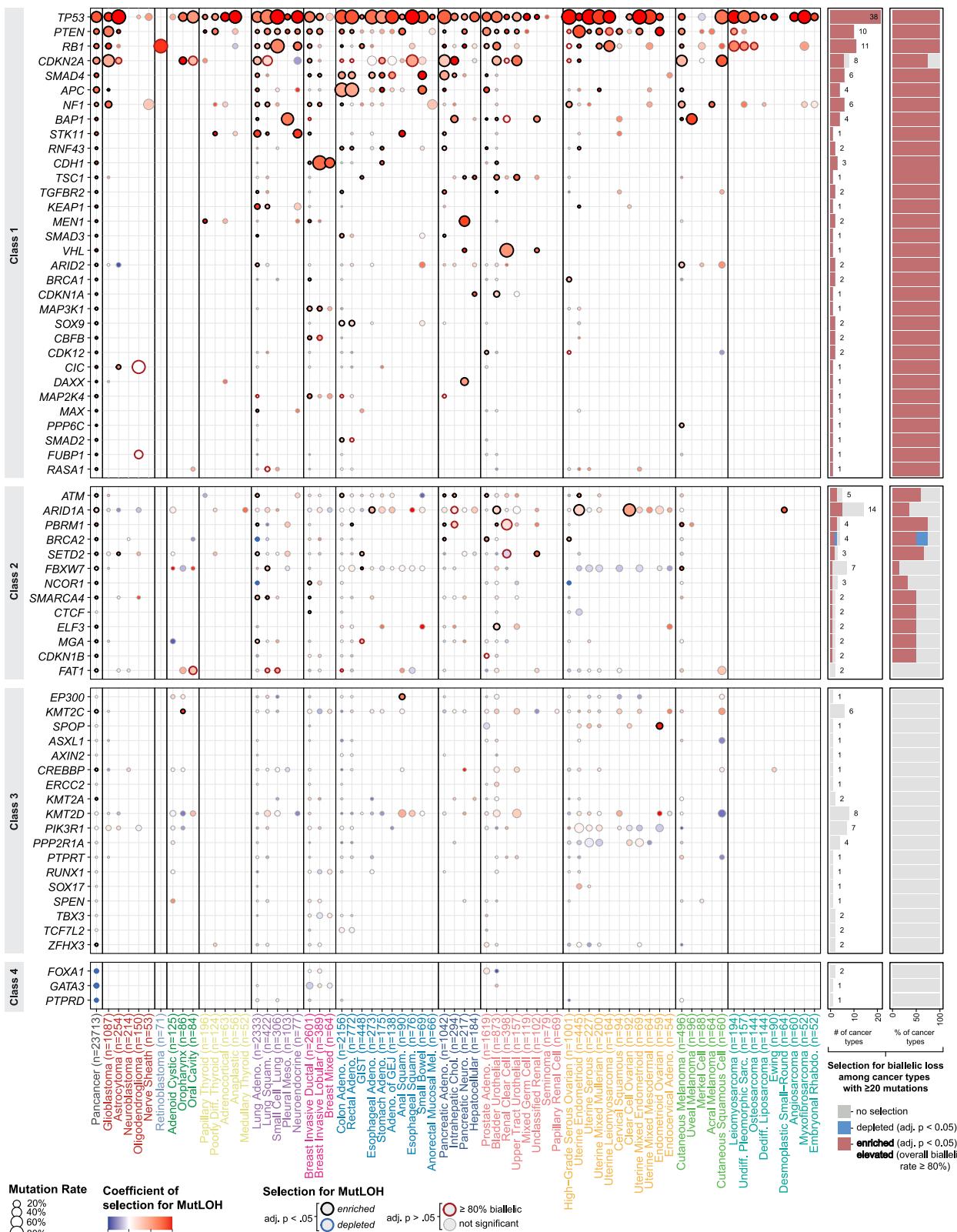
Establishing the functional significance of infrequently mutated genes is often challenging. We hypothesized that enrichment for MutLOH (as in **Figure 3**) could be used as a metric of evolutionary selection, and thereby functional relevance, of genes mutated in non-canonical contexts. For example, *BAP1* was under significant selection for MutLOH both in cancers where it is highly mutated (mesotheliomas, renal cell cancers, uveal melanomas, and intrahepatic cholangiocarcinomas), as well as those with comparatively few *BAP1* mutations, such as prostate adenocarcinoma (0.2% of tumors, OR = 163, $q = 3 \times 10^{-4}$) and LUAD (0.7%, OR = 26, $q = 3 \times 10^{-5}$). Similarly, *STK11* mutations were infrequent among cancers other than those of thoracic origin but, when present, were often observed with concomitant loss of the WT allele (e.g., in pancreatic adenocarcinoma, OR = 8.3, $q = 1.9 \times 10^{-3}$; in breast invasive ductal carcinoma, OR = 9.8, $q = 4 \times 10^{-4}$). In hepatocellular carcinomas, *TSC1* and *TSC2* mutations, although infrequent (<3%), were associated with a unique and aggressive form of HCC that is responsive to *MTOR* inhibition.²⁷ Consistent with this, we find universal biallelic loss of *TSC1* (5/5 mutations biallelic) and *TSC2* (5/5 mutations biallelic) in HCC.

Focusing on genes mutated in <10% of samples in a particular cancer type (i.e., in a non-canonical cancer lineage), we noted an

Figure 2. Landscape of biallelic alterations by tumor suppressor gene and cancer type

Tumors with fusion events in the corresponding genes and heterozygous oncogenic TSG mutations with a composite VUS mutation in the same gene were excluded because their biallelic status cannot be robustly ascertained. Heterozygous LOH was considered as wild type (WT) when calculating biallelic inactivation rates for somatic oncogenic alterations. For a complete list of biallelic rates for all 224 TSGs across 96 cancer types evaluated in this study, refer to **Table S2B**. Color corresponds to the biallelic inactivation rate of each gene within the cancer type. The bar chart below shows the total number of cancer types in which the gene was found altered in at least 10 tumors and an alteration rate of at least 1% within the corresponding cancer type. Of these, the number of cancer types in which the gene presented with a biallelic rate >80% is shown in dark green.

See also **Figure S2**.



(legend on next page)

enrichment for biallelic loss of *APC* in prostate (OR = 36, $q = 9 \times 10^{-30}$) and lung (OR = 34, p value = 2×10^{-12}) adenocarcinomas, which were among the most statistically significant findings (Figure 4A). Biallelic loss of *APC* is an early and obligate event for tumorigenesis in most colorectal cancers that promotes stabilization and nuclear translocation of the transcription factor β -catenin.²⁸ In total, we identified seven tumor lineages (glioblastomas, cutaneous melanomas, and adenocarcinomas of prostate, lung, pancreas, stomach, and esophagus) in which *APC* was mutated in fewer than 10% of cases but was significantly enriched for biallelic alterations via MutLOH, which has been described in part in prior work.^{17,29,30} In The Cancer Genome Atlas (TCGA), prostate adenocarcinomas (OR = 48, p value = 4.9×10^{-6}) were similarly enriched for biallelic loss of *APC* (Figure 4B). However, statistical significance was not reached for the LUAD patients in TCGA (OR = 2.7, p value = 0.18) or Tracking Cancer Evolution through therapy (Rx) (TRACERx, OR = 3, p value = 0.065) cohorts, likely due to small sample sizes. Despite this, consistent with selection for biallelic loss of *APC*, LUADs in the TCGA with biallelic but not heterozygous alterations in *APC* showed significantly higher levels of β -catenin protein, whereas *CTNNB1* gene expression levels remain unchanged (Figure 4C). These data indicate that *APC* is a rare but positively selected driver gene in both prostate and LUADs.

Wnt pathway mutations arise late in lung and prostate adenocarcinomas

We sought to understand the genomic contexts in which Wnt pathway mutations arose in lung and prostate cancers. As in colorectal cancers,³¹ we observed mutual exclusivity between mutations *APC*, *CTNNB1*, and other Wnt genes (Figures 4D and 4E). We also observed shared patterns of co-mutations between different Wnt genes and other canonical drivers. For example, in lung cancer, both *APC^{MUT}* and *CTNNB1^{MUT}* tumors were significantly enriched for mutations in *EGFR*, *SMAD4*, and *PTEN* compared with tumors WT for Wnt alterations. In prostate cancer, *SPOP* mutations were significantly enriched in *APC^{MUT}* tumors compared with Wnt WT tumors (41% vs. 11%, Benjamini-Hochberg [BH]-corrected p value = 9×10^{-25}) and were elevated in *CTNNB1^{MUT}* tumors (18%, BH-corrected p value = 0.1).

To understand the evolutionary origin of Wnt pathway mutations in lung and prostate cancers, we examined patients with two or more clonally related tumor specimens collected longitudinally over the course of their clinical care. For these patients, we reasoned that absence of a mutation in an early specimen (but presence in a later specimen) implied late evolutionary emergence of a mutation (Figure 4F; see STAR Methods). Consistent with their role as early drivers, we found no instances

of exclusively late-arising *EGFR*-mutant lung cancers (0/280 patients), *SPOP*-mutant prostate adenocarcinoma (0/23 patients), or *APC*-mutant colorectal cancers (0/118 patients) (Figure 4G). In contrast, in patients who acquired specific resistance mutations in response to treatment in lung,³² colorectal,³³ and prostate cancers,³⁴ these mutations were detected in only the later sequenced biopsies (Figure 4G). Unlike truncal/early *APC* mutations in colorectal cancers, in nearly a third of each of the lung (35%, 8/23) and prostate (30%, 7/23) adenocarcinomas we evaluated, we observed *APC* mutations in only one of the sequenced biopsies (Figures 4G and 4H). Similarly, *CTNNB1* mutations were also late arising in 49% (19/39) and 38% (5/13) of lung and prostate adenocarcinomas, respectively (Figures 4G and 4H). Despite the later acquisition of *APC* mutations, both the biallelic rates and the selection pressure to lose the WT were indistinguishably high in both primary and metastatic tumors of both lung and prostate adenocarcinomas (Figure S4; Table S3C; STAR Methods). Together, these observations suggest that Wnt pathway mutations significantly co-occur with canonical drivers in lung (*EGFR*) and prostate (*SPOP*) cancers and that in many cases, these mutations arise late in tumor evolution, suggesting that they mediate tumor progression or treatment resistance.

Selection for MutLOH reclassifies functional VUSs

Mutational recurrence at a residue is a singularly important determinant for prediction of oncogenic potential of mutant alleles. Unlike in oncogenes, missense mutations in TSGs often do not cluster at single residues and are rather dispersed across the length of the gene,³⁵ rendering many putative oncogenic alleles to be classified as VUSs.¹³ We hypothesized that selection for biallelic inactivation could be used as a metric to identify genes with strong enrichment for putatively functional VUSs.

We therefore quantified selection for MutLOH in gene/cancer-type pairs with at least 10 tumors with VUSs and at least 10 tumors with OncoKB-annotated oncogenic mutations (Figure 5A). We identified 15 gene/cancer-type pairs with evidence of positive selection for MutLOH in the context of a VUS (BH-corrected p value < 0.05; Table S3D). LUAD had the highest number of genes (*ATM*, *KEAP1*, *STK11*, and *CDKN2A*) with significant enrichment for MutLOH among VUSs. Interestingly, we identified three examples of universal MutLOH among VUSs, including *MEN1* in pancreatic neuroendocrine cancers (25/25 cases with MutLOH, $q = 4.2 \times 10^{-6}$), a tumor type in which loss of *MEN1* is a pathognomonic genetic event, and *CBFB* in breast invasive ductal carcinoma (30/30, $q = 8.6 \times 10^{-6}$). Thus, although VUSs demonstrate no widespread selective pressure for loss of the WT allele in most genes, this selection is near complete in select gene/cancer-type contexts.

Figure 3. Selection for biallelic inactivation across genes and cancer types

All tumors with fusions, homozygous deletions, and composite mutations were excluded from consideration when determining each gene's signal for selection within a cancer type. Circles with black outlines denote statistically significant enrichment for biallelic inactivation, whereas blue outlines denote significant depletion (adjusted p value < 0.05). The circle with red outline denotes genes and cancer-type pairs with at least 20 tumors with any oncogenic alteration and an overall biallelic inactivation rate of $\geq 80\%$ (as in Figure 2), but statistical significance for selection for MutLOH was not reached. The two bar charts to the side summarize the selection patterns observed among cancer types with at least 20 oncogenic mutations in the given gene among cancer types in which selection for MutLOH was evaluated. See STAR Methods for TSG classification criteria.

See also Figures S2–S4.

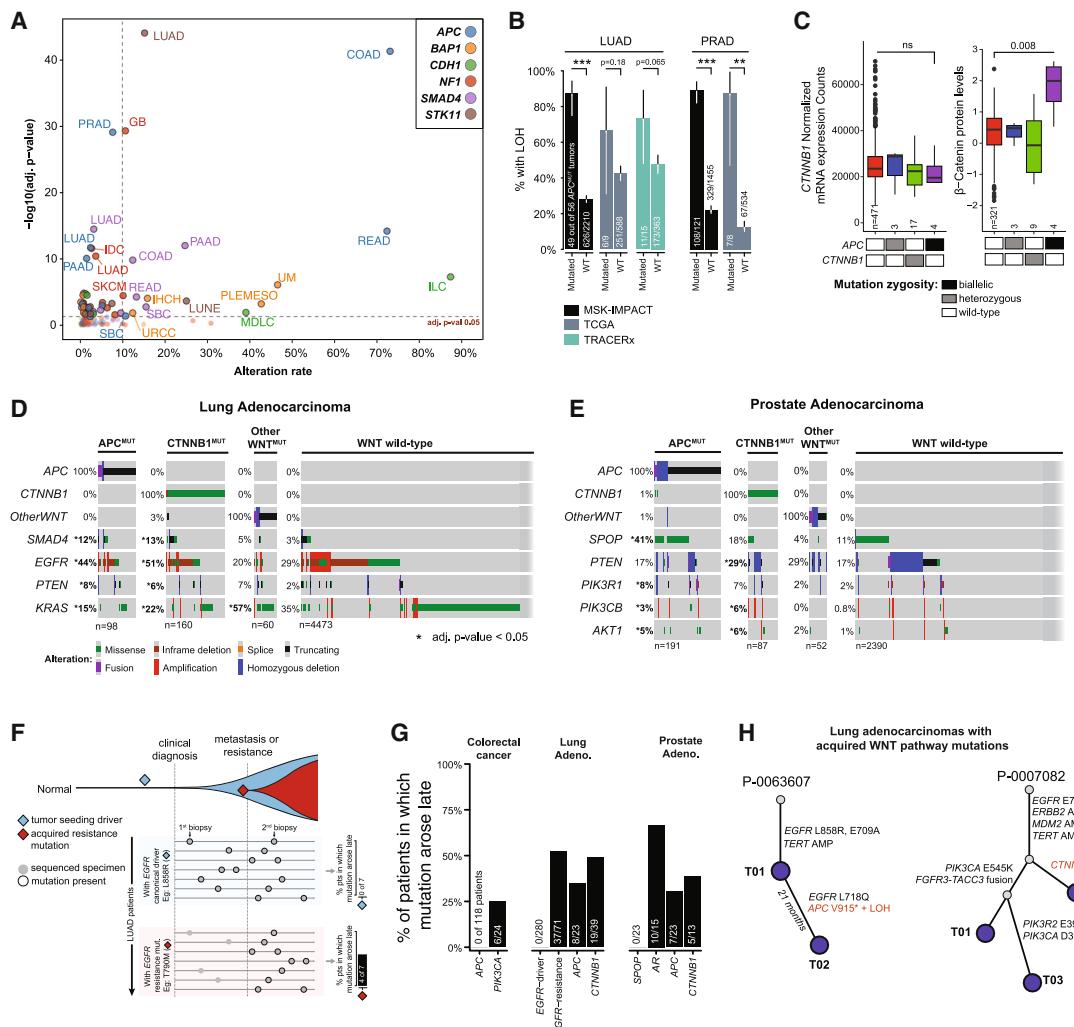


Figure 4. Biallelic inactivation among rare drivers identifies late-arising APC mutations in several cancers

(A) Scatterplot of enrichment for MutLOH (as calculated in Figure 3) as a function of alteration rate of TSGs in each labeled cancer type.

(B) Fraction of APC mutated and WT patients with LOH at the APC locus for MSK-IMPACT, TCGA, and TRACERx cohorts. *** p value < 0.001, ** p value < 0.01, two-sided Fisher's exact test was used. LOH, loss of heterozygosity; LUAD, lung adenocarcinoma; PRAD, prostate adenocarcinoma. Error bars represent 95% binomial confidence interval (CI).

(C) CTNNB1 gene and protein expression for LUAD patients in TCGA for different APC zygosity and CTNNB1 mutation groups. Wilcoxon test was used. The center line of the boxplots is the median, and the lower and the upper hinges represent the interquartile range (IQR). Upper and lower whiskers extend to 1.5x the IQR.

(D and E) Oncoprint showing alterations in key (D) LUAD and (E) PRAD cancer genes across different Wnt pathway mutation groups. Alterations that are enriched or depleted in APC^{MUT}, CTNNB1^{MUT} or OtherWnt^{MUT} compared with Wnt WT patients are indicated by *** (adjusted p values < 0.05). Other Wnt genes include AMER1, AXIN1, AXIN2, GSK3B, LZTR1, RNF43, TCF7L2, and ZNR3.

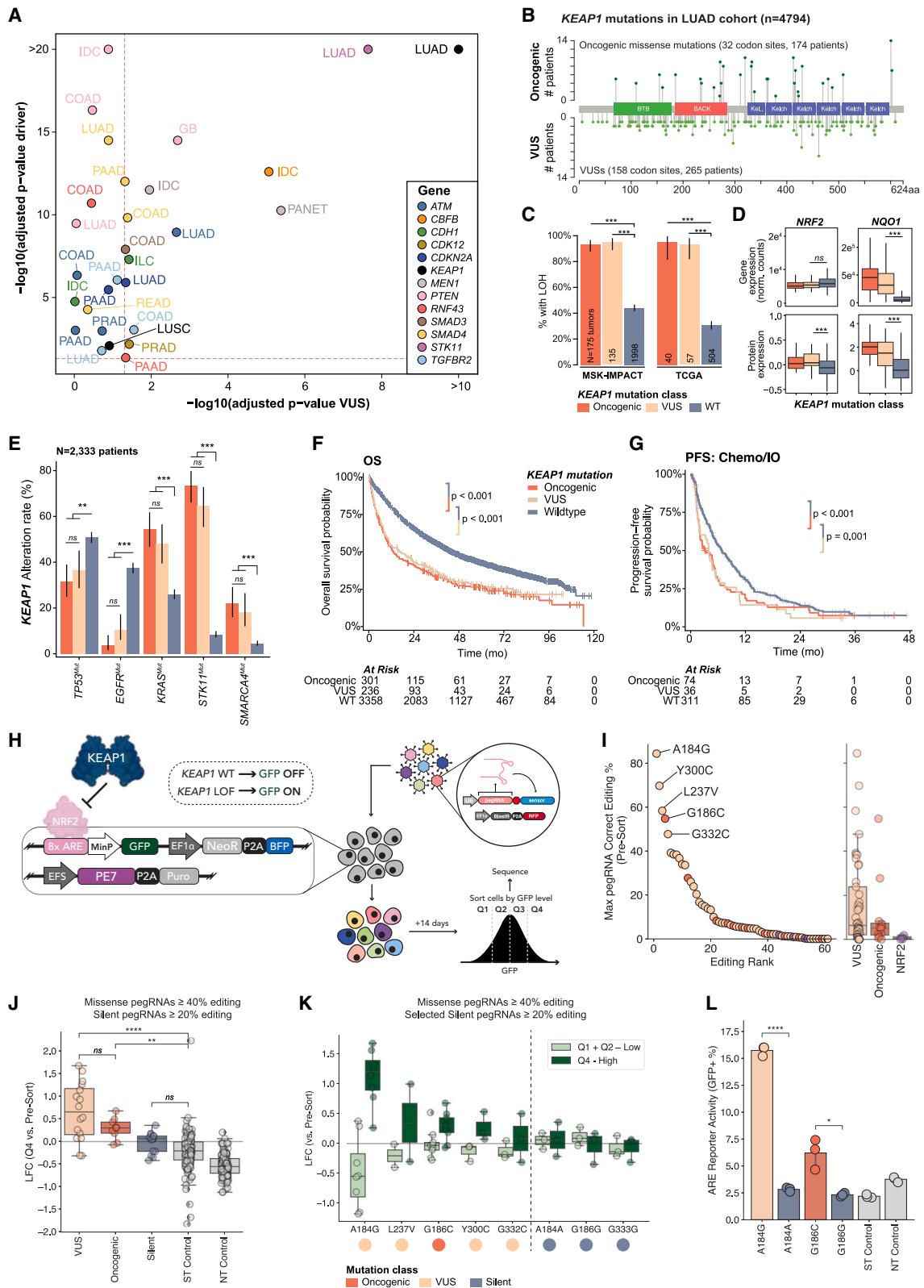
(F) Schematic depicting approach to identify late-arising mutations in tumor lineages. In routine clinical sequencing early driver mutations (blue diamond), such as EGFR drivers in lung cancers, are expected to be detected in every biopsy sequenced. However, resistance mutations (red diamond) such as EGFR "gatekeeper" mutations in lung cancers are acquired on treatment and are absent in biopsies taken before the tumors acquire resistance.

(G) APC and CTNNB1 alterations are evaluated corresponding to (F) to determine if they arise late in the given tumor lineages. APC, EGFR, and SPOP are known early drivers in colorectal, lung, and prostate cancers, respectively. PIK3CA, EGFR-resistance (gatekeeper), and AR mutations are known acquired mutations in the indicated diseases.

(H) Select lung cancer patients with late-arising/acquired APC and CTNNB1 mutations.

The most statistically significant selection for MutLOH among VUSs was in KEAP1 in LUAD (Figure 5A). This was notable given that only 17% of all missense mutations in KEAP1 were classified as oncogenic by OncoKB, in part because the majority of all

VUSs were observed in two or fewer tumors (Figure 5B). In total, 8% of all MSK-IMPACT patients with LUAD harbored OncoKB-annotated oncogenic mutations in KEAP1, and an additional 5% had KEAP1 missense VUSs. The strong selection for



(legend on next page)

MutLOH was evident among both oncogenic mutations and VUSs in both the MSK-IMPACT and the TCGA cohorts of LUADs (Figure 5C).

In homeostatic conditions, *KEAP1* mediates the ubiquitination and subsequent degradation of the transcription factor NRF2 (also known as *NFE2L2*). However, in the presence of reactive oxygen species or loss-of-function mutations in *KEAP1*, NRF2 translocates to the nucleus and binds antioxidant response elements (AREs) to activate the transcription of cytoprotective anti-oxidant genes.³⁶ Consistent with its function in regulating NRF2 at the protein level, tumors in the TCGA LUAD cohort with *KEAP1* VUSs, as well as those with *KEAP1* oncogenic mutations, showed significant increases in NRF2 protein, but not RNA, abundance relative to WT ($p = 2 \times 10^{-5}$; Figure 5D). Similarly, both RNA and protein expression of the NRF2 target gene *NQO1* were also elevated (RNA VUS vs. WT, $p = 7 \times 10^{-20}$; protein VUS vs. WT, $p = 7 \times 10^{-11}$). Finally, both *KEAP1* oncogenic mutations and *KEAP1* VUSs showed similar patterns of co-mutations with other key drivers of LUAD (Figure 5E). Altogether, these data imply that *KEAP1* VUSs in lung cancers phenocopy aspects of *KEAP1* oncogenic mutations by activating NRF2 signaling.

KEAP1 mutations, often in conjunction with *STK11*, define a genetically distinct group of LUADs with a poor response to systemic therapy and overall poor prognosis.³⁷ Based on the functional and genomic data above, we hypothesized that patients with tumors harboring *KEAP1* VUSs would show clinical outcomes indistinguishable from those with *KEAP1* oncogenic alleles. Consistent with this, MSK-IMPACT patients with LUAD harboring *KEAP1* VUSs had similar overall survival as those with *KEAP1* oncogenic mutations when compared to patients with *KEAP1* WT tumors (VUS vs. WT, median overall survival

[mOS] of 17 vs. 44 months, $HR_{adj} = 1.7$ [95% confidence interval (CI): 1.4–2.0], p value = 2.1×10^{-9} ; oncogenic vs. WT, mOS 13 vs. 44 months, Adjusted Hazard Ratio (HR_{adj}) = 1.7 [95% CI: 1.4–1.9], p value = 1.5×10^{-11}) (Figures 5F and S5; Tables S4A and S4B), an effect which was not evident in VUSs in other genes (see STAR Methods; Figures S5A and S5B). In addition to OS, we also sought to evaluate whether *KEAP1* VUS would be similarly associated with poor progression-free survival (PFS) on standard-of-care chemoimmunotherapy in advanced non-small cell lung cancers (NSCLC) (chemo-IO cohort, $n = 421$ patients).³⁸ Again, patients with tumors harboring either VUS or oncogenic driver mutations in *KEAP1* had significantly worse PFS when compared to patients with *KEAP1*^{WT} tumors (VUS vs. WT, mPFS of 4 vs. 6.8 months, $HR_{adj} = 2.0$ [95% CI: 1.3–3.0], p value = 1.3×10^{-3} ; oncogenic vs. WT, mPFS 3 vs. 6.8 months, $HR_{adj} = 1.7$ [95% CI: 1.3–2.3], p value = 5.8×10^{-4}), further suggesting that most *KEAP1* VUS are likely oncogenic (Figures 5G and S5C).

Prime editing screen establishes that *KEAP1* VUSs activate NRF2

We sought to experimentally determine if *KEAP1* VUS phenocopy annotated driver variants by using prime editing to engineer and screen *KEAP1* variants. To do so, we adapted a reporter system to quantify the effects of somatic alterations in *KEAP1* to suppress or activate NRF2³⁹ (Figure 5H). We constructed a genetic reporter containing eight copies of the NRF2-targeted antioxidant response element (ARE) next to a minimal promoter linked to GFP (8× ARE-GFP), such that GFP expression indicates functional NRF2 transcriptional activity (Figure 5H). We transduced this reporter into NCI-H1299 NSCLC cells stably expressing PE7, a recently developed prime editor.⁴⁰ These cells

Figure 5. Enrichment of biallelic inactivation among VUSs

(A) Enrichment for biallelic losses among TSGs with VUS vs. known oncogenic mutations. Dashed lines indicate the adjusted p value = 0.05. See Table S1A for cancer type definitions.

(B) Lollipop plot showing sites of oncogenic and VUS missense mutations in *KEAP1* in LUAD (excluding tumors with TMB higher than 90th percentile in LUAD).

(C) LOH rates of *KEAP1* mutated and WT tumors in LUAD in MSK-IMPACT and TCGA cohorts. *** p value < 0.001, two-sided Fisher's exact test was used. Error bars represent 95% binomial CI.

(D) Gene and protein expression of NRF2 and *NQO1* in LUAD tumors in TCGA with either *KEAP1* oncogenic mutations, *KEAP1* VUSs, or *KEAP1* WT. *** p value < 0.001, Wilcoxon test was used.

(E) Co-mutation patterns of *KEAP1* VUSs with genes known to be co-occurring or mutually exclusive with *KEAP1* oncogenic mutations.

** p value < 0.01, *** p value < 0.001, ns, not significant, two-sided Fisher's exact test was used. Error bars represent 95% binomial CI.

(F) Kaplan-Meier curves showing overall survival (OS, in months) for patients with lung adenocarcinoma in the MSK-IMPACT cohort with tumors harboring *KEAP1* oncogenic mutations or VUSs compared with *KEAP1*^{WT}. p values are computed from a multivariate Cox proportional hazards model accounting for significant clinico-genomic covariates (see STAR Methods and Figure S5B for full model).

(G) Kaplan-Meier curves showing progression-free survival (PFS) of advanced NSCLC patients receiving first-line chemoimmunotherapy ($n = 421$) by *KEAP1* mutation class. p values are computed from a multivariate Cox proportional hazards model accounting for significant clinico-genomic covariates (see STAR Methods and Figure S5C for full model).

(H) Schematic of prime editing screen. BlastR, blasticidin S selection marker; EF1 α , elongation factor 1-alpha promoter; EFS, elongation factor 1-alpha short promoter; MinP, minimal promoter; NeoR, neomycin selection marker; P2A, peptide 2A RT; Puro, puromycin selection marker.

(I) Maximum pegRNA correct editing percentage for missense mutation-inducing pegRNAs in the library.

(J) The log₂ fold change (LFC) of the highest GFP-expressing bin (Q4) relative to pre-sort populations for different *KEAP1* variant or control classes. Statistics shown for t test of independent samples with Bonferroni correction. ** p value ≤ 0.01, *** p value ≤ 0.0001, ns, not significant (p value > 0.05).

(K) The LFC for low (Q1 and Q2) and high (Q4) GFP-expressing cells relative to the pre-sort population for missense pegRNAs with ≥40% editing and selected silent pegRNAs with ≥20% editing.

(L) Flow cytometry-based validation of the ARE-reporter activity (GFP+ %) of individual pegRNA-expressing NCI-H1299 cells 10 days after pegRNA transduction (ST, safe-targeting; NT, non-targeting). Statistics shown for t test of independent samples with Bonferroni correction. * p value ≤ 0.05, *** p value ≤ 0.0001, ns, not significant (p value > 0.05). Data are presented as mean values with a 95% CI. For (D) and (I–K), boxplots display median line, IQR boxes, and 1.5x IQR whiskers.

See also Figures S5 and S6.

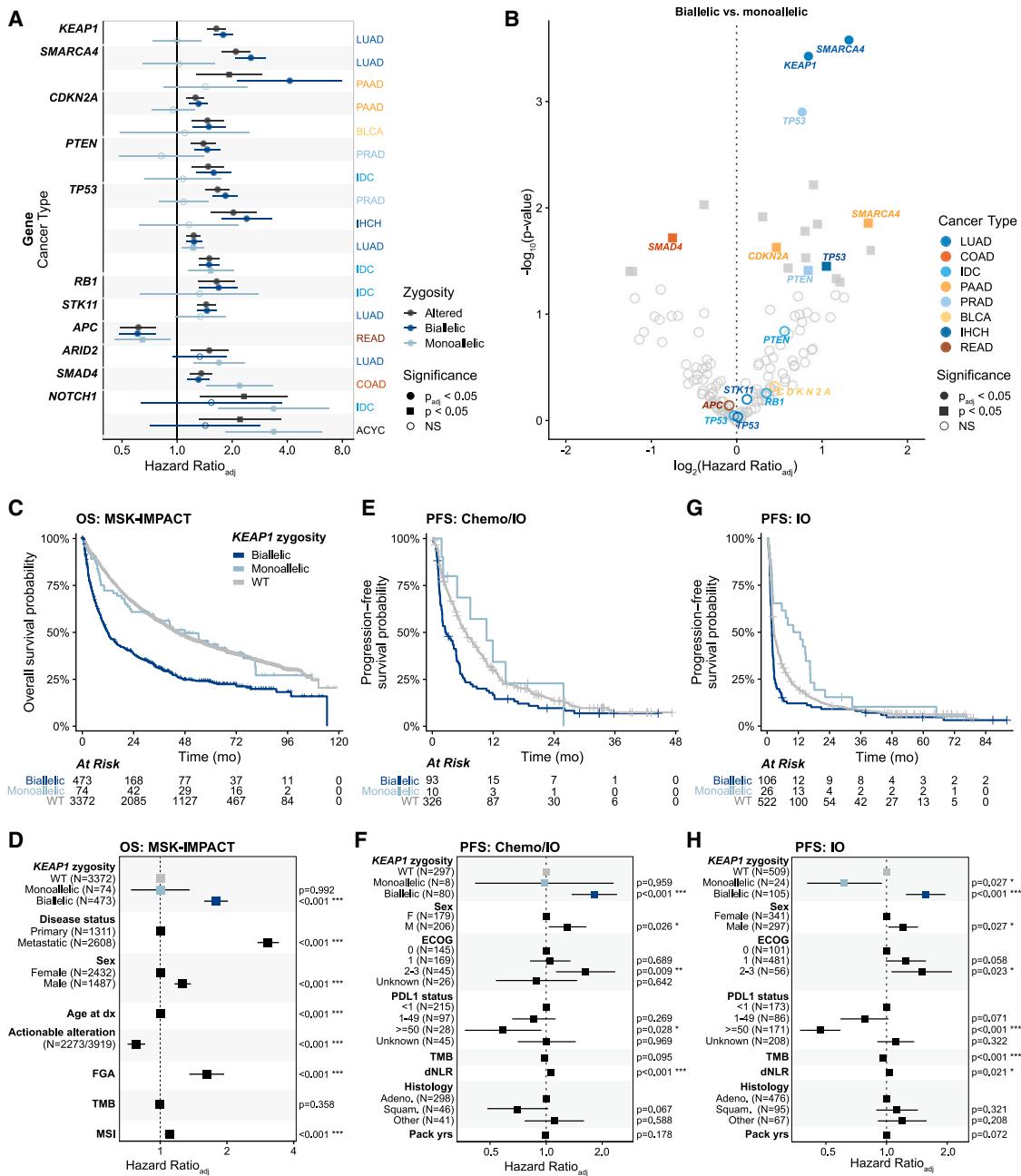


Figure 6. Zygosity as a prognostic and predictive biomarker

(A) OS among patients in the MSK-IMPACT cohort with tumors harboring biallelic, monoallelic, or any oncogenic alteration in a given gene compared with WT (see STAR Methods). Only genes that showed statistically significant difference in OS in either biallelic or monoallelic groups compared with WT are shown (see Table S4C for full list). Hazard ratios are from multivariate Cox proportional hazards models of OS by zygosity and alteration within each gene/subtype pair, adjusting for significant clinico-genomic covariates (see STAR Methods). LUAD, Lung Adenocarcinoma; PAAD, Pancreatic Adenocarcinoma; BLCA, Bladder Urothelial Carcinoma; PRAD, Prostate Adenocarcinoma; IDC, Breast Invasive Ductal Carcinoma; IHC, Intrahepatic Cholangiocarcinoma; READ, Rectal Adenocarcinoma; COAD, Colon Adenocarcinoma; ACYC, Adenoid Cystic Carcinoma.

(B) Volcano plot of OS for patients in the MSK-IMPACT cohort with tumors with biallelic inactivation relative to those with monoallelic inactivation among all gene/subtype pairs with sufficient sample size (see STAR Methods; Table S4D). Pairs in which biallelic inactivation was associated with significantly shorter OS compared with WT are highlighted and labeled.

(C) Kaplan-Meier curve of OS by KEAP1 zygosity status among patients in the MSK-IMPACT cohort with LUAD.

(D) Forest plot of multivariate Cox regression model of OS by KEAP1 zygosity as in (C).

(legend continued on next page)

carry WT copies of both *KEAP1* and *NFE2L2*,⁴¹ allowing us to validate the NRF2 reporter by treating them with tert-butylhydroquinone (tBHQ).^{42,43} As expected, we observed a robust, dose-dependent induction of 8× ARE-driven GFP expression (Figure S6A). We also validated the prime editing activity of NCI-H1299-PE7 cells with Lenti-PEAR-mCherry, a reporter construct where GFP is turned on in the event of successful prime editing^{44,45} (Figure S6B).

We adopted a recently described prime editing “sensor” approach for high-throughput screening. Prime editing sensors contain a prime editing guide RNA (pegRNA) and a synthetic copy of the endogenous target site where the guide is predicted to install specific mutations, thereby providing an integrated readout of the efficiency of a given pegRNA.^{44,46} We constructed a library of 500 pegRNA sensors targeting 59 missense variants in *KEAP1*, including 47 VUS and 12 annotated oncogenic variants. We further included four *NRF2* hotspot mutations, silent substitution control mutations for each mutated codon, and both safe- and non-targeting control pegRNAs, with multiple pegRNAs designed for each variant (Figures S6C and S6D). We delivered this pegRNA-sensor library to NCI-H1299 cells stably expressing PE7 and the 8× ARE-GFP reporter followed by antibiotic selection of cells with successful integrations. After allowing 14 days for editing to occur, we used fluorescence-activated cell sorting to isolate cells into four equally sized bins based on GFP expression levels (Q1–Q4) (Figure 5H). DNA sequencing of integrated pegRNA-sensor cassettes was used to quantify the abundance of pegRNAs and their editing efficiency at their cognate sensor sites in both the pre-sort and sorted cell populations, identifying efficient pegRNAs for many of the missense-inducing pegRNAs in the library (Figures 5I, S6E, and S6F). Given that NRF2 activation requires the loss of function of both *KEAP1* alleles, we restricted our analysis to five pegRNAs with at least 40% editing efficiency with a high likelihood of installing the desired edit at both *KEAP1* alleles (Figures 5J, 5K, and S6G). Consistent with the hypothesis that *KEAP1* VUS phenocopy known oncogenic driver mutations, we found that highly efficient pegRNAs that generated oncogenic mutations or VUS (but not silent substitutions) all exhibited NRF2 activation, as evidenced by their enrichment in the highest GFP-expressing bin (Q4) (Figures 5J and 5K).

The variant that was most strongly enriched in the high GFP-expressing population (Q4) was A184G, a novel VUS that was observed in one patient, and also had the highest efficiency pegRNA (>80% sensor editing) (Figures 5I–5K). We further tested this variant, alongside the highest scoring annotated oncogenic driver mutation in *KEAP1* (G186C) and silent substitution-generating pegRNAs (A184A and G186G), by individually transducing these pegRNAs into NCI-H1299-PE7-8×ARE-GFP cells. We validated that the A184G and G186C pegRNAs introduced their edits at the endogenous *KEAP1* locus (Figure S6H) followed by flow cytometric assessment of their relative induction of the 8× ARE-GFP reporter. We found that both A184G and G186C exhibited significant GFP expression relative to A184A and G186G, respectively, and that A184G activated the ARE-GFP reporter more strongly than G186C (Figure 5L), an effect that was not ascribable to differences in editing efficiency (Figure S6I). These functional data confirm that specific *KEAP1* VUSs in LUAD phenocopy known oncogenic driver mutations.

***KEAP1* mutant zygosity is a prognostic biomarker to standard therapies in LUAD**

KEAP1 mutations are characteristically associated with inferior outcomes and therapeutic resistance in lung cancer, but the role of mutation zygosity in mediating this association in *KEAP1* and in other cancer driver genes is not understood.^{47–50} We therefore evaluated (in a multivariate manner controlling for disease status, age of diagnosis, sex, fraction of genome altered (FGA), tumor mutation burden (TMB), MSI status, and the presence of clinically targetable alterations) the association between OS and mutation status/zygosity for all genes with at least 10 monoallelic and 10 biallelic alterations in a given cancer type (in all, $n = 150$ gene/cancer type pairs; see STAR Methods). Interestingly, we observed several instances in which biallelic, but not monoallelic, alterations exhibited significantly different outcomes compared with WT patients (Figure 6A; Tables S4C–S4E). For example, in prostate adenocarcinomas, when compared with WT tumors, only the patients with biallelic, but not monoallelic, alterations in *TP53* showed worse outcomes of overall survival (biallelic vs. WT, $HR_{adj} = 1.8$, q value = 8×10^{-14} ; monoallelic vs. WT, $HR_{adj} = 1.1$, q value = 0.6).

Next, we systematically identified gene/cancer-type pairs in which outcome differences between tumors with biallelic and monoallelic alterations were statistically significant (Figure 6B). In addition to *TP53* in prostate adenocarcinomas, patients with tumors with biallelic alterations in *KEAP1* and *SMARCA4* in LUAD had significantly worse OS compared with those with monoallelic alterations in these respective genes (*KEAP1* biallelic vs. monoallelic, $HR_{adj} = 1.8$, q value = 4×10^{-4} ; *SMARCA4* biallelic vs. monoallelic, $HR_{adj} = 2.5$, q value = 3×10^{-4}). Strikingly, the overall survival among *KEAP1* monoallelic tumors was indistinguishable from that of *KEAP1* WT tumors (monoallelic vs. WT, mOS 48 vs. 44 months, $HR_{adj} = 1.0$, p value = 0.99) (Figures 6C and 6D). These data directly indicate that *KEAP1* zygosity is a critical determinant of the prognostic value of *KEAP1* mutations, independent of other clinical covariates.

KEAP1-mutated tumors have been associated with reduced efficacy of standard-of-care first-line immune checkpoint inhibitors both as monotherapy or in combination with chemotherapy (referred to as chemoimmunotherapy).^{37,38} Based on the data in Figures 6A–6D, we hypothesized that *KEAP1* zygosity status may also be prognostic of responses to these therapies and sought to evaluate PFS in two cohorts of advanced NSCLC patients treated with chemoimmunotherapy (chemo-IO cohort

(E) Kaplan-Meier curve of PFS on first-line chemoimmunotherapy among patients with advanced NSCLC.

(F) Forest plot of multivariate Cox regression model of PFS on first-line chemoimmunotherapy by *KEAP1* zygosity as in (E).

(G) Kaplan-Meier curve of PFS on first-line immunotherapy alone by *KEAP1* zygosity among patients with advanced NSCLC.

(H) Forest plot of multivariate Cox regression model of PFS on first-line immunotherapy alone by *KEAP1* zygosity as in (G). For (A), (D), (F), and (H), error bars represent 95th percentile CI around adjusted HR estimate.

as in Figure 5G, $n = 385$ patients with evaluable zygosity)³⁸ or anti-PD(L1) immunotherapy alone (IO-cohort, $n = 638$ patients with evaluable zygosity).⁵¹ In the chemo-IO cohort, after adjusting for Eastern Cooperative Oncology Group (ECOG) performance status, programmed death ligand 1 (PD-L1) expression, TMB, derived neutrophil to lymphocyte ratio (dNLR), smoking history, and histology, only the patients with tumors harboring biallelic alterations in *KEAP1* demonstrated significantly worse PFS compared with those with *KEAP1* WT tumors (mPFS of 2.7 vs. 6.7 months, $HR_{adj} = 1.8$, $p = 3.1 \times 10^{-5}$) (Figures 6E and 6F). No difference in PFS between patients with tumors harboring monoallelic *KEAP1* alterations vs. WT was observed (mPFS of 11 vs. 6.7 months, $HR_{adj} = 0.98$, $p = 0.96$). More striking differences in PFS by *KEAP1* zygosity were observed in the IO-cohort (Figures 5G and 5H). Although patients with *KEAP1* biallelic tumors showed significantly worse PFS compared with those with WT tumors (mPFS of 1.8 vs. 2.7 months, $HR_{adj} = 1.6$, $p = 1.04 \times 10^{-4}$), monoallelic *KEAP1* inactivation was associated with favorable outcomes on IO when compared with the same WT tumors (mPFS of 11 vs. 2.7 months, $HR_{adj} = 0.6$, $p = 0.027$). In total, these data argue that the zygosity of *KEAP1* mutations, rather than *KEAP1* mutation status alone, may be an important prognostic biomarker of overall survival in lung cancer and, more importantly, may be a predictive biomarker of response to standard-of-care immunotherapies.

DISCUSSION

Loss-of-function somatic alterations to TSGs represent one of the two fundamental genetic events underlying oncogenesis. Although complete loss (biallelic inactivation) is the cornerstone for the two-hit model of TSG-mediated tumorigenesis, the extent to which this model applies to most TSGs has remained incompletely understood. Here, we used allele-specific analysis of somatic mutation and copy-number data to conduct a census of the incidence of diverse mechanisms of biallelic loss across 224 TSGs and 96 detailed cancer types.

Our results broadly stratified all TSGs into four classes based on their tendency for biallelic inactivation. The majority of Class 1 TSGs underwent near universal biallelic inactivation in every cancer type in which they were mutated, suggesting that loss of both alleles in these genes is obligate for impairing their tumor-suppressive effect. In contrast, Class 2 TSGs (e.g., *ATM*, *ARID1A*, and *FBXW7*), although broadly mutated across cancers, exhibited highly lineage-restricted and divergent preferences for retention of the WT allele across lineages. Although these genes may demonstrate haploinsufficiency or dosage sensitivity in some contexts,^{18,20,52–54} our data directly implicate selection for biallelic inactivation in at least some lineages and argue for a context-dependent role in oncogenesis for these tumor-suppressive genes. In contrast, Class 3 and 4 TSGs invariably showed a lack of selection for biallelic inactivation in every cancer type in which they were commonly mutated. Although biological mechanism of action for a majority of these genes remains to be understood, some (e.g., *PPP2R1A* and *SPOP*) have been shown to acquire dominant-negative oncogenic mutations that constitutively inhibit their WT alleles without requiring additional somatic hits to the gene locus.^{22,55} Interestingly, Class 4 genes demonstrated

a particular preference for retaining the WT allele, the biological rationale for which remains to be understood. Taken together, these data suggest that the propensity to lose the WT allele is dictated by the biochemical mode of tumor suppression.

Haploinsufficiency or dosage sensitivity of TSGs has been proposed to explain the retention of heterozygosity observed in many TSGs.¹⁰ Although our study was not designed to evaluate haploinsufficiency, our data nevertheless challenge how the evidence on dosage sensitivity of key TSGs that arose in murine models translates to human cancers. For example, heterozygous loss of *PTEN* has been suggested to be sufficient to promote tumor development in murine models of prostate cancer,⁸ breast cancer,⁵⁶ and astrocytomas.⁵⁷ However, in our patient cohort, we saw a strong selection for biallelic losses of *PTEN* in each of these diseases, suggesting a dependence on the complete loss of the protein. A plausible reconciliation of the two observations is that partial losses of *PTEN* are oncogenic early in tumor initiation (as observed in mice), but as tumors progress, there is an ultimate dependence on loss of WT.¹¹ Our findings thus argue for careful interpretation of haploinsufficiency in tumor suppressors in human cancers.

We demonstrated that signals of selection for biallelic inactivation of genes such as *APC* in unexpected lineages can reveal new insights into the progression of those cancers. Although Wnt pathway dysregulation in lung cancers has been shown to be associated with metastasis,⁵⁸ poor outcomes,⁵⁹ resistance to cisplatin,⁶⁰ and EGFR inhibitors,⁶¹ Wnt activation in this disease is often thought to be mediated by increased secretion of Wnt ligands by the surrounding “niche” cells.⁶² We report here that Wnt pathway mutations, although infrequent in lung cancers, show hallmarks of late-arising and biologically important mutations.

Finally, we also leveraged signatures of selection for biallelic inactivation to identify potentially functional alleles currently overlooked by contemporary, clinically operational frameworks for oncogenicity annotation (e.g., OncoKB). Although such frameworks combine information on mutational recurrence with functional predictions (e.g., whether a mutation is likely to introduce a truncating allele to a putative TSG), they face significant difficulty in classifying non-recurrent missense mutations. Using selection for biallelic inactivation, we identified several genes in which variants of uncertain significance show compelling evidence of function and/or patient outcomes that mirror those of oncogenic alleles, prompting their reclassification. Most importantly, we demonstrated that only *KEAP1*-mutant tumors with biallelic *KEAP1* inactivation, but not those that retained the WT allele, were associated with poor overall survival and response to standard-of-care therapies in lung cancer. Our discoveries here complement both prior observations in TSGs (*SMARCA4* LUAD⁵⁰), as well as in oncogenes (*KRAS*^{53,64}) that mutant allele dosage can mediate response to both targeted therapies, as well as chemo- and immunotherapies. These data therefore support the routine clinical reporting of loss of heterozygosity as a guide to interpretation of the clinical significance of TSG mutations.

Limitations of the study

This work employed targeted clinical tumor sequencing data to detect evidence of selection for biallelic inactivation. Although our clinical sequencing assay robustly captures intragenic

homozygous deletions, we have relatively reduced sensitivity to detect focal heterozygous losses. Nevertheless, future studies leveraging whole-genome sequencing of tumor specimens are uniquely poised to characterize and understand the role of LOH-only events in cancer. We focused on evidence of selection for the most common form of biallelic inactivation (MutLOH), but future studies should develop more comprehensive statistical models that integratively capture the likelihood of other modes of biallelic inactivation, such as homozygous deletion and composite mutation. Our data were largely derived from single-site sequencing, hindering the ability to accurately call subclonal biallelic inactivation that may arise in response to therapy or otherwise over the course of tumor evolution. Although our TSG classification was robust to varying criteria, we envision that future studies with larger cohorts, and especially those with tumor types underrepresented in our study, are required to completely resolve allele-specific patterns of inactivation in infrequently mutated TSGs and determine their class assignment.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Chaitanya Bandlamudi (bandlamc@mskcc.org).

Materials availability

All experimental materials used in this study are commercially available as detailed in the [key resources table](#). Unique materials developed for this study may be requested through the [lead contact](#).

Data and code availability

All TSG mutation and zygosity data for samples analyzed in this study are made available via [Table S1](#). The raw sequencing data for the MSK-IMPACT cohort are protected for privacy reasons and are not broadly available. However, raw data may be requested from bandlamc@mskcc.org and will require additional institutional approvals. Data and code to generate the figures is available at <https://github.com/samgould2/KEAP1-mutLOH-prime-editing-sensor> (prime editing screen) and https://github.com/reznik-lab/tsg_biallelic/ (all other analyses).

Gene and protein expression data from TCGA are available through the Genomic Data Commons (<https://portal.gdc.cancer.gov/>). TRACERx cohort data are from Frankell et al. (<https://doi.org/10.1038/s41586-023-05783-5>).

ACKNOWLEDGMENTS

We thank our patients and their families for participating in this study. We are grateful to the members of the MSK Molecular Diagnostics Service in the Department of Pathology and Laboratory Medicine who generated the clinically reported MSK-IMPACT genomic data. E.R. acknowledges vigorous and lifelong scientific discussions with Edita Reznik and Grigory Reznik. A.E. was supported by the Cancer Research Society Next Generation of Scientists Award, an American Society Young Investigator Award, and the Canadian Institutes of Health Research; R.T. was supported by NIH T32-CA009207 and an ASCO Young Investigator Award; S.I.G. was supported by NIH T32GM136540 and the MIT School of Science Fellowship in Cancer Research; and Y.R.M.-G. has received funding from the Andrew Sabin Family Foundation related to this work. F.J.S.-R. is an HHMI Hanna Gray Fellow and was supported by the V Foundation for Cancer Research (V2022-028), NCI Cancer Center Support grant P30-CA1405, the Ludwig Center at MIT (2036636), Koch Institute Frontier awards (2036648 and 2036642), and the MIT Research Support Committee (3189800). E.R. was supported by NIH R37 CA276200, DOD HT9425-23-1-0995, a Kidney Cancer Association Young Investigator Award, and the MSK Society. C.B. was supported by a Druckenmiller Center for Lung Cancer Research Award. This work was also supported by NIH/NCI

Cancer Center Support Grant P30 CA008748, NIH Award R01 CA227534, and the Marie Josée and Henry R. Kravis Center for Molecular Oncology.

AUTHOR CONTRIBUTIONS

Conceptualization, M.Z., M.A.P., E.R., and C.B.; methodology, M.Z., M.A.P., S.I.G., F.J.S.-R., E.R., and C.B.; software, M.Z., M.A.P., S.I.G., E.R., and C.B.; investigation and validation, M.Z., M.A.P., S.I.G., F.J.S.-R., E.R., and C.B.; formal analysis and visualization, M.Z., M.A.P., S.I.G., K.G., F.J.S.-R., E.R., and C.B.; data curation, M.Z., M.A.P., A.E., A.S., R.T., M.M., K.G., A.J.S., and C.B.; resources: S.I.G., F.J.S.-R., D.B.S., M.F.B., and J.C.; funding acquisition, E.R. and C.B.; writing – original draft, M.Z., M.A.P., E.R., and C.B.; writing – review and editing, M.Z., M.A.P., S.I.G., F.J.S.-R., E.R., C.B., D.B.S., N.S., M.F.B., and A.J.S. with input from all authors; supervision, E.R. and C.B.

DECLARATION OF INTERESTS

M.A.P. reports stock ownership in Amgen. Y.R.M.-G. reports travel, accommodation, and expenses from AstraZeneca and Loxo Oncology/Eli Lilly. She acknowledges honoraria from Virology Education and Projects in Knowledge (for a CME program funded by an educational grant from Amgen). She acknowledges associated research funding to the institution from Mirati Therapeutics, Loxo Oncology at Eli Lilly, Elucida Oncology, Taiho Oncology, Hengrui USA, Ltd/Jiangsu Hengrui Pharmaceuticals, Lutzana Biotechnology, Endeavor Biomedicines, and AbbVie. She is an employee of Memorial Sloan Kettering Cancer Center, which has an institutional interest in Elucida. She acknowledges royalties from Rutgers University Press and Wolters Kluwer. She acknowledges food/beverages from Endeavor Biomedicines. D.B.S. reports personal fees from Pfizer, Scorpion Therapeutics, FORE Therapeutics, Function Oncology, Fog Pharma, Elsie Biotechnologies, Rain Oncology, and BridgeBio outside the submitted work. M.F.B. declares consulting activity from AstraZeneca, Eli Lilly, and Paige AI. A.J.S. reports grants and personal fees from BMS, Merck, Iovance Biotherapeutics, and Amgen; personal fees from Johnson & Johnson, KSQL Therapeutics, Enara Bio, Perceptive Advisors, Oppenheimer and Co, Umoja Biopharma, Legend Biotech, Prelude Therapeutics, Immunocore, Lyell Immunopharma, and Heat Biologics; and grants from GSK, PACTpharma, Achilles Therapeutics, and Harpoon Therapeutics outside of the submitted work.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
 - Study cohort
 - Sample filtering for zygosity analysis
- [METHOD DETAILS](#)
 - Mutation profiling
 - Allele-specific copy number inference
 - Assessing biallelic inactivation at each gene locus
 - Mutational timing using patients with multiple biopsies
 - APC mutation zygosity in TRACERx cohort
 - Tumor suppressor gene classification
 - Plasmids
 - Prime editing sensor library design & cloning
 - Lentivirus production
 - Cell culture and cell line generation
 - Prime editing sensor screening protocol
 - Fluorescence-activated cell sorting and analysis
 - Genomic DNA extraction, library deconvolution, and sequencing
 - Sequence analysis
 - pegRNA-sensor analysis
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Assessing enrichment for biallelic inactivation
 - Gene and protein expression analysis in the TCGA
 - Survival analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.11.010>.

Received: November 3, 2023

Revised: August 14, 2024

Accepted: November 10, 2024

Published: December 18, 2024

REFERENCES

- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558. <https://doi.org/10.1126/science.1235122>.
- Knudson, A.G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* 68, 820–823. <https://doi.org/10.1073/pnas.68.4.820>.
- Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305, 779–784. <https://doi.org/10.1038/305779a0>.
- Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* 87, 159–170. [https://doi.org/10.1016/s0092-8674\(00\)81333-1](https://doi.org/10.1016/s0092-8674(00)81333-1).
- Shuin, T., Kondo, K., Torigoe, S., Kishida, T., Kubota, Y., Hosaka, M., Nagashima, Y., Kitamura, H., Latif, F., and Zbar, B. (1994). Frequent somatic mutations and loss of heterozygosity of the Von Hippel-Lindau tumor suppressor gene in primary human renal cell carcinomas. *Cancer Res.* 54, 2852–2855.
- Srinivasan, P., Bandlamudi, C., Jonsson, P., Kemel, Y., Chavan, S.S., Richards, A.L., Penson, A.V., Bielski, C.M., Fong, C., Syed, A., et al. (2021). The context-specific role of germline pathogenicity in tumorigenesis. *Nat. Genet.* 53, 1577–1585. <https://doi.org/10.1038/s41588-021-00949-1>.
- Jonsson, P., Bandlamudi, C., Cheng, M.L., Srinivasan, P., Chavan, S.S., Friedman, N.D., Rosen, E.Y., Richards, A.L., Bouvier, N., Selcuklu, S.D., et al. (2019). Tumour lineage shapes BRCA-mediated phenotypes. *Nature* 571, 576–579. <https://doi.org/10.1038/s41586-019-1382-1>.
- Kwabi-Addo, B., Giri, D., Schmidt, K., Podsypanina, K., Parsons, R., Greenberg, N., and Ittmann, M. (2001). Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression. *Proc. Natl. Acad. Sci. USA* 98, 11563–11568. <https://doi.org/10.1073/pnas.201167798>.
- Willis, A., Jung, E.J., Wakefield, T., and Chen, X. (2004). Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene* 23, 2330–2338. <https://doi.org/10.1038/sj.onc.1207396>.
- Berger, A.H., Knudson, A.G., and Pandolfi, P.P. (2011). A continuum model for tumour suppression. *Nature* 476, 163–169. <https://doi.org/10.1038/nature10275>.
- Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 155, 948–962. <https://doi.org/10.1016/j.cell.2013.10.011>.
- Donehower, L.A., Soussi, T., Korkut, A., Liu, Y., Schultz, A., Cardenas, M., Li, X., Babur, O., Hsu, T.-K., Lichtarge, O., et al. (2019). Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome. *Cell Rep.* 28, 1370–1384.e5. <https://doi.org/10.1016/j.celrep.2019.07.001>.
- Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* 2017, PO.17.00011. <https://doi.org/10.1200/PO.17.00011>.
- Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713. <https://doi.org/10.1038/nm.4333>.
- Gorelick, A.N., Sánchez-Rivera, F.J., Cai, Y., Bielski, C.M., Biederstedt, E., Jonsson, P., Richards, A.L., Vasan, N., Penson, A.V., Friedman, N.D., et al. (2020). Phase and context shape the function of composite oncogenic mutations. *Nature* 582, 100–103. <https://doi.org/10.1038/s41586-020-2315-8>.
- Yamulla, R.J., Nalubola, S., Flesken-Nikitin, A., Nikitin, A.Y., and Schimenti, J.C. (2020). Most Commonly Mutated Genes in High-Grade Serous Ovarian Carcinoma Are Nonessential for Ovarian Surface Epithelial Stem Cell Transformation. *Cell Rep.* 32, 108086. <https://doi.org/10.1016/j.celrep.2020.108086>.
- Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al. (2018). The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* 50, 645–651. <https://doi.org/10.1038/s41588-018-0078-z>.
- Shi, H., Tao, T., Abraham, B.J., Durbin, A.D., Zimmerman, M.W., Kadoc, C., and Look, A.T. (2020). ARID1A loss in neuroblastoma promotes the adrenergic-to-mesenchymal transition by regulating enhancer-mediated gene expression. *Sci. Adv.* 6, eaaz3440. <https://doi.org/10.1126/sciadv.aaz3440>.
- Wu, J.N., and Roberts, C.W.M. (2013). ARID1A mutations in cancer: another epigenetic tumor suppressor? *Cancer Discov.* 3, 35–43. <https://doi.org/10.1158/2159-8290.CD-12-0361>.
- Sun, X., Wang, S.C., Wei, Y., Luo, X., Jia, Y., Li, L., Gopal, P., Zhu, M., Nasrour, I., Chuang, J.-C., et al. (2017). Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. *Cancer Cell* 32, 574–589.e6. <https://doi.org/10.1016/j.ccr.2017.10.007>.
- Zhang, Y., Kwok-Shing Ng, P., Kucherlapati, M., Chen, F., Liu, Y., Tsang, Y.H., de Velasco, G., Jeong, K.J., Akbani, R., Hadjipanayis, A., et al. (2017). A pan-cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell* 31, 820–832.e3. <https://doi.org/10.1016/j.ccr.2017.04.013>.
- Haesen, D., Abbasi Asbagh, L., Derua, R., Hubert, A., Schrauwen, S., Hoorn, Y., Amant, F., Waelkens, E., Sablina, A., and Janssens, V. (2016). Recurrent PPP2R1A Mutations in Uterine Cancer Act through a Dominant-Negative Mechanism to Promote Malignant Cell Growth. *Cancer Res.* 76, 5719–5731. <https://doi.org/10.1158/0008-5472.CAN-15-3342>.
- Coulouarn, C., Factor, V.M., Andersen, J.B., Durkin, M.E., and Thorgerisson, S.S. (2009). Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* 28, 3526–3536. <https://doi.org/10.1038/onc.2009.211>.
- Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* 43, 27–33. <https://doi.org/10.1038/ng.730>.
- Parolia, A., Cieslik, M., Chu, S.-C., Xiao, L., Ouchi, T., Zhang, Y., Wang, X., Vats, P., Cao, X., Pitchiaya, S., et al. (2019). Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* 571, 413–418. <https://doi.org/10.1038/s41586-019-1347-4>.
- Adams, E.J., Karthaus, W.R., Hoover, E., Liu, D., Gruet, A., Zhang, Z., Cho, H., Diloreto, R., Chhangawala, S., Liu, Y., et al. (2019). FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. *Nature* 571, 408–412. <https://doi.org/10.1038/s41586-019-1318-9>.
- Ho, D.W.H., Chan, L.K., Chiu, Y.T., Xu, I.M.J., Poon, R.T.P., Cheung, T.T., Tang, C.N., Tang, V.W.L., Lo, I.L.O., Lam, P.W.Y., et al. (2017). TSC1/2 mutations define a molecular subset of HCC with aggressive behaviour and treatment implication. *Gut* 66, 1496–1506. <https://doi.org/10.1136/gutjl-2016-312734>.
- Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in

- colon cancer by mutations in beta-catenin or APC. *Science* 275, 1787–1790. <https://doi.org/10.1126/science.275.5307.1787>.
29. Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B., et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455, 1069–1075. <https://doi.org/10.1038/nature07423>.
 30. Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* 48, 607–616. <https://doi.org/10.1038/ng.3564>.
 31. Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt signaling in cancer. *Oncogene* 36, 1461–1473. <https://doi.org/10.1038/onc.2016.304>.
 32. Cooper, A.J., Sequist, L.V., and Lin, J.J. (2022). Third-generation EGFR and ALK inhibitors: mechanisms of resistance and management. *Nat. Rev. Clin. Oncol.* 19, 499–514. <https://doi.org/10.1038/s41571-022-00639-9>.
 33. Sartore-Bianchi, A., Martini, M., Molinari, F., Veronese, S., Nichelatti, M., Artale, S., Di Nicolantonio, F., Saletti, P., De Dosso, S., Mazzucchelli, L., et al. (2009). PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res.* 69, 1851–1857. <https://doi.org/10.1158/0008-5472.CAN-08-2466>.
 34. Watson, P.A., Arora, V.K., and Sawyers, C.L. (2015). Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* 15, 701–711. <https://doi.org/10.1038/nrc4016>.
 35. Martínez-Jiménez, F., Muñoz, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., et al. (2020). A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* 20, 555–572. <https://doi.org/10.1038/s41568-020-0290-x>.
 36. Taguchi, K., and Yamamoto, M. (2017). The KEAP1-NRF2 System in Cancer. *Front. Oncol.* 7, 85. <https://doi.org/10.3389/fonc.2017.00085>.
 37. Ricciuti, B., Arbour, K.C., Lin, J.J., Vajdi, A., Vokes, N., Hong, L., Zhang, J., Tolstorukov, M.Y., Li, Y.Y., Spurr, L.F., et al. (2022). Diminished Efficacy of Programmed Death-(Ligand)1 Inhibition in STK11- and KEAP1-Mutant Lung Adenocarcinoma Is Affected by KRAS Mutation Status. *J. Thorac. Oncol.* 17, 399–410. <https://doi.org/10.1016/j.jtho.2021.10.013>.
 38. Alessi, J.V., Elkrief, A., Ricciuti, B., Wang, X., Cortellini, A., Vaz, V.R., Lamberti, G., Frias, R.L., Venkatraman, D., Fulgenzi, C.A.M., et al. (2023). Clinicopathologic and Genomic Factors Impacting Efficacy of First-Line Chemotherapy in Advanced NSCLC. *J. Thorac. Oncol.* 18, 731–743. <https://doi.org/10.1016/j.jtho.2023.01.091>.
 39. Wyler, E., Franke, V., Menegatti, J., Kocks, C., Boltengagen, A., Praktikno, S., Walch-Rückheim, B., Bosse, J., Rajewsky, N., Grässer, F., et al. (2019). Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. *Nat. Commun.* 10, 4878. <https://doi.org/10.1038/s41467-019-12894-z>.
 40. Yan, J., Oyler-Castrillo, P., Ravisankar, P., Ward, C.C., Levesque, S., Jing, Y., Simpson, D., Zhao, A., Li, H., Yan, W., et al. (2024). Improving prime editing with an endogenous small RNA-binding protein. *Nature* 628, 639–647. <https://doi.org/10.1038/s41586-024-07259-6>.
 41. Gong, M., Li, Y., Ye, X., Zhang, L., Wang, Z., Xu, X., Shen, Y., and Zheng, C. (2020). Loss-of-function mutations in KEAP1 drive lung cancer progression via KEAP1/NRF2 pathway activation. *Cell Commun. Signal.* 18, 98. <https://doi.org/10.1186/s12964-020-00568-z>.
 42. Zagorski, J.W., Turley, A.E., Dover, H.E., VanDenBerg, K.R., Compton, J.R., and Rockwell, C.E. (2013). The Nrf2 activator, tBHQ, differentially affects early events following stimulation of Jurkat cells. *Toxicol. Sci.* 136, 63–71. <https://doi.org/10.1093/toxsci/kft172>.
 43. Li, W., and Kong, A.-N. (2009). Molecular mechanisms of Nrf2-mediated antioxidant response. *Mol. Carcinog.* 48, 91–104. <https://doi.org/10.1002/mc.20465>.
 44. Gould, S.I., Wuest, A.N., Dong, K., Johnson, G.A., Hsu, A., Narendra, V.K., Atwa, O., Levine, S.S., Liu, D.R., and Sánchez Rivera, F.J. (2024). High-throughput evaluation of genetic variants with prime editing sensor libraries. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-024-02172-9>.
 45. Simon, D.A., Tálas, A., Kulcsár, P.I., Biczók, Z., Krausz, S.L., Váradyi, G., and Welker, E. (2022). PEAR, a flexible fluorescent reporter for the identification and enrichment of successfully prime edited cells. *eLife* 11, e69504. <https://doi.org/10.7554/eLife.69504>.
 46. Gould, S.I. (2024). Prime editing sensors enable multiplexed genome editing. *Nat. Rev. Genet.* 25, 454. <https://doi.org/10.1038/s41576-024-00737-7>.
 47. Saleh, M.M., Scheffler, M., Merkelbach-Bruse, S., Scheel, A.H., Ulmer, B., Wolf, J., and Buettnner, R. (2022). Comprehensive Analysis of TP53 and KEAP1 Mutations and Their Impact on Survival in Localized- and Advanced-Stage NSCLC. *J. Thorac. Oncol.* 17, 76–88. <https://doi.org/10.1016/j.jtho.2021.08.764>.
 48. Zavitsanou, A.-M., Pillai, R., Hao, Y., Wu, W.L., Bartnicki, E., Karakousi, T., Rajalingam, S., Herrera, A., Karatza, A., Rashidfarrokhi, A., et al. (2023). KEAP1 mutation in lung adenocarcinoma promotes immune evasion and immunotherapy resistance. *Cell Rep.* 42, 113295. <https://doi.org/10.1016/j.celrep.2023.113295>.
 49. Jamaspishvili, T., Berman, D.M., Ross, A.E., Scher, H.I., De Marzo, A.M., Squire, J.A., and Lotan, T.L. (2018). Clinical implications of PTEN loss in prostate cancer. *Nat. Rev. Urol.* 15, 222–234. <https://doi.org/10.1038/nrurol.2018.9>.
 50. Schoenfeld, A.J., Bandlamudi, C., Lavery, J.A., Montecalvo, J., Namakydoust, A., Rizvi, H., Egger, J., Concepcion, C.P., Paul, S., Arcila, M.E., et al. (2020). The Genomic Landscape of SMARCA4 Alterations and Associations with Outcomes in Patients with Lung Cancer. *Clin. Cancer Res.* 26, 5701–5708. <https://doi.org/10.1158/1078-0432.CCR-20-1825>.
 51. Thummalapalli, R., Ricciuti, B., Bandlamudi, C., Muldoon, D., Rizvi, H., Elkrief, A., Luo, J., Alessi, J.V., Pecci, F., Lamberti, G., et al. (2023). Clinical and Molecular Features of Long-term Response to Immune Checkpoint Inhibitors in Patients with Advanced Non-Small Cell Lung Cancer. *Clin. Cancer Res.* 29, 4408–4418. <https://doi.org/10.1158/1078-0432.CCR-23-1207>.
 52. Kemp, Z., Rowan, A., Chambers, W., Wortham, N., Halford, S., Sieber, O., Mortensen, N., von Herbay, A., Gunther, T., Ilyas, M., et al. (2005). CDC4 mutations occur in a subset of colorectal cancers but are not predicted to cause loss of function and are not associated with chromosomal instability. *Cancer Res.* 65, 11361–11366. <https://doi.org/10.1158/0008-5472.CAN-05-2565>.
 53. Davis, H., and Tomlinson, I. (2012). CDC4/FBXW7 and the “just enough” model of tumorigenesis. *J. Pathol.* 227, 131–135. <https://doi.org/10.1002/path.4004>.
 54. Akhoondi, S., Sun, D., von der Lehr, N., Apostolidou, S., Klotz, K., Maljkova, A., Cepeda, D., Fiegl, H., Dafou, D., Marth, C., et al. (2007). FBXW7/hCDC4 is a general tumor suppressor in human cancer. *Cancer Res.* 67, 9006–9012. <https://doi.org/10.1158/0008-5472.CAN-07-1320>.
 55. Ma, J., Shi, Q., Cui, G., Sheng, H., Botuyan, M.V., Zhou, Y., Yan, Y., He, Y., Wang, L., Wang, Y., et al. (2021). SPOP mutation induces replication over-firing by impairing Geminin ubiquitination and triggers replication catastrophe upon ATR inhibition. *Nat. Commun.* 12, 5779. <https://doi.org/10.1038/s41467-021-26049-6>.
 56. Alimonti, A., Carracedo, A., Clohessy, J.G., Trotman, L.C., Nardella, C., Egia, A., Salmena, L., Sampieri, K., Haveman, W.J., Brogi, E., et al. (2010). Subtle variations in Pten dose determine cancer susceptibility. *Nat. Genet.* 42, 454–458. <https://doi.org/10.1038/ng.556>.
 57. Kwon, C.-H., Zhao, D., Chen, J., Alcantara, S., Li, Y., Burns, D.K., Mason, R.P., Lee, E.Y.-H.P., Wu, H., and Parada, L.F. (2008). Pten haploinsufficiency accelerates formation of high-grade astrocytomas. *Cancer Res.* 68, 3286–3294. <https://doi.org/10.1158/0008-5472.CAN-07-6867>.
 58. Nguyen, D.X., Chiang, A.C., Zhang, X.H.-F., Kim, J.Y., Kris, M.G., Ladanyi, M., Gerald, W.L., and Massagué, J. (2009). WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* 138, 51–62. <https://doi.org/10.1016/j.cell.2009.04.030>.

59. Stewart, D.J. (2014). Wnt signaling pathway in non-small cell lung cancer. *J. Natl. Cancer Inst.* 106, djt356. <https://doi.org/10.1093/jnci/djt356>.
60. Zhang, J., Liu, J., Li, H., and Wang, J. (2016). β -Catenin signaling pathway regulates cisplatin resistance in lung adenocarcinoma cells by upregulating Bcl-xL. *Mol. Med. Rep.* 13, 2543–2551. <https://doi.org/10.3892/mmr.2016.4882>.
61. Casás-Selves, M., Kim, J., Zhang, Z., Helfrich, B.A., Gao, D., Porter, C.C., Scarborough, H.A., Bunn, P.A., Chan, D.C., Tan, A.C., et al. (2012). Tankyrase and the canonical Wnt pathway protect lung cancer cells from EGFR inhibition. *Cancer Res.* 72, 4154–4164. <https://doi.org/10.1158/0008-5472.CAN-11-2848>.
62. Tammela, T., Sanchez-Rivera, F.J., Cetinbas, N.M., Wu, K., Joshi, N.S., Helenius, K., Park, Y., Azimi, R., Kerper, N.R., Wesselhoeft, R.A., et al. (2017). A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. *Nature* 545, 355–359. <https://doi.org/10.1038/nature22334>.
63. Bielski, C.M., Donoghue, M.T.A., Gadiya, M., Hanrahan, A.J., Won, H.H., Chang, M.T., Jonsson, P., Penson, A.V., Gorelick, A., Harris, C., et al. (2018). Widespread Selection for Oncogenic Mutant Allele Imbalance in Cancer. *Cancer Cell* 34, 852–862.e4. <https://doi.org/10.1016/j.ccr.2018.10.003>.
64. Burgess, M.R., Hwang, E., Mroue, R., Bielski, C.M., Wandler, A.M., Huang, B.J., Firestone, A.J., Young, A., Lacap, J.A., Crocker, L., et al. (2017). KRAS Allelic Imbalance Enhances Fitness and Modulates MAP Kinase Dependence in Cancer. *Cell* 168, 817–829.e15. <https://doi.org/10.1016/j.cell.2017.01.020>.
65. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120. <https://doi.org/10.1038/ng.2764>.
66. Frankell, A.M., Dietzen, M., Al Bakir, M., Lim, E.L., Karasaki, T., Ward, S., Veeriah, S., Collier, E., Huebner, A., Bunkum, A., et al. (2023). The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* 616, 525–533. <https://doi.org/10.1038/s41586-023-05783-5>.
67. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* 2, 401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095>.
68. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6, pl1. <https://doi.org/10.1126/scisignal.2004088>.
69. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
70. Shen, R., and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* 44, e131. <https://doi.org/10.1093/nar/gkw520>.
71. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554. <https://doi.org/10.1186/s13059-014-0554-4>.
72. Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M.D., Wendl, M.C., and Ding, L. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 30, 1015–1016. <https://doi.org/10.1093/bioinformatics/btt755>.
73. Middha, S., Zhang, L., Nafa, K., Jayakumaran, G., Wong, D., Kim, H.R., Sadowska, J., Berger, M.F., Delair, D.F., Shia, J., et al. (2017). Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precis. Oncol.* 2017, PO.17.00084. <https://doi.org/10.1200/PO.17.00084>.
74. AACR Project GENIE Consortium (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* 7, 818–831. <https://doi.org/10.1158/2159-8290.CD-17-0151>.
75. Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z.Y., Won, H.H., Scott, S.N., et al. (2015). Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J. Mol. Diagn.* 17, 251–264. <https://doi.org/10.1016/j.jmoldx.2014.12.006>.
76. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
77. Vázquez-García, I., Uhrlitz, F., Ceglia, N., Lim, J.L.P., Wu, M., Mohibullah, N., Niyazov, J., Ruiz, A.E.B., Boehm, K.M., Bojilova, V., et al. (2022). Ovarian cancer mutational processes drive site-specific immune evasion. *Nature* 612, 778–786. <https://doi.org/10.1038/s41586-022-05496-1>.
78. Grigoriadis, K., Huebner, A., Bunkum, A., Collier, E., Frankell, A.M., Hill, M.S., Thol, K., Birkbak, N.J., Swanton, C., Zaccaria, S., et al. (2024). CONI-PHER: a computational framework for scalable phylogenetic reconstruction with error correction. *Nat. Protoc.* 19, 159–183. <https://doi.org/10.1038/s41596-023-00913-9>.
79. Akama-Garren, E.H., Joshi, N.S., Tammela, T., Chang, G.P., Wagner, B.L., Lee, D.-Y., Rideout, W.M., Papagiannakopoulos, T., Xue, W., and Jacks, T. (2016). A Modular Assembly Platform for Rapid Generation of DNA Constructs. *Sci. Rep.* 6, 16836. <https://doi.org/10.1038/srep16836>.
80. Sánchez-Rivera, F.J., Diaz, B.J., Kastenhuber, E.R., Schmidt, H., Katti, A., Kennedy, M., Tem, V., Ho, Y.-J., Leibold, J., Paffenholz, S.V., et al. (2022). Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat. Biotechnol.* 40, 862–873. <https://doi.org/10.1038/s41587-021-01172-3>.
81. Wiznerowicz, M., and Trono, D. (2003). Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* 77, 8957–8961. <https://doi.org/10.1128/jvi.77.16.8957-8951.2003>.
82. Meier, J.A., Zhang, F., and Sanjana, N.E. (2017). GUIDES: sgRNA design for loss-of-function screens. *Nat. Methods* 14, 831–832. <https://doi.org/10.1038/nmeth.4423>.
83. Morgens, D.W., Wainberg, M., Boyle, E.A., Ursu, O., Araya, C.L., Tsui, C.K., Haney, M.S., Hess, G.T., Han, K., Jeng, E.E., et al. (2017). Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* 8, 15178. <https://doi.org/10.1038/ncomms15178>.
84. Nelson, J.W., Randolph, P.B., Shen, S.P., Everette, K.A., Chen, P.J., Anzalone, A.V., An, M., Newby, G.A., Chen, J.C., Hsu, A., et al. (2022). Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* 40, 402–410. <https://doi.org/10.1038/s41587-021-01039-7>.
85. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37, 224–226. <https://doi.org/10.1038/s41587-019-0032-3>.
86. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadiy, S., Liu, D.L., Kantheti, H.S., Saghatelian, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 173, 321–337.e10. <https://doi.org/10.1016/j.cell.2018.03.035>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human tumor and matched normal (blood) samples	This paper	N/A
Chemicals, peptides, and recombinant proteins		
10% FBS	Gibco	Cat. #26140079
1X Penicillin-Streptomycin	Sigma-Aldrich	Cat. #P4333
Blasticidin S	ThermoFisher	Cat. #A1113903
Lipofectamine 2000	Invitrogen	Cat. #11668030
Polybrene Transfection Reagent	Sigma-Aldrich	Cat. #TR-1003
Puromycin	Sigma-Aldrich	Cat. #P7255
RPMI-1640 media	Gibco	Cat. #11875093
Critical commercial assays		
DNeasy Blood & Tissue Kit	Qiagen	Cat. #69504
MiSeq Reagent Kit v2	Illumina	Cat. #MS-102-2003
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	Cat. #E2621
Q5 High Fidelity 2X Master Mix	New England Biolabs	Cat. #M0492S
QIAquick PCR Purification Kit	Qiagen	Cat. #28104
QIAquick Gel Extraction Kit	Qiagen	Cat. #28704
Deposited data		
Gene and protein expression data from TCGA	Cancer Genome Atlas Research Network et al. ⁶⁵ ; Genomic Data Commons	https://doi.org/10.1038/ng.2764 ; https://portal.gdc.cancer.gov/
TRACERx cohort	Frankell et al. ⁶⁶	https://doi.org/10.1038/s41586-023-05783-5
Experimental models: Cell lines		
Human (male): NCI-H1299	Koch Institute ES Cell Core	RRID:CVCL_0060
Oligonucleotides		
F1 primer 5'-CGCTCTTCCGATCTCTA GCGTCGAGTTAGGAATT-3'	IDT	N/A
F2 primer 5'-AATGATAACGGCGACCACCGAGAT CTACACCGCTCTTCCGATCTCTAGCGT-3'	IDT	N/A
R1 primer 5'-CTGAACCGCTCTTCCGATCTTG TGGAAAGGACGAAACACC-3'	IDT	N/A
R2 primer 5'-CAAGCAGAACAGACGGCATACGAGA TNNNNNNNNCCTGCTAACCGCTCTCCGATCT-3'	IDT	N/A
Index primer 5'-CGGTGTTCGTCCTTCCACAA AGATCGGAAGAGCGGTTACGCAGG-3'	IDT	N/A
Read 1 primer 5'-ACCGCTCTTCCGATCTCTA GCGTCGAGTTAGGAATT-3'	IDT	N/A
Read 2 primer 5'-CCTGCTGAACCGCTCTTCC GATCTTGTGGAAAGGACGAAACACCG-3'	IDT	N/A
Oligonucleotide library	Twist Biosciences	N/A
Recombinant DNA		
pCMV-PE7	Yan et al. ⁴⁰	Addgene #214812
pMD2.G	Addgene	Cat. #12259
pREP-8xARE-GFP-SV40-BFP	Wyler et al. ³⁹	Addgene #134910
psPAX2	Addgene	Cat. #12260

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lenti-Trono-BR	Gould et al. ⁴⁴	N/A
Lenti-PEAR-mCherry	Gould et al. ⁴⁴	N/A
Lenti-UPEmS-tevo	Gould et al. ⁴⁴	N/A
Lenti-EF1a-PE7-P2A-Puro	This paper	N/A
Lenti-Trono-8xARE-GFP-EFS-Neo-P2A-BFP	This paper	N/A
Software and algorithms		
BD FACSDiva v9.0	BD Biosciences	N/A
cBioPortal	Cerami et al. ⁶⁷ ; Gao et al. ⁶⁸	https://www.cbiportal.org/
DESeq2 v.1.38.3	Love et al. ⁶⁹	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
FACETS	Shen and Seshan ⁷⁰	https://github.com/mskcc/facets-suite
fastq-join (version 1.3.1)	https://doi.org/10.2174/1875036201307010001	https://github.com;brwnj/fastq-join
FlowJo 10.9.0	BD Biosciences	N/A
logistf (v1.26.0)	https://cran.r-project.org/web/packages/logistf/index.html	https://doi.org/10.32614/CRAN.package.logistf
MAGECK (v0.5.9.5)	Li et al. ⁷¹	https://sourceforge.net/projects/mageck/
MSIsensor	Niu et al. ⁷²	https://github.com/ding-lab/msisensor
OncokB	Chakravarty et al. ¹³	https://github.com/oncokb/oncokb
PEGG (version 2.1.0)	Gould, et al. ⁴⁴	https://github.com/samgould2/PEGG2.0
Python (version 3.9.12)	Python Software Foundation	http://www.python.org
R (version 4.2.2)	https://cran.r-project.org/	N/A
survival (version 3.7-0)	https://cran.r-project.org/web/packages/survival/	https://doi.org/10.32614/CRAN.package.survival

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study cohort

Our cohort comprised solid tumor specimens from 48,179 patients across 67 major cancer types and 498 detailed histologies that were biopsied and sequenced as part of routine clinical care at Memorial Sloan Kettering Cancer Center between November 2013 and August 2021. Sequencing was performed using the Food and Drug Administration (FDA)-authorized MSK-IMPACT (Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets) assay. The study was approved by the MSKCC Institutional Review Board (IRB), and all patients provided written informed consent for tumor sequencing and review of medical records. At the time of sequencing, 56% of the patients had active metastatic disease. The patients included in the cohort (N=23,713; see below) were 54% female and 71% non-Hispanic White (self-reported), with a median age of 58 years at diagnosis.

Sample filtering for zygosity analysis

We applied several exclusion criteria to retain samples suitable for robust zygosity analysis across tumor suppressors and cancer types. First, all tumor samples with purity less than 30% were excluded from this analysis. Second, all tumors with copy number fits that did not pass the copy number fit (FACETS) quality thresholds described below were excluded. Third, tumors with 'extremely' high tumor mutational burden in each cancer type were excluded. Since, the distribution of TMB varies across cancer types, we determined that a 90th percentile threshold of TMB in each cancer type represents a reasonable cut-off above which tumors may harbor mutational processes that may confound the zygosity analysis. Tumors with microsatellite instability defined by MSIsensor score ≥ 10 were also excluded.^{72,73} Fourth, similar to TMB, we reasoned that some tumors may harbor a high number of focal amplifications and deletions that may be introduced by high levels of genomic instability. Therefore, all patients with a total of 20 or more focal amplifications/deletions were excluded from this analysis. Finally, we limited our analysis to only the cancer types in which we had 50 or more tumors pre-filtering. After applying all filters, 23,713 patients (49% of the cohort) were retained for zygosity analysis in this study. TSG somatic alteration and zygosity data are included in Tables S3, S4, and S5. Nearly all (99.8%, 23,666 of 23,713) of patients are included in the AACR-GENIE public cohort.⁷⁴

We also reasoned that high rates of loss of heterozygosity at loci harboring germline pathogenic variants in cancer predisposition genes may confound observations associated with evaluating biallelic inactivation associated with somatic mutations.⁶ We therefore excluded specific gene and tumor sample pairs in which a pathogenic germline variant was identified. We restrict this exclusion to

cancer predisposition genes that have either high or moderate penetrance.⁶ Germline variant calling for all genes was performed as previously described using a clinically validated pipeline.⁷⁵ Pathogenicity was assessed using an approach guided by American College of Medical Genetics and Genomics (ACMG) criteria which incorporated a combination of expert curation in a clinical setting, database annotations (ClinVar), and *in silico* predictions for functional impact and allele frequencies in healthy individuals (Mehine et al., unpublished data).⁶ In all, we identified a pathogenic variant in one of 44 high or moderate penetrance genes in 1,558 patients. When assessing zygosity or signals of selection for LOH associated with somatic mutations in each individual TSG, patients with germline pathogenic variants in the corresponding gene were excluded from consideration.

METHOD DETAILS

Mutation profiling

Matched tumor and normal sequencing was performed using the MSK-IMPACT clinical sequencing assay that profiles up to 505 cancer genes across four versions of the assay (IM3, 341 genes, n=2,302 patients; IM5, 410 genes, n=8,202; IM6, 468 genes, n=30,586; IM7, 505 genes, n=7,089).⁷⁵ All samples were required to have at least 100x sequencing coverage and a purity (pathologist estimated or FACETS derived, see below) of at least 30%. The median tumor purity among the 23,713 tumor specimens analyzed in this study was 54% (interquartile range of 41% to 69%). For patients with multiple biopsies, the specimen sequenced with the latest version of the panel or with the most recently sequenced biopsy was chosen. Tumor and matched blood normal specimens were sequenced to a median depth of 615X and 484X, respectively. All somatic alterations including substitutions, small insertions/deletions, gene-level focal amplifications/homozygous deletions and, in select genes, structural rearrangements were identified using a clinically validated pipeline.¹⁴ Somatic alterations were assessed to be ‘oncogenic’ or ‘driver’ if they were annotated as oncogenic or likely oncogenic by OncoKB. An alteration is defined as being at least likely oncogenic in OncoKB based on any of multiple sources of evidence suggesting its role in promoting cell proliferation or other hallmarks of cancer (<https://sop.oncokb.org>).^{13,76} Briefly, for every somatic mutation identified on the MSK-IMPACT clinical assay, evidence for oncogenicity in OncoKB is curated from *in silico* predictions of known cancer hotspots, experimental data reported in prior studies, clinical response or resistance to targeted therapies and based on whether it has a clear biological effect.¹³ Of note, all loss of functions mutations in tumor suppressor genes are also considered to be oncogenic. All oncoprints and lollipop plots were created using the cBioPortal.^{67,68} TSGs (n = 224) were assigned to pathways according to the table below.

Pathway	Gene list
TP53	ATM, TP53
Cell cycle	CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C, RB1
Epigenetic	ARID1A, ARID1B, ARID2, ASXL1, ASXL2, CTCF, DAXX, DNMT3A, DNMT3B, EED, EZH2, MEN1, NSD1, PBRM1, SETD2, SMARCA4, SMARCB1, SUZ12, TET1, TET2, KMT2A, KMT2B, KMT2C, KMT2D
PI3K	INPP4B, PIK3R1, PIK3R2, PIK3R3, PPP2R1A, PTEN, STK11, TSC1, TSC2
WNT	APC, AXIN1, AXIN2, RNF43, TCF7L2
TGF-Beta	SMAD2, SMAD3, SMAD4, TGFBRI1, TGFBRI2
NOTCH	CREBBP, EP300, FBXW7, NCOR1, NOTCH1, NOTCH2, NOTCH3, NOTCH4, SPEN
RTK-RAS	CBL, NF1, RASA1, ERF1, ERF, SPRED1
HIPPO	FAT1, LATS1, LATS2, NF2
DDR	ATR, BARD1, BLM, BRCA2, BRIP1, CHEK1, CHEK2, ERCC2, ERCC3, ERCC4, ERCC5, FANCA, FANCC, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PARP1, PMS1, PMS2, POLE, RAD50, RAD51, RAD51B, RAD51C, RAD51D, RECQL, BRCA1, XRCC2, RAD21, POLD1, MSH3, NTHL1, RECQL, RTEL1, SLX4, TP53BP1
NRF2	CUL3, KEAP1
MYC	MAX, MGA
unassigned	ALOX12B, ARID5B, B2M, BAP1, BBC3, BCL2L11, BMPR1A, CASP8, CBFB, CDC73, CDH1, CDK12, CIC, DICER1, DIS3, EPCAM, EPHA3, EPHB1, ETV6, FH, FLCN, FOXA1, FOXL2, FOXP1, FUBP1, GATA3, GRIN2A, HNF1A, IFNGR1, JAK1, KLF4, MAP2K4, MAP3K1, MITF, MYOD1, NKX3-1, NPM1, PAX5, PHOX2B, PMAIP1, PRDM1, PTCH1, PTPRD, PTPRS, PTPRT, RUNX1, RYBP, SDHA, SDHAF2, SDHB, SDHC, SDHD, SHQ1, SOCS1, SOX17, SOX9, SPOP, SUFU, TBX3, TMEM127, TNFAIP3, TNFRSF14, TOP1, TP63, VHL, WT1, ID3, BCL10, INHA, MST1, HLA-A, EPHA7, INHBA, PPP6C, SH2B3, FOXO1, NFKBIA, ZFHXB3, ANKRD11, GPS2, HOXB13, TCF3, CEBPA, CYLD, DUSP4, ELF3, HLA-B, INPPL1, SESN1, SESN2, SESN3, APLNR, ARHGAP35, CMTR2, CTR9, ETAA1, FOXF1, HLA-C, LZTR1, MAD2L2, MTAP, POT1, REST, SERPINB3, SETDB1, SLFN11, SMARCA2, SMARCE1, SPRTN, TRIP13, ZNRF3

Allele-specific copy number inference

Allele-specific copy number status at each gene locus was determined using FACETS algorithm (v0.5.14)⁷⁰ in a two-step approach described previously⁶³ (<https://github.com/mskcc/facets-suite>). Briefly, the first run (low-sensitivity, cval=100) determined the log-ratio corresponding to the diploid state (total copy number of 2) of the tumor genome. Using this diploid log-ratio, a second run is generated in a high sensitivity mode (cval = 50) to infer locus specific copy number state for each gene. All copy number fits inferred via this approach were subjected to a series of filters to identify and exclude samples with low confidence fits. These include: insufficient evidence supporting the diploid state, hypersegmentation, large fraction of the genome identified as homozygous deletions, has very high ploidy (eg: > 7), has invalid purity estimates such as NA or FACETS default of 0.3, high fraction of the genome identified as being subclonal, or, has fraction of the genome where the integer copy number estimate is discordant with the allelic imbalance observed with variant allele log odds ratio (<https://github.com/taylor-lab/facets-preview>). Copy number segments from FACETS calls with a minor copy number of 0 are identified as LOH events. Focal deletions are defined as copy number segments smaller than 10 megabases in size with a total copy number of 1 and harboring 10 or fewer genes.⁷⁷ Homozygous deletions were detected using a clinical validated pipeline as previously described.⁷⁵ Briefly, genes with mean segment fold-change (ratio of normalized sequencing depth in tumor to normal) of -2 or less at one or more exons with an adjusted statistical significance of p-value < 0.05 were considered as harboring homozygous deletions. Statistical significance for a homozygous deletion call is determined by comparing the fold-change of the candidate segment to the observed distribution of segments that are clustered around the segment fold-change of 1.

Assessing biallelic inactivation at each gene locus

Each gene locus is considered to be biallelically inactivated if it harbored (1) a homozygous deletion, (2) an oncogenic mutation and a structural rearrangement, (3) two somatic loss of function mutations, or (4) an oncogenic mutation with a concomitant copy-number loss of heterozygosity (MutLOH, see below) (Figure 1A).

A gene locus was considered to be MutLOH if any of the three criteria calculated from the allele-specific copy number inference using FACETS. First, if the total copy number (*tcn*) at the locus with an oncogenic mutation was estimated to be 1 (minor copy number, *mcn*, has to be 0). Second, if the *mcn* was 0 and the *tcn* was greater than 1, the mutation was required to be in at least two copies for the locus to be MutLOH. To determine this, we first calculated the expected variant allele frequency (VAF) of the oncogenic mutation as if it were present on only one of the copies:

$$(\Phi * tcn) / (2 * (1 - \Phi) + (\Phi * 1) + (\Phi * (tcn - 1))), \text{ where } \Phi \text{ is the purity of the tumor.}$$

Then, the locus was classified as MutLOH if the observed VAF of the oncogenic mutation is greater than the upper bound of the 95th percentile binomial confidence interval of this expected VAF if the oncogenic mutation was present on only one copy. Third, if the observed VAF of the oncogenic mutation was consistent with it being present on all copies of the locus. We calculated the expected VAF of the oncogenic mutation as if it was present on *tcn* copies as:

$(\Phi * tcn) / (2 * (1 - \Phi) + (\Phi * tcn))$. Then, the locus was considered to be MutLOH if the observed VAF of the oncogenic mutation is greater than the lower bound of the 95th percentile binomial confidence interval of the expected VAF if it was present on all copies.

All other loci with at least one somatic mutation but those that were not biallelic by the criteria described above are considered as ‘heterozygous.’ Remaining loci were considered as ‘wild-type’. These include all heterozygous copy number loss (LOH) events (*mcn*=0) as their oncogenic effect cannot be established to distinguish from events that are simply a consequence of genomic instability. Exemplifying this, focal LOH events arise at similar frequencies in both oncogenes and tumor suppressors (Figure S1).

Mutational timing using patients with multiple biopsies

We identified all lung, colon, rectal and prostate adenocarcinoma patients whose tumors were profiled multiple times using MSK-IMPACT targeted sequencing. All tumor specimens were required to have at least 300x coverage and a TMB that is less than 20 nonsynonymous mutations per megabase. Tumor specimens in which no mutations were identified using our clinical sequencing were excluded from this analysis. All patients with germline pathogenic mutations or those with multiple independently diagnosed cancer types were also excluded. Only specimens that shared at least one somatic mutation were considered to be clonally related and eligible for analysis. For somatic mutations not seen in every biopsy of the patient, we reasoned that a subset may still be detectable at levels below clinical sequencing thresholds. We therefore re-genotyped all mutations observed in any one of each patient’s biopsies in all of their sequenced tumor biopsies. We then considered a mutation to be present in a biopsy if there were at least 5 reads supporting the variant allele. For patients with more than two clonally related biopsies, we preferentially selected the two biopsies in which the genes of interest (*APC*, *EGFR*, *PIK3CA*, *CTNNB1*, *SPOP* and *AR*) were present in one but not the other. This is particularly relevant for truncal mutations such as *APC*/colorectal, *EGFR*/lung and *SPOP*/prostate where the driver mutations are seen in all of the biopsies of every patient with these mutations.

APC mutation zygosity in TRACERx cohort

The TRACERx cohort comprises 421 untreated, early-stage non-small cell lung cancer patients.⁶⁶ Bulk multi-region whole exome sequencing (WES) was performed on primary tumors, and any associated lymph node lesions that were sampled at the time of primary surgical resection. Bulk WES data is processed through a bioinformatics pipeline in order to infer mutations present in each tumor region, and tumor subclonal architecture and tumor phylogenies are reconstructed using the method CONIPHER.⁷⁸

CONIPHER infers mutation clusters, their prevalence in each tumor region, and their evolutionary ordering. Mutations themselves are classified as being ‘truncal’ if their associated mutation cluster has been assigned to the trunk node of the reconstructed phylogenetic tree. Mutations assigned to any other mutation cluster are classified as ‘subclonal’. Loss of heterozygosity at the *APC* locus was inferred for 378 tumors evaluable for copy-number analysis. In all, 15 of 378 tumors had at least one truncal *APC* loss of function (LOF) mutation (two tumors were observed with two truncal *APC* LOF mutations each). The LOH rate among *APC* mutant tumors was 73% (11 of 15) while only about half (48%, 173/363) of *APC* wild-type tumors were observed with an LOH (two-sided Fisher’s exact test, p=0.065).

Tumor suppressor gene classification

TSGs were grouped into four classes based on their selection patterns across cancer types as described in Figure S3A. Within a given cancer type, a given gene was considered to exhibit ‘positive selection’ for biallelic inactivation in a given cancer type if either there was a statistically significant enrichment for MutLOH (Figure 3) or the overall biallelic rate (Figure 2) was 80% or higher. Only enrichment scores for cancer types in which the TSG was mutated in 20 or more tumors, or biallelic rates for cancer types in which the TSG was altered (including mutations, homozygous deletions or composite mutations) in 20 or more tumors were considered for classification. TSGs were considered to undergo ‘negative selection’ for biallelic inactivation when there was significant depletion of MutLOH at the cancer-type or pan-cancer level. TSGs were determined to exhibit ‘no selection’ when no significant enrichment for MutLOH (at the cancer-type level) or depletion (at the cancer-type or pan-cancer level) was observed. Genes were then classified as: Class 1, if there was positive selection in >=80% of evaluated cancer types and no negative selection; Class 2, if there was positive selection in at least one cancer type; Class 3, if there was absence of positive or negative selection in any cancer type; and finally, Class 4, if no positive selection was observed in any cancer type but negative selection was observed at either the cancer type or the pan-cancer level. In all, 66 TSGs were assigned to one of the four classes with all other TSGs remaining unclassified (Table S3A). See Methods S1 and Figure S3 for detailed evaluation of the criteria used in the TSG classification.

Plasmids

All new plasmids were generated using Gibson Assembly strategies⁷⁹ using NEBuilder® HiFi DNA Assembly Master Mix (NEB #E2621) following the manufacturer’s protocol. All new plasmids, along with detailed maps and sequences, will be made available through Addgene. The sequence for PE7 that was used to generate the Lenti-EF1a-PE7-P2A-Puro plasmid was obtained from pCMV-PE7 (Addgene #214812). The NRF2 reporter plasmid was constructed by first transferring the U6-sgRNA-EFS-Blast-P2A-BFP cassette from pUSEBB⁸⁰ into the higher titer pLV backbone,⁸¹ then replacing the blasticidin S selection marker with a neomycin selection marker via Gibson Assembly to generate Lenti-Trono-Neo-P2A-BFP, and finally transferring the 8xARE-GFP cassette from pREP-8xARE-GFP-SV40-BFP (Addgene #134910) into the Lenti-Trono-Neo-P2A-BFP backbone, replacing the U6-sgRNA cassette to generate the final Lenti-Trono-8xARE-GFP-EFS-Neo-P2A-BFP reporter plasmid. Libraries of pegRNA-sensors were cloned into the Lenti-Trono-BR backbone, which was previously generated by transferring the U6-sgRNA-EFS-Blast-P2A-TurboRFP cassette from pUSEBR into the higher titer pLV backbone.⁴⁴ Lenti-PEAR-mCherry, the reporter plasmid used for validating prime editing activity, was also previously assembled.⁴⁴ The Lenti-UPEmS-tevo plasmid was used to assemble individual pegRNAs via Golden Gate Assembly for validation experiments.⁴⁴ All plasmids were validated via whole-plasmid sequencing.

Prime editing sensor library design & cloning

Prime editing sensor libraries were designed using the Python package PEGG (version 2.1.0).⁴⁴ As input, we provided the 42 most frequently observed VUS in *KEAP1* (40/42 were prime editing amenable), top 10 most frequent driver mutations (10/10 prime editing amenable), and 10 germline mutations (9/10 prime editing amenable), as well as 4 hotspot variants in NRF2, and silent substitutions (SNPs or oligonucleotide substitutions) at each mutated codon location. We generated 5 pegRNAs per mutation for each missense variant (when suitable NGG PAM sequences were present), prioritizing diversity among protospacers in generating these pegRNA designs. We generated 2 pegRNAs for each silent substitution mutation, and included 41 non-targeting controls, whose protospacers have no target site in the genome,⁸² and 50 safe-targeting controls, who target intergenic regions that lack annotated genomic elements.⁸³ The pegRNA-sensor cassette included a 60 nucleotide long sensor site in the reverse complement orientation relative to the protospacer, and also contained the tevopreneQ1 motif to improve pegRNA stability.⁸⁴ The pegRNA-sensor cassettes were filtered to exclude EcoRI and Esp3I sites and polyT termination sequences of greater than or equal to 4 consecutive thymines. The full details and selection criteria for the pegRNAs can be found in the associated GitHub repository: <https://github.com/samgould2/KEAP1-mutLOH-prime-editing-sensor>. The oligonucleotide library was ordered from Twist Biosciences. The library was cloned following the same library cloning protocol presented in Gould et al.⁴⁴

Lentivirus production

Lentiviruses were produced by co-transfection of HEK293T cells with the relevant lentiviral transfer vector and packaging vectors psPAX2 (Addgene, catalog no. 12260) and pMD2.G (Addgene, catalog no. 12259) using Lipofectamine 2000 (Invitrogen, catalog no. 11668030). Viral supernatants were collected at 48- and 72-hours post-transfection and stored at -80°C.

Cell culture and cell line generation

NCI-H1299 cells were received from Koch Institute ES Cell Core, where they were mycoplasma tested and STR validated. The NCI-H1299 cells were maintained in RPMI-1640 media (Gibco, catalog no. 11875093) supplemented with 10% FBS and 1X Penicillin-Streptomycin (ThermoFisher). To generate NCI-H1299-PE7 cells via spinfection, 500,000 NCI-H1299 were plated in each well of a 6-well plate with 2 mL of fresh Lenti-EF1a-PE7-P2A-Puro plasmid lentivirus and polybrene transfection reagent (Sigma-Aldrich, catalog no. TR-1003) was added to a final concentration of 8 µg/mL, and the cells were spun at 800 g for 2 hours before being incubated overnight. The following day, this spinfection protocol was repeated to maximize prime editor expression, the NCI-H1299-PE7 cells were pooled, selected with 10 µg/mL puromycin, and validated for prime editing activity. Subsequently, to generate NCI-H1299-PE7-8xARE-GFP cells, the NCI-H1299-PE7 cells were transduced with Lenti-Trono-8xARE-GFP-EFS-Neo-P2A-BFP. After 3 days, we sorted the BFP positive population to isolate cells expressing this reporter construct. These cells were used for the subsequent screen.

Prime editing sensor screening protocol

For each of the three replicates, we plated one million NCI-H1299-PE7-8xARE-GFP cells per well in three 6-well plates in RPMI-1640 media with 10 µg/mL puromycin, resulting in 17 million plated cells per replicate to achieve 10,000X representation assuming an MOI of 0.3. We then added 50 µL of the prime editing sensor lentivirus to achieve a final volume of 3 mL in each well. After 24 hours, cells from each replicate were lifted, combined, and plated in five 15-cm plates, and 10 µg/mL blasticidin S was added to the media. A small population of cells from each replicate was split and not treated with blasticidin S to experimentally determine the MOI. After 72 hours, we measured the RFP positive fraction of this unselected population to quantify the MOI, which we calculated to be 0.342 (~30% of cells transduced). Puromycin and blasticidin S were maintained at 10 µg/mL throughout the screen for the rest of the cells. Cells were replated every 3 days, maintaining ≥ 10,000X representation at each time-point during the screen. Fourteen days post-transduction, we generated a 10,000X representation cell pellet from each replicate (5 million cells), and then sorted the remaining cells from each replicate into 4 equally sized bins on the basis of their GFP fluorescence level.

Fluorescence-activated cell sorting and analysis

The BD FACSCelesta Cell Analyzer in tube or plate reader format, with BD FACSDiva v9.0 software used for data collection, was used for the validation of *NRF2* 8xARE-GFP reporter construct, the Lenti-PEAR-mCherry prime editing reporter, and validation of individual pegRNAs for ARE-GFP reporter activity. Downstream analysis was performed using FlowJo 10.9.0 to identify single cells and quantify fluorescence. The BD FACSAria III Cell Sorter was used for cell sorting.

Genomic DNA extraction, library deconvolution, and sequencing

Genomic DNA (gDNA) was extracted from each sample using the DNEasy blood and tissue kit (Qiagen) following the manufacturer's protocol. We then performed two rounds of PCR to amplify the pegRNA-sensor cassette and add barcodes and Illumina adaptors for next-generation sequencing. In the first round of PCR, we performed 4 PCR reactions for each sample, using 20 µL of gDNA, 25 µL of Q5 High Fidelity 2X Master Mix (NEB), and 2.5 µL of the forward (F1) and reverse primers (R1), which were at a concentration of 10 µM. The plasmid library was amplified using 10 ng of the plasmid library as a template. The sequence of the F1 primer was 5'-CGCTCTTCCGATCTCTAGCGTTCGAGTTAGGAATT-3' and the R1 primer was 5'-CTGAACCGCTTCCGATCTTGTTGGAAAG GACGAAACACC-3'. The PCR program for the first PCR amplification reaction was (1) 98°C x 2 minutes, [(2) 98°C x 10 seconds, (3) 60°C x 30 seconds, (4) 72°C x 30 seconds] x 25 cycles, (5) 72°C x 2 minutes, (6) 4°C Hold. We then PCR purified the samples, pooling samples together, using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's protocol, before gel extracting the appropriately sized band with the QIAquick Gel Extraction Kit (Qiagen), again following the manufacturer's protocol. The PCR1 products were measured with a NanoDrop 2000 (ThermoFisher) and were then used as template for PCR2. We then performed the second round of PCR amplification (PCR2), running two PCR reactions per sample. Each reaction contained 10 ng of purified PCR1 product as template, 25 µL of Q5 High Fidelity 2X Master Mix, 0.5 µL of the forward (F2) and reverse primers (R2), which were at a concentration of 10 µM, and were completed to 50 µL with nuclease-free water. The sequence of the F2 primer was 5'-AAT GATACGGCGACCACCGAGATCTACACCGCTTCCGATCTAGCGT-3', and the R2 primer was 5'-CAAGCAGAACGGCA TACGAGATNNNNNNNNN CCTGCTGAACCGCTTCCGATCT-3', with a different eight nucleotide barcode used to identify each sample. The PCR program for the second PCR amplification reaction (PCR2) was (1) 98°C x 2 minutes, [(2) 98°C x 10 seconds, (3) 67°C x 30 seconds, (4) 72°C x 30 seconds] x 10 cycles, (5) 72°C x 2 minutes, (6) 4°C Hold. These PCR reactions were pooled for each sample, PCR purified and subsequently gel extracted, before being run on an Illumina MiSeq v2 sequencer with the 2x250 paired end sequencing kit. The sequencing reaction used the following custom primers: Read 1: 5'-ACCGCTTCCGATCTC TAGCGTTCGAGTTAGGAATT-3', Read 2: 5'-CCTGCTGAACCGCTTCCGATCTTGTTGGAAAGGACGAAACACCG-3', Index 1: 5'-CGGTGTTCGTCTTCCACAAAGATCGGAAGAGCGGTTAGCAGG-3'.

Sequence analysis

Sequences from different samples were sorted into different fastq files on the basis of their 8 nucleotide barcodes in the index 1 read. Subsequently, reads from each sample were joined using the fastq-join (1.3.1) algorithm with the default parameters enabled (<https://doi.org/10.2174/1875036201307010001>). To identify the pegRNA counts in each sample, we filtered reads below an average Phred

quality score of 30 and matched (with no mismatches allowed) the protospacer and 3' extension regions of the read to the known sequences of each pegRNA in our library. For each sample, we sorted the 60 nucleotide sensor site from each read into a different fastq corresponding to each pegRNA. We used the last 8 nucleotides of each sensor site, which uniquely identified each sensor, to determine whether recombination of the sensor site had occurred, filtering out sensors without a perfect match in this region. We then used Crispresso2 to quantify the editing at each sensor site for each pegRNA in each sample,⁸⁵ using HDR mode with a quantification window center of 5-55. The full sequence analysis pipeline is available at the following GitHub repository: <https://github.com/samgould2/KEAP1-mutLOH-prime-editing-sensor>

pegRNA-sensor analysis

We used MAGeCK (v0.5.9.5)⁷¹ to normalize read counts between samples and determine the log₂ fold-change (LFC) of each pegRNA in each sorted bin (Q1-Q4) relative the pre-sort populations, using the paired-end mode with the non-targeting control guides designated as controls. We then filtered pegRNAs with a normalized control count <30 reads to reduce spuriously enriching pegRNAs. We further removed noise from this dataset by filtering out pegRNAs that exhibited enrichment (LFC>0.1) in all of the sorted bins (Q1-Q4) relative to the pre-sort populations, as these pegRNAs likely represented PCR-amplification bias. When quantifying editing, we limited our analysis to pegRNA-sensor pairs with at least 50 sensor reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Assessing enrichment for biallelic inactivation

Evaluation of selection for biallelic inactivation within TSGs and cancer types was limited to only MutLOH. Tumors with homozygous deletions, structural rearrangements (fusions) and multiple loss of function mutations (composite mutations) in TSGs were excluded from this analysis as robustly assessing the null distribution of these types of events is often intractable without whole-genome sequencing. Enrichment for MutLOH was measured by comparing the rate of LOH among tumors with oncogenic loss-of-function mutations in a given TSG to those that were wild-type using a multivariable logistic regression model that accounted for disease status (primary/metastatic), tumor mutational burden and fraction of genome altered as the measure of genomic instability. To mitigate the problem of sparse-data bias, we performed the logistic regression by applying Firth's bias reduction using the "logistf" function in the *logistf* R package (v1.26.0) (doi: 10.32614/CRAN.package.logistf). Enrichment was measured for each TSG at pan-cancer as well as cancer type level. Only the associations in which the TSG is mutated in at least 3 tumors in a given cancer type are evaluated. Genes on sex chromosomes were excluded from this analysis. In all, including pan-cancer associations, 1550 associations were evaluated for signals of selection for MutLOH among oncogenic loss of function mutations in TSGs (Table S3A). Correction for multiple hypothesis testing was performed using the Benjamini-Hochberg method. Similar regression model accounting for TMB, FGA and disease status was adopted to evaluate selection for MutLOH in VUSs (Figure 5). Only tumors with nonsynonymous substitutions that are VUSs and those with no other oncogenic loss of function alteration in the evaluated gene were considered for this analysis. In all 210 gene and cancer types with 3 or more VUS mutations were evaluated (Table S3D).

For pathway analysis in Figure S2B, we specifically asked if the mutations in a given pathway in a given cancer type are statistically significantly enriched or depleted for harboring a loss of heterozygosity event. Our pathway definitions and gene composition were derived from the TCGA pan-cancer study that investigated oncogenic signaling pathways in cancer.⁸⁶ Pathway annotations for each of the 224 tumor suppressor genes are included in STAR Methods. As with analysis at gene-level (Figure 3), we restricted this analysis to cancer types with 50 or more mutated tumors and with at least 3 mutations in a given pathway. In all, after adjusting for tumor mutational burden, fraction of genome altered and disease status, of the 502 pathway and cancer type pairs we evaluated, 248 were significantly enriched (n=246) or depleted (n=2) for MutLOH (adjusted p-value < 0.05; Table S3B; Figure S2B).

To evaluate patterns of selection for MutLOH between primary and metastatic tumors, we conducted screens for selection for MutLOH independently in primary (n=9,789) and metastatic (n=13,337) tumor cohorts using a logistic regression model that accounted for FGA and TMB. All p-values were adjusted using the Benjamini Hochberg method. We limited this screen to 144 gene and cancer type pairs in which the respective genes are mutated in at least 10 tumors each of primary and metastasis patients in that disease (Table S3C; Figure S4). Of these 144, 59 pairs (41%) were selected for MutLOH in both primary and metastatic cohorts while 31 pairs (22%) were statistically significant for selection for MutLOH in only the primary (n=11) or metastatic (n=21) cohorts of the corresponding cancer type. For 27 of 31 pairs, the differences in MutLOH rates between primary and metastatic tumors was less than 20% highlighting the difficulty in biologically interpreting the differences in selection for MutLOH (Figure S4B).

Gene and protein expression analysis in the TCGA

Gene expression data (transcript counts) for 485 lung adenocarcinoma patients in TCGA was obtained from Genomic Data Commons.⁶⁵ We used the Bioconductor package DESeq2⁶⁹ to perform differential expression analysis. We compared resultant normalized counts for *CTNNB1* between *APC*^{WT}/*CTNNB1*^{WT} and *APC*^{Biallelic}/*CTNNB1*^{WT} using the Wilcoxon test to ascertain statistical significance. For protein expression analysis, we downloaded RPPA TCGA data for 351 patients with LUAD, and compared protein expression for β-catenin between *APC*^{WT}/*CTNNB1*^{WT} and *APC*^{Biallelic}/*CTNNB1*^{WT} patients also with the Wilcoxon test. Same gene and protein expression data from lung adenocarcinoma patients in TCGA were used to evaluate differences in expression of key genes across tumors with different *KEAP1* mutation classes (VUS vs. oncogenic). We used the Wilcoxon test to compare normalized

counts for gene expression analysis and RPPA values for protein expression between: (1) *KEAP1^{VUS}* and *KEAP1^{WT}* patients, and (2) *KEAP1^{Oncogenic}* and *KEAP1^{WT}* patients, for NRF2 and NQO1.

Survival analyses

Overall survival (OS) was measured as the time from the initial date of tumor sequencing to the date of last follow-up or death. For the OS analysis, all patients in our cohort with their biopsied specimens harboring a tumor purity of 15% or higher were considered. Patients with tumors exhibiting high tumor mutation burden (>90th percentile by cancer type) or copy number alteration burden (>=20 focal amplifications or deletions) as described above, were excluded. For patients with multiple sequenced biopsies, the earliest collected specimen with a FACETS fit that passed all QC criteria (see above) was used. All multivariate models of OS were performed using Cox proportional hazards models accounting for primary/metastatic disease status, sex, age at diagnosis, presence of an OncoKB Level 1 actionable alteration, FGA, TMB, and MSI score. Patients who were missing data for any covariates were excluded from multivariate models. For evaluating differences in outcomes between either mutation class (Figures 5F and S5) or by zygosity (Figures 6A and 6B), we considered only the cancer types with at least 100 or more total patients. For the VUS versus WT analysis (Figure S5), each evaluated gene was required to harbor at least 10 oncogenic mutations and 10 VUSs in that cancer type, with at least 5 OS events in each group (n=290 evaluable gene and cancer type pairs). Similarly, for the outcomes by zygosity analysis in Figure 6, we required the gene to harbor at least 10 tumors each with monoallelic and biallelic alterations and 5 patients with OS events in each group (n=150 evaluable pairs). Individual multivariate models were constructed for each pair with sufficient sample size, and p-values were adjusted using FDR correction based on the number of evaluated pairs. In Figure 6, for all genes and cancer type pairs evaluated, except for *KEAP1* in LUAD, only oncogenic alterations are considered. Based on multiple lines of evidence demonstrating that *KEAP1* VUSs in LUAD phenocopy oncogenic mutations, both types of *KEAP1* alterations (oncogenic and VUS) in LUAD are considered as ‘Altered’ when evaluating differences in outcomes by *KEAP1* zygosity. For Figure 6A, the biallelic group also includes tumors with homozygous deletions and composite mutations, and the altered group includes tumors with both monoallelic and biallelic alterations.

Progression-free survival (PFS) on first line chemoimmunotherapy (chemo/IO) was evaluated on 556 patients with NSCLC treated at MSKCC.³⁸ After excluding tumors from patients with pathogenic germline mutations, high TMB or FGA, or insufficient purity (<15%) as described above, as well as those without any mutations, PFS in 421 patients was evaluated for differences by mutation class (oncogenic, VUS or wild-type) (Figure 5G). After excluding patients for whom clinical data was missing for any covariate, N=378 patients were included in the multivariate model of PFS by mutation class (Figure S5C). PFS differences by zygosity was evaluated in 429 patients in whom zygosity was evaluable (Figure 6E), with N=385 in the multivariate model (Figure 6F). Similarly, of 923 patients with NSCLC treated with first line immunotherapy (IO) at MSK, PFS was evaluated on 654 patients with *EGFR^{WT}* tumors for which zygosity was evaluable⁵¹ (Figure 6G). N=638 were included in the multivariate model (Figure 6H) after excluding patients for whom covariate data was missing. Progression-free survival (PFS) for both lines of treatment was measured as the time from start of treatment to progression or death. For patients that did not progress, the time of last disease assessment was used as the censoring time. PFS by *KEAP1* mutation status or zygosity was evaluated using Cox proportional hazards models accounting for sex, ECOG (Eastern Cooperative Oncology Group) performance status, PD-L1 level, tumor mutation burden (TMB), derived neutrophil to lymphocyte ratio (dNLR), smoking pack years, and histology.

All survival analyses were performed using the survival package in R (version 3.7-0).

Supplemental figures

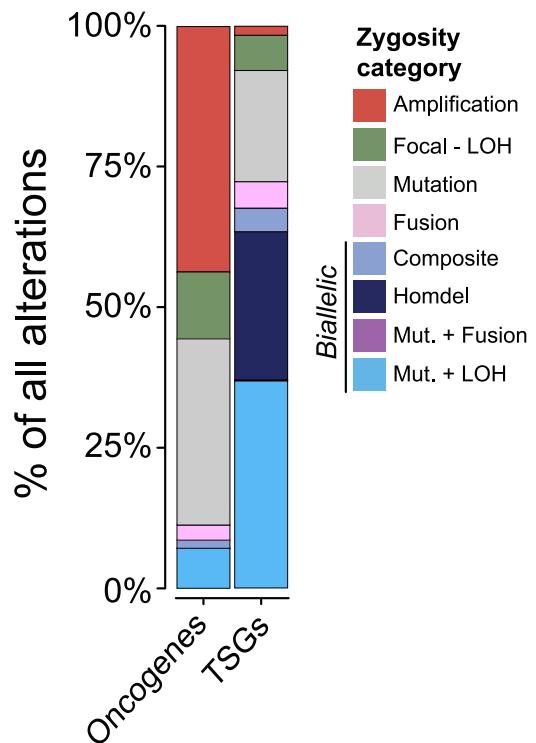
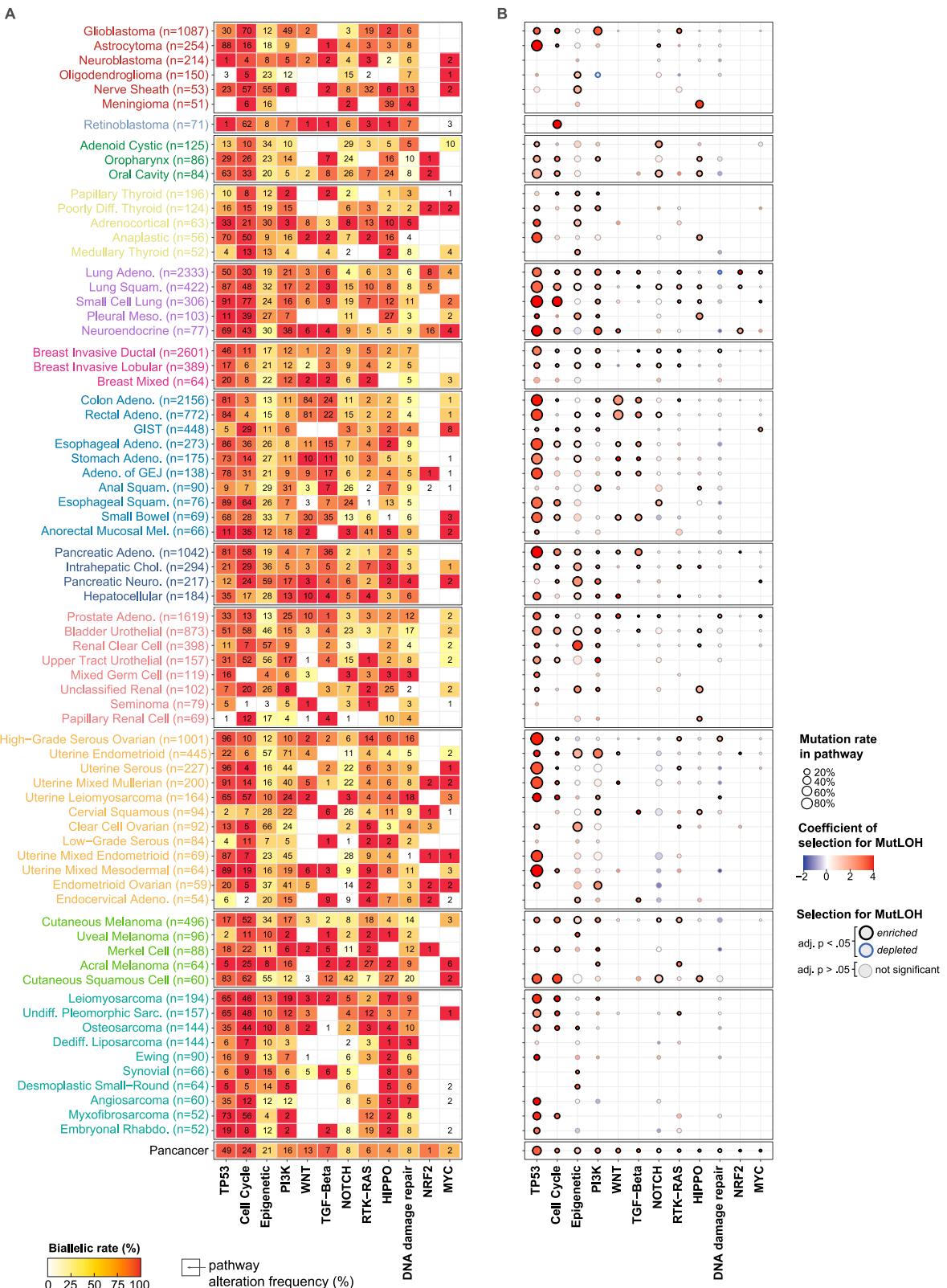


Figure S1. Zygosity changes associated with somatic alterations in oncogenes and tumor suppressors, related to Figure 1

Focal-LOH events are determined as described in [STAR Methods](#). Focal copy-number deletions were observed at similar rates in both oncogenes and tumor suppressors. Of note, Focal-LOH events include only those without any other somatic oncogenic alterations in that gene. Any Focal-LOH events in tumors that also harbor a concomitant loss-of-function mutation are also considered as Mut. + LOH (see [STAR Methods](#)).



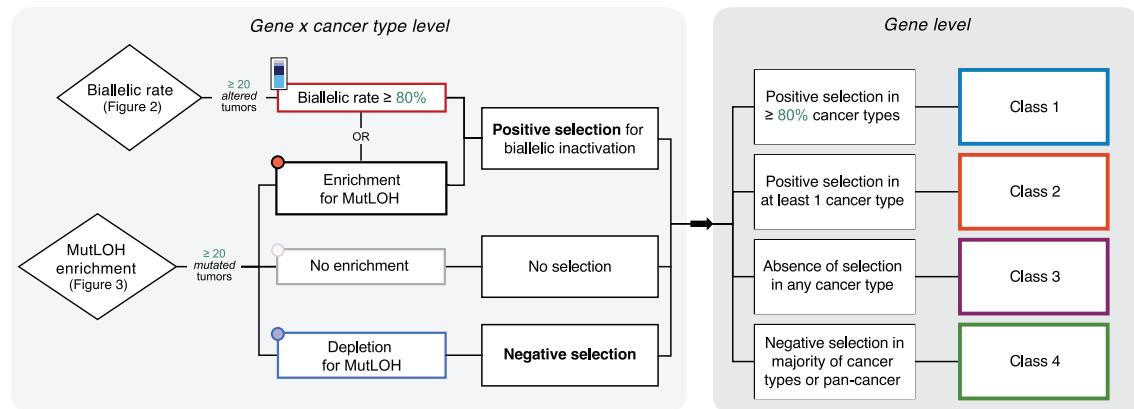
(legend on next page)

Figure S2. Pathway-level patterns of biallelic inactivation among tumor suppressor genes across cancers, related to Figures 2 and 3

(A) Overall alteration rate of each pathway across cancer types is shown. See [STAR Methods](#) for gene and pathway assignment. A pathway is considered altered if any one of the genes in the pathway in that tumor type harbors a loss-of-function mutation. Tumors with fusion events in the corresponding genes were excluded in this figure as their biallelic status cannot be robustly ascertained. Pathway alteration rates less than 1% are not shown. For a complete list of biallelic rates of pathways across cancers, refer to [Table S2C](#).

(B) Evidence for selection for MutLOH among mutations within a given pathway in each cancer type is shown. Similar to the analysis in [Figure 3](#), all tumors with fusions, homozygous deletions, and composite mutations were excluded from consideration when determining enrichment of selection for MutLOH among mutations in a given pathway in a given disease. The circles with black outlines denote those pathways and cancer types in which we find statistically significant (adjusted p value < 0.05) enrichment for MutLOH, whereas blue outlines denote depletion. For a complete list of pathway-level enrichment for MutLOH, refer to [Table S3B](#).

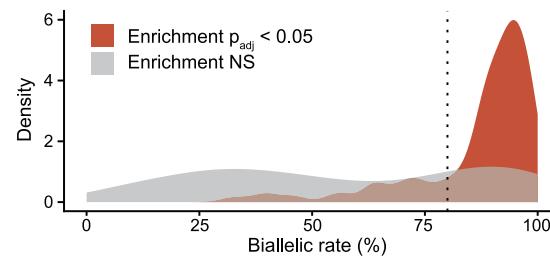
A



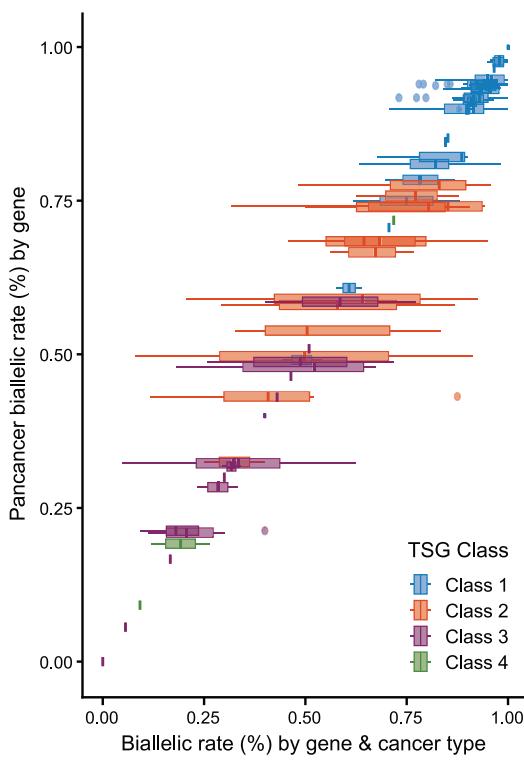
B

		Biallelic rate		
Enrichment for MutLOH		< 80%	≥ 80%	
Positive (FDR < 0.05)		27 (100%)	115 (75%)	142 (78%)
None (FDR ≥ 0.05)		0 (0%)	39 (25%)	39 (22%)
		27 (15%)	154 (85%)	N = 181

C



D



E

Biallelic rate threshold: 80%	Class 1 N = 32	Class 2 N = 13	Class 3 N = 18	Class 4 N = 3
70%	32 (100%)	1 (7.7%)	0 (0%)	0 (0%)
90%	0 (0%)	12 (92%)	2 (11%)	0 (0%)
	0 (0%)	0 (0%)	16 (89%)	0 (0%)
	0 (0%)	0 (0%)	0 (0%)	3 (100%)

F

Min. n: 20	Class 1	Class 2	Class 3	Class 4
70%	27 (84%)	1 (7.7%)	0 (0%)	0 (0%)
90%	5 (16%)	12 (92%)	4 (22%)	0 (0%)
	0 (0%)	0 (0%)	14 (78%)	0 (0%)
	0 (0%)	0 (0%)	0 (0%)	3 (100%)

G

Lineage threshold: 80%	Class 1	Class 2	Class 3	Class 4
70%	32 (100%)	1 (7.7%)	0 (0%)	0 (0%)
90%	0 (0%)	12 (92%)	0 (0%)	0 (0%)
	0 (0%)	0 (0%)	18 (100%)	0 (0%)
	0 (0%)	0 (0%)	0 (0%)	3 (100%)

n (%)

Figure S3. Tumor suppressor gene classification by biallelic patterns across cancer types, related to Figure 3

(A) Schematic describing the TSG classification approach (see STAR Methods). “Altered” tumors and biallelic rate includes homozygous deletions and composite mutations, as in Figure 2.

(B) Biallelic rate threshold and selection for MutLOH for the 181 gene and cancer-type pairs used in TSG classification that exhibited positive selection.

(C) Distribution of biallelic rate for all genes and cancer types evaluated for TSG classification, colored by whether significant selection for MutLOH was observed (adjusted p value < 0.05) or not.

(legend continued on next page)

-
- (D) Variation in distribution of biallelic inactivation rates by gene across cancer types (see [Methods S1](#)). Boxplots display median line, IQR boxes, and 1.5x IQR whiskers.
- (E) Confusion matrix of TSG class assignments ([Figure 3](#)) when varying the biallelic rate threshold to either 70% or 90%.
- (F) Changes to TSG class assignments ([Figure 3](#)) when varying the minimum mutation or alteration count threshold to either 10 or 30.
- (G) Changes to TSG class assignments ([Figure 3](#)) when varying the threshold for the fraction of cancer types exhibiting positive selection for a gene to be considered as Class 1.
- For (E)–(G), concordant class assignments are highlighted in green and discordant in red.

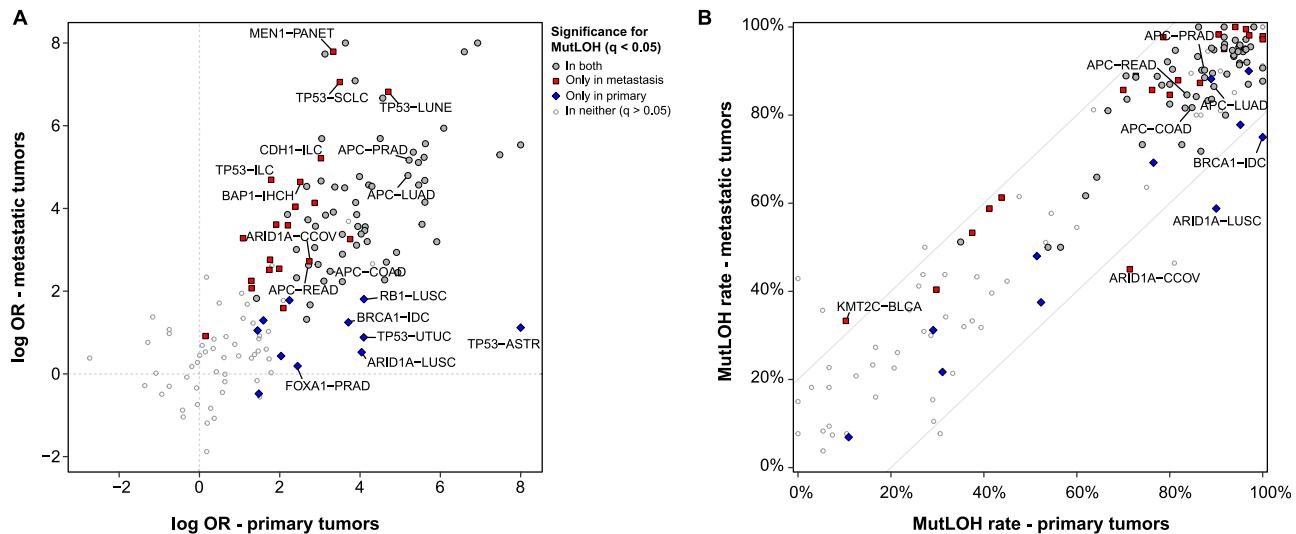


Figure S4. Differences in selection for MutLOH between primary and metastatic tumors, related to Figure 3

(A) Log-odds ratio from the logistic regression model evaluating selection for MutLOH after adjusting for fraction of genome altered and tumor mutational burden in primary ($n = 9,789$) and metastatic ($n = 13,337$) disease cohorts (see [STAR Methods](#)). Data points are labeled with gene and cancer type if the difference in log-odds ratio between primary and metastatic tumors is greater than 2 or if the differences in rates of MutLOH between the disease states differs by more than 20% or, finally, if the gene is *APC* (related to [Figure 5](#)). See [Table S1A](#) for cancer type definitions.

(B) Scatterplot showing MutLOH rates (i.e., proportion of mutated tumors that also harbor an LOH at that locus) in primary and metastatic tumors. All gene and cancer-type pairs between the two slanted lines have differences in MutLOH rates less than 20%.

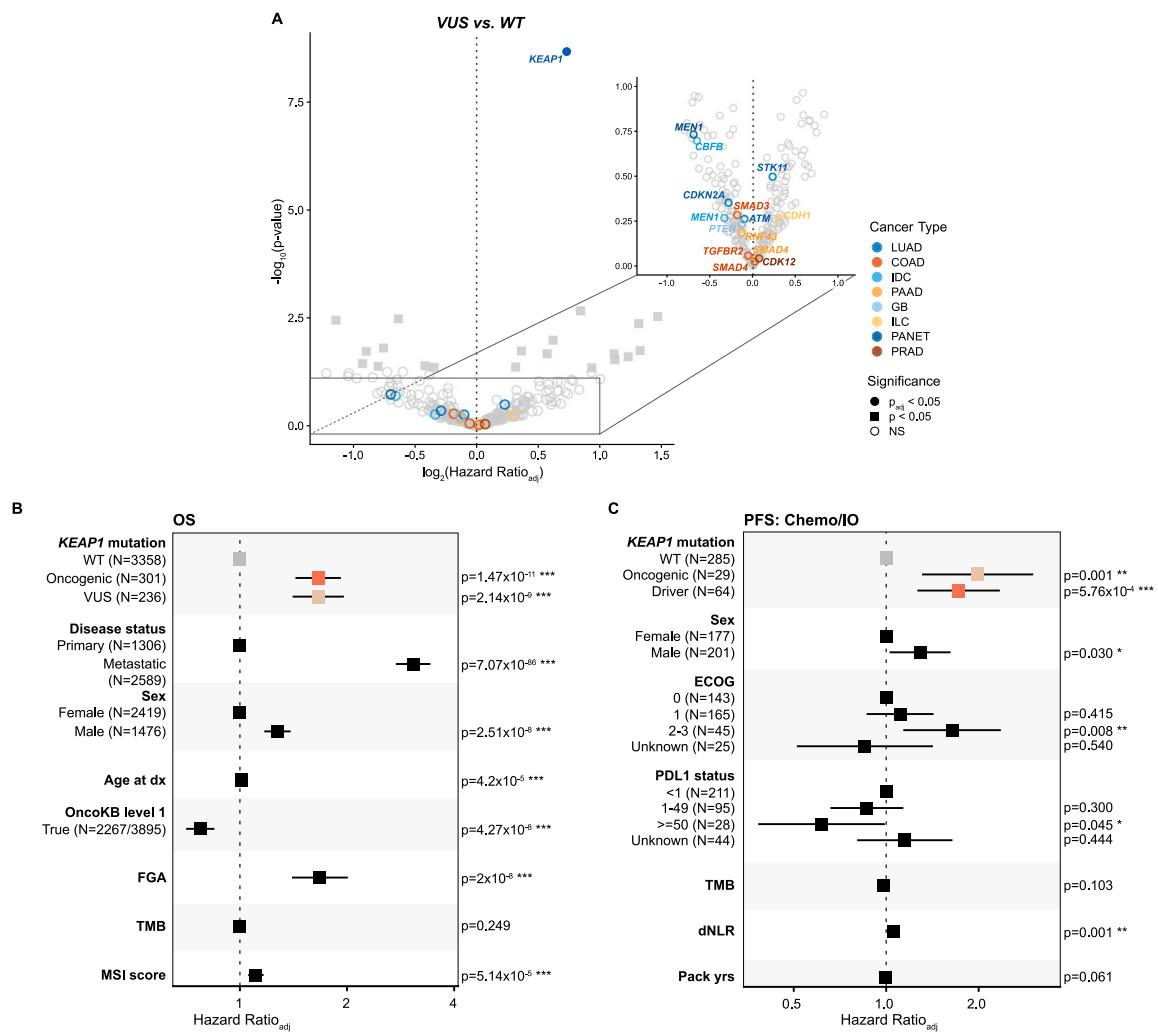


Figure S5. Overall and progression-free survival differences by TSG mutation class, related to Figure 5

(A) Volcano plot displaying comparisons of OS in patients with tumors harboring VUS vs. WT in all gene/cancer-type pairs with sufficient sample size (see [STAR Methods](#)). Each point represents the \log_2 of the adjusted hazard ratio vs. $-\log_{10} p$ value from a multivariate Cox proportional hazards models of OS by mutation class, adjusting for disease status, age, sex, FGA, MSI score, and TMB within each gene/subtype pair. The 15 pairs where significant selection for biallelic inactivation in VUS was observed are highlighted ([Figure 5A](#)). p values were adjusted for multiple testing using the FDR method with $n = 290$ comparisons. See [Table S4B](#) for a complete list. NS, not significant. See [Table S1A](#) for cancer type definitions.

(B) Forest plot of OS by KEAP1 mutation class, corresponding to [Figure 5F](#).

(C) Forest plot of PFS on chemoimmunotherapy by KEAP1 mutation class, corresponding to [Figure 5G](#). For (B) and (C), error bars represent 95% CI around adjusted HR estimate.

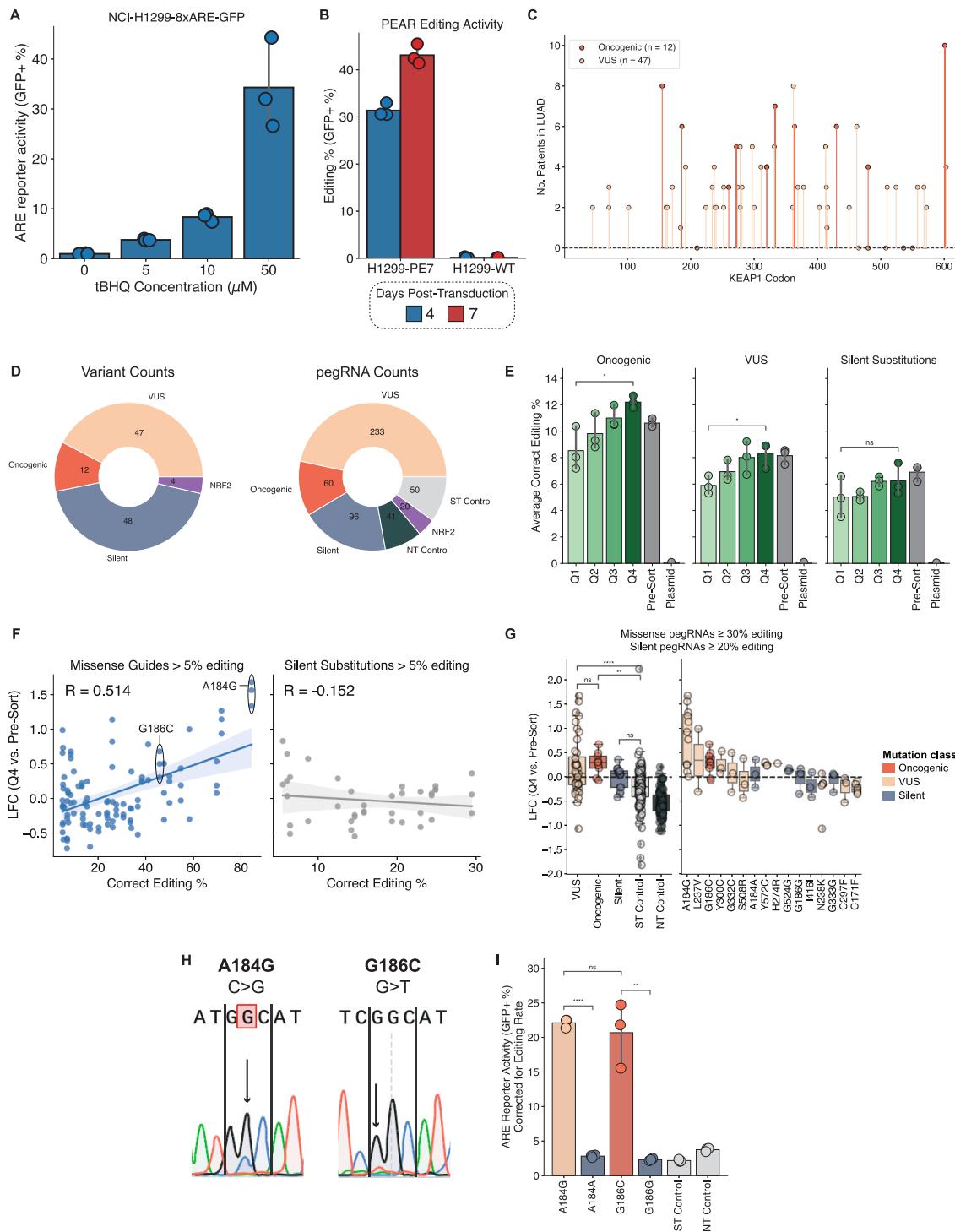


Figure S6. A prime editing sensor screen coupled to quantitative assessment of NRF2 activity validates KEAP1 VUSs, related to Figure 5

(A) Activation of the 8x ARE-GFP NRF2 activity reporter in NCI-H1299 cells treated with increasing doses of tBHQ. Plots show the GFP positive cell fraction of cells harboring the reporter construct. Data are presented as mean values with a 95% CI.

(B) Quantification of prime editing activity (GFP positive cell percentage) in H1299-PE7 and H1299-WT cells at 4 and 7 days post-transduction with the Lenti-PEAR-mCherry reporter. In this system, cells turn on GFP in the event of successful prime editing. Data are presented as mean values with a 95% CI.

(C) Plot of codon locations for *KEAP1* missense variants included in the library, colored by variant class. Variants R320L and R320Q were observed in four patients each and are overlapping in the plot.

(legend continued on next page)

-
- (D) Summary of the variants included in the prime editing sensor library (left) and the number of pegRNAs for each variant or control class in the library (right). NT control, non-targeting control; Silent, silent substitution control; ST control, safe-targeting control.
- (E) The average correct sensor editing percentage for oncogenic *KEAP1* variants, VUS, and silent substitution variants for each sorted population (Q1–Q4), as well as the pre-sorted populations and plasmid pool. Statistics shown for t test of independent samples with Bonferroni correction. Data are presented as mean values with a 95% CI.
- (F) Scatterplot of correct editing percentage (pre-sort) and LFC of Q4 relative to the pre-sort populations of missense-inducing pegRNAs (left) and silent substitution-inducing pegRNAs (right) for pegRNAs with >5% editing.
- (G) The \log_2 fold change (LFC) of the highest GFP-expressing bin (Q4) relative to pre-sort populations for different *KEAP1* variant or control classes. We used sensor editing measurements to focus this analysis on samples showing $\geq 30\%$ and $\geq 20\%$ sensor editing in the pre-sort population for missense and silent pegRNAs, respectively (left), and expanded to show individual variants (right). Statistics shown for t test of independent samples with Bonferroni correction. Boxplots display median line, IQR boxes, and 1.5x IQR whiskers.
- (H) Representative Sanger sequencing results of the endogenous *KEAP1* locus for the A184G and G186C pegRNAs.
- (I) Validation flow cytometry analyses for ARE-reporter activity (GFP+ %) of individual pegRNA-expressing NCI-H1299 cells (ST, safe-targeting; NT, non-targeting). Here, ARE-reporter activity is normalized to the expected number of homozygous variants by dividing by the square of the sensor editing rate for missense-inducing pegRNAs. Statistics shown for t test of independent samples with Bonferroni correction. * p value ≤ 0.05 , ** p value ≤ 0.01 , *** p value ≤ 0.0001 , ns, not significant (p value >0.05). Data are presented as mean values with a 95% CI.