



Gene Fusions Create Partner and Collateral Dependencies Essential to Cancer Cell Survival

Riaz Gillani^{1,2,3,4}, Bo Kyung A. Seong^{1,2}, Jett Crowdis^{2,5}, Jake R. Conway^{2,6}, Neekesh V. Dharia^{1,2,3,4}, Saif Alimohamed⁷, Brian J. Haas², Kyuho Han⁸, Jihye Park^{2,5}, Felix Dietlein^{2,5}, Meng Xiao He^{2,6}, Alma Imamovic^{2,5}, Clement Ma¹, Michael C. Bassik^{8,9,10}, Jesse S. Boehm², Francisca Vazquez², Alexander Gusev^{5,11}, David Liu^{2,5}, Katherine A. Janeway^{1,3,4}, James M. McFarland², Kimberly Stegmaier^{1,2,3,4}, and Eliezer M. Van Allen^{2,5,12}

ABSTRACT

Gene fusions frequently result from rearrangements in cancer genomes. In many instances, gene fusions play an important role in oncogenesis; in other instances, they are thought to be passenger events. Although regulatory element rearrangements and copy number alterations resulting from these structural variants are known to lead to transcriptional dysregulation across cancers, the extent to which these events result in functional dependencies with an impact on cancer cell survival is variable. Here we used CRISPR-Cas9 dependency screens to evaluate the fitness impact of 3,277 fusions across 645 cell lines from the Cancer Dependency Map. We found that 35% of cell lines harbored either a fusion partner

dependency or a collateral dependency on a gene within the same topologically associating domain as a fusion partner. Fusion-associated dependencies revealed numerous novel oncogenic drivers and clinically translatable alterations. Broadly, fusions can result in partner and collateral dependencies that have biological and clinical relevance across cancer types.

Significance: This study provides insights into how fusions contribute to fitness in different cancer contexts beyond partner-gene activation events, identifying partner and collateral dependencies that may have direct implications for clinical care.

Introduction

Structural variants (SV), including insertions, deletions, copy number alterations, translocations, and other complex rearrangements, play an integral role in oncogenesis (1–3). These somatic events are particularly enriched in pediatric cancers, which are classically characterized by otherwise low mutational burdens; however, their importance to cancer spans across the spectra of age and histology (4, 5). Prior work demonstrated that SVs contribute to transcriptional dysregulation through *cis*-regulatory element rearrangement (such as “enhancer-

hijacking”), copy number alterations leading to changes in gene dosage, and other variations on these phenomena (6–11). Changes in RNA expression within gene sets have been used to interpret SVs relevant for cancer pathogenesis. However, it is not known whether these expression changes at the level of individual genes create functional dependencies with an impact on cancer cell survival (12).

Genome-scale CRISPR-Cas9 knockout screening has enabled exploration of the functional importance of individual genes in numerous cancer contexts, and a collaborative effort has conducted CRISPR-Cas9 loss-of-function screening across established cancer cell lines to develop the Cancer Dependency Map (DepMap) project ($n = 769$ cell lines to date; refs. 13–15). Picco and colleagues previously demonstrated that CRISPR-Cas9 screening could be applied to understand the functional impact of partner genes in well-known and some less-well-characterized fusions across 371 cell lines (16). However, the full extent to which fusions create (i) functional dependencies on partner genes, or (ii) “collateral dependencies” on genes within the same topologically associating domains (TAD) as the fusion partners is unknown. We hypothesized that a significant number of fusions create dependencies on fusion partners as well as collateral genes through *cis*-regulatory element rearrangement, copy number change, or variations on these two processes (Fig. 1A). We therefore integrated fusion calls and dependency data in cell lines from DepMap to characterize the impact of fusions on cancer cell survival as mediated through partner and collateral dependencies, reasoning that a subset of these occurrences would reveal novel insights into cancer biology and therapeutic target development opportunities with potential for clinical translation.

¹Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. ²Broad Institute of Harvard and MIT, Cambridge, Massachusetts. ³Department of Pediatrics, Harvard Medical School, Boston, Massachusetts. ⁴Boston Children’s Hospital, Boston, Massachusetts. ⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. ⁶Harvard Medical School, Boston, Massachusetts. ⁷Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, North Carolina. ⁸Department of Genetics, Stanford University School of Medicine, Stanford, California. ⁹Program in Cancer Biology, Stanford University School of Medicine, Stanford, California. ¹⁰Program in Chemistry, Engineering and Medicine for Human Health (ChEM-H), Stanford University, Stanford, California. ¹¹Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. ¹²Center for Cancer Genomics, Dana-Farber Cancer Institute, Boston, Massachusetts.

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Eliezer M. Van Allen, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215. Phone: 617-632-6656; E-mail: eliezerm_vanallen@dfci.harvard.edu

Cancer Res 2021;81:3971–84

doi: 10.1158/0008-5472.CAN-21-0791

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 International (CC BY-NC-ND).

©2021 The Authors; Published by the American Association for Cancer Research

Materials and Methods

DepMap data source

We used a publicly available collection of annotated cell lines previously compiled and characterized by the Cancer Cell Line Encyclopedia (CCLE), as well as associated fusion calls, genome-scale

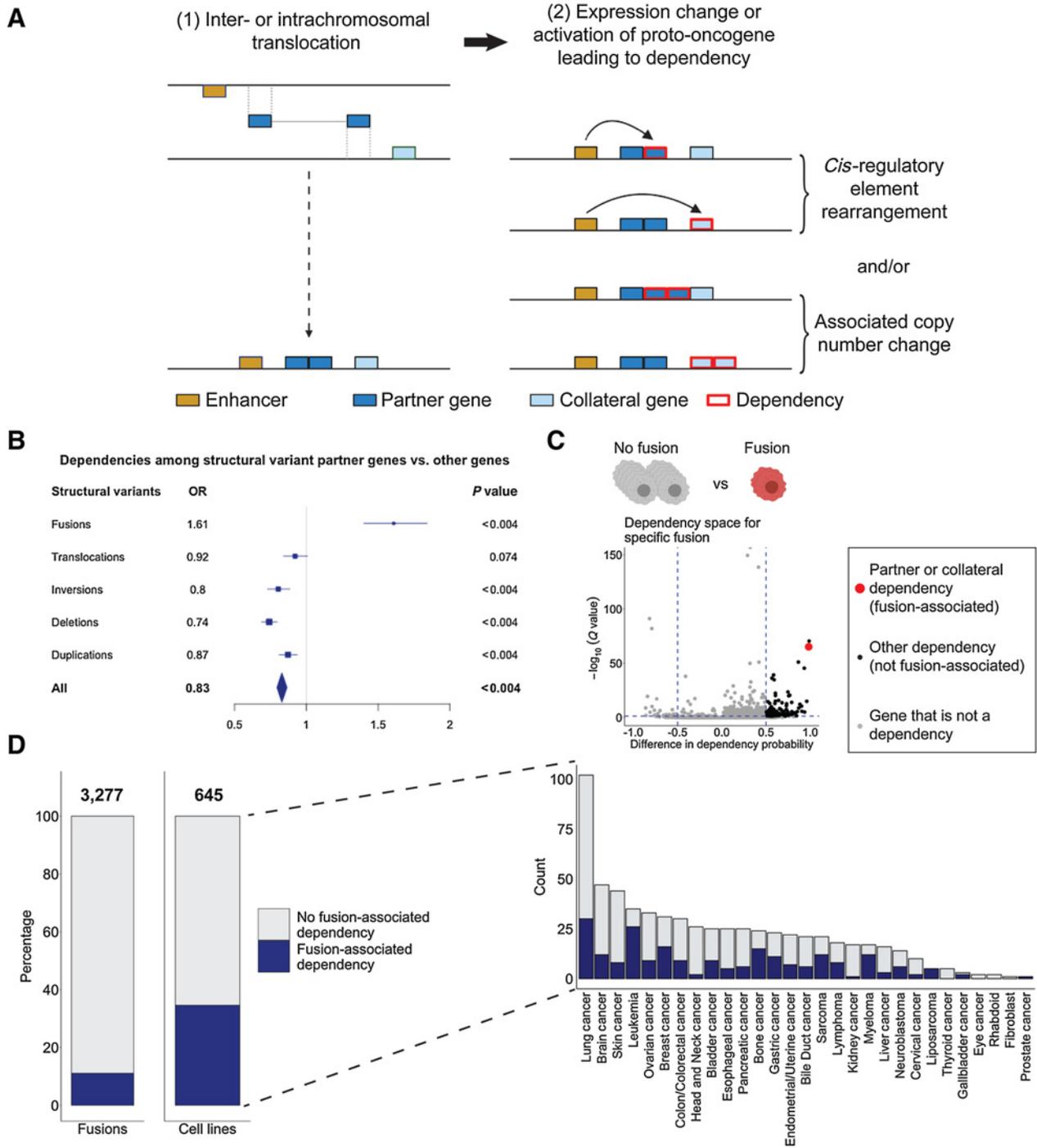


Figure 1. Hypothesis, approach, and identification of fusion-associated dependencies. **A**, Conceptual hypothesis illustrating how translocations lead to fusion-associated dependencies. **B**, Aggregate ORs and *P* values for enrichment of partners genes among absolute dependencies, stratified by category of structural variant (Fisher exact test; FWER < 0.05 to ascertain significance, using a Bonferroni correction results in a significance threshold of *P* < 0.004). **C**, Illustration of differential dependency evaluation for each fusion of interest. **D**, Left, proportion of all fusions and all cell lines identified as having at least one differential fusion-associated dependency (partner or collateral). Right, detailed count of cell lines with at least one differential fusion-associated dependency (partner or collateral) relative to total, broken out by disease type.

CRISPR knockout screening (dependency data), RNA expression, copy number alterations, mutation calls, and sgRNA locations from the following DepMap release (Supplementary Table S1): “DepMap, Broad (2020): DepMap 20Q2 Public. figshare. Dataset.” <https://doi.org/10.6084/m9.figshare.12280541.v4>. All subsequent genomic analyses related to these cell lines were based on alignment to hg38.

Fusion calls and CRISPR-Cas9 dependency data

RNA sequencing (RNA-seq)-based methods have been used extensively for fusion detection (17). The STAR-Fusion pipeline version 1.6.0 was previously used to identify fusions from RNA-seq data across cell lines in the CCLE (<https://github.com/STAR-Fusion/STAR-Fusion/wiki>). Among 1,299 cell lines, 21,999 unique fusions were identified after a preliminary filter was applied to remove the following as described in the DepMap documentation. We applied further filtering to select for high-confidence fusion calls (18) by focusing only on fusions with FPPM > 0.1, those with 5' GT and 3' AG dinucleotide breakpoints associated with canonical splicing (19, 20), those with >5 junction reads, and those that involved protein-coding partners or the IgH locus. This produced a list of 5,093 high-confidence fusions in 1,075 cell lines.

We used dependency data for 18,119 genes across 769 cell lines, processed as previously described in DepMap documentation to obtain CERES gene scores (correcting for gene independent CRISPR-Cas9 cutting in copy number amplified regions), and subsequently converted to dependency probability scores intended to infer the probability that a score represents a true dependency. As per DepMap documentation, because dependency probability scores take screening quality into account and may be utilized to identify which cell lines are more sensitive to CRISPR knockout when stratifying by the presence of a biomarker, they were used for all analyses. We took the intersection of cell lines with high-confidence fusion calls and those with dependency data to arrive at a starting set of 3,277 fusions across 645 cell lines.

Using fusions as a biomarker to identify associated dependencies across DepMap through genome-scale screening

For each of the 3,277 fusions, all cell lines were stratified by the presence or absence of the fusion of interest, and the mean dependency probability score for each of 18,119 protein-coding genes was calculated for each group. As 94% of fusions were observed in only a single cell line, a two-sample *t* test with the assumption of equal variance was carried out as a screen to identify genes that were likely to be dependencies based on the difference in scores between both groups. We drew upon the assumption that for each fusion, all cell lines were drawn from the same underlying population with a uniform variance. Correction for multiple-hypothesis testing was done using the Benjamini-Hochberg method to arrive at *Q* values (using $Q < 0.05$ to identify genes having a significant effect on cell survival). Building upon the consensus a prior threshold for a true dependency in the DepMap data set of an absolute dependency probability >0.5, we additionally only focused on differential dependencies for this analysis (those with a dependency probability score difference > 0.5 between cell lines with the fusion of interest and those without), to stringently highlight the genes having the most significant unique impact on cell lines with the fusion of interest. *Q* values were $-\log_{10}$ transformed for data visualization.

To identify associated dependencies for each fusion, we selected fusion partners and genes in close proximity to fusion partners. Prior studies have shown that TAD boundaries are largely invariant between

cell types (21). Although rearrangements that result in fusions will inevitably disrupt many TAD boundaries in cancer cells, we used TAD boundaries from a normal endothelial cell line as a generalizable approximation of genes in close proximity to fusion partners (22, 23). Genes in close proximity were identified on the basis of TADs from a Hi-C experiment done on an endometrial microvascular endothelial cell line; we downloaded the BED file with the identifier ENCF633ORE from ENCODE and utilized the UCSC LiftOver tool to convert hg19 coordinates for TAD boundaries to hg38 coordinates (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>; refs. 24, 25). Using gene coordinates from the BioMart package, SQL queries were carried out to assign each gene to a single TAD, with the exception of genes falling in TAD boundaries, which were assigned to upstream and/or downstream TADs within 40 kb of the starting gene coordinates (Supplementary Tables S2 and S3; ref. 26). Fusion-associated dependencies that were in close proximity to fusion partners were defined as collateral dependencies based on this assignment.

We defined a fusion-dependency pairing as a unique combination of a partner or collateral dependency with a given fusion. To obtain a conservative total count of fusion-dependency pairings, we counted only unidirectional transcripts for fusions that had both forward and reverse fusion transcripts detected by RNA-seq. Of the 104 fusions identified to contribute the 112 unique partner fusion-dependency pairings, STAR-Fusion was able to predict in-frame versus out-of-frame status for 75; manual curation to include *BCL2-IgH* and *RUNX1-IgH* as out-of-frame fusions led to a total of 77 fusions for which in-frame versus out-of-frame status was available. Of the 77 fusions with a predicted protein coding consequence, 50 (65%) were predicted to be in-frame whereas 27 (35%) were predicted to be out-of-frame (Supplementary Table S4). To obtain a count of fusion-dependency pairings accounting for complex rearrangements within TADs, we removed instances where a dependency was associated with multiple fusions in the same cell line (retaining partner dependencies in the rare cases that a gene was a partner and collateral dependency in the same cell line); this count was used as a comparator for permutation schemes.

Structural variant analysis in cell lines

Structural variant calls from WGS data were available for 329 cell lines using the SvABA structural variant caller as previously described (1). For the 209 cell lines with dependency data for which structural variant data were available, we evaluated whether fusions (those with associated dependencies and those without) were seen as part of a rearrangement, or if one or both partners were seen as part of separate rearrangements.

For the 103 fusion-cell contexts with associated dependencies and concurrent WGS structural variant data, we evaluated whether fusions appeared to be associated with simple or complex structural variants. For each fusion-cell context, we quantified the number of structural variants in the same TAD that did not involve at least one of the fusion partners. We used the ShatterSeek analysis package (27) to synthesize copy number and structural variant data in five different cell lines harboring fusion-associated dependencies (NCIN87, DU4475, THP1, HCC38, and PANC1) to visualize the relationship of fusions with simple and complex structural variants in these cell lines.

For the PANC1 cell line, we utilized ICE-normalized Hi-C data (Dekker and colleagues 2016) from the ENCODE portal (23) with the identifiers ENCF463HGQ and ENCF358MNA and the 3D Genome Browser (28) to visualize chromatin interactions.

Enrichment of SV partner and collateral genes among absolute dependencies

We analyzed structural variant calls involving at least one coding partner across 209 cell lines that also had dependency data. For each cell, we (i) identified genes that were absolute dependencies (those with a dependency probability score > 0.5 , consistent with the DepMap consensus a priori threshold for a true dependency), (ii) identified genes that were “partners” in a SV in that cell line, and (iii) identified genes that were “collateral” genes relative to a SV in that cell line, based on our previous definition of normal TADs.

Across all cell lines, we observed the total instances where a partner gene was an absolute dependency (vs. not a dependency) compared with other genes, calculated ORs, and determined significance using Fisher exact tests. We repeated this analysis independently for the collateral genes. We carried out this analysis for structural variants in aggregate, as well as for fusions, translocations (interchromosomal rearrangements detected by WGS), inversions, deletions, and duplications individually. Fusions in this analysis were those identified from RNA-seq using STAR-Fusion (as previously described) and with additional evidence in the WGS data of a SV correlate (manifesting as one or multiple of the other SV categories of deletion, duplication, inversion, or translocation). Fusion calls meeting these criteria were identified in 162 of the 209 cell lines with WGS and dependency data. Desiring a family-wise error rate (FWER) < 0.05 , we used a Bonferroni correction for the 12 hypotheses tested to obtain a threshold of $P < 0.004$ to ascertain significant associations. Analysis of ORs among individual cells was focused only on the 162 cell lines with fusion calls, and distributions were compared using two-sided t tests.

Copy number and mutational analysis in fusion-associated dependencies

Gene level copy number and mutation data were matched to each fusion-dependency-cell line context. Mutations identified included silent, missense, splice site, nonsense, and in-frame deletions. Copy number data were \log_2 transformed with a pseudo count of 1. We used < 0.5 as the threshold for copy number loss, and > 1.5 as the threshold for copy number gain.

Identifying COSMIC fusion genes, COSMIC cancer census genes, and kinases

Known Catalogue of Somatic Mutations in Cancer (COSMIC) fusions were utilized from the COSMIC fusion census (<https://cancer.sanger.ac.uk/cosmic/fusion>). A list of known cancer driver genes from the COSMIC cancer census (<https://cancer.sanger.ac.uk/cosmic/curation>) and a curated list of kinases from a prior study (29) were used to annotate remaining fusions with potentially interesting biology.

Evaluating FOXR1 overexpression and fusion-calling in clinical samples

A total of 12,747 clinical tumor samples with \log_2 -normalized TPM RNA expression data available through the UCSC Treehouse Childhood Cancer Project (Treehouse Tumor Compendium V11 Public PolyA; ref. 30) were screened for significant *FOXR1* overexpression, using a threshold of \log_2 (TPM + 1) > 2 . The histologies of specific samples with *FOXR1* overexpression were identified, and given the recurrence of neuroblastoma as a histologic subtype, four neuroblastoma samples from the TARGET study were selected for further evaluation of the presence of a *FOXR1* fusion (sample IDs: TARGET-30-PASUCB, TARGET-30-PASPBZ, TARGET-30-PASSWW, TARGET-30-PARBAJ). Two independent fusion-callers, STAR-Fusion and FusionCatcher (31), were used to call fusions from

hg19-aligned FASTQ RNA-seq files of these samples available for controlled access download through dbGaP (phs000218.v22.p8). WGS-based copy number variant calls from the Complete Genomics CNV pipeline (https://target-data.nci.nih.gov/Public/NBL/WGS/L3/copy_number/CGI/), specifically at the 11q23.3 locus where *FOXR1* is located, were additionally evaluated in these samples.

Experimental validation of FOXR1 fusion dependency in cell lines

143B (obtained from the Broad Institute; used within 6 months of collection; authenticated by L. Guenther using STR profiling; no mycoplasma testing conducted) and CALU6 (obtained from ATCC; used within 6 months of collection; authenticated by ATCC using STR profiling; mycoplasma testing conducted by ATCC, mycoplasma not detected) cells were grown in Eagle's minimum essential medium (EMEM) supplemented with 10% FBS and penicillin-streptomycin (Thermo Fisher Scientific, MT30002CI). To validate *FOXR1* fusion dependencies, we used CRISPR-Cas9 sgRNAs targeting *FOXR1* to knockout *PFAH1B2-FOXR1* in 143B cells and *RPS25-FOXR1* in CALU6 cells. The sgRNA sequences are as follows: sg*FOXR1-1* 5' GAGACCTCCAGCTTTCCAGG 3'; sg*FOXR1-2* 5' GGAAGATGC-CAGCTGCTCAG 3'; sg*FOXR1-3* 5' TGAGACCTCCAGCTTTC-CAG 3'. We infected cells with either nontargeting (NT) or *FOXR1*-targeting sgRNA, selected cells with puromycin (1 $\mu\text{g}/\text{mL}$), and assessed growth using CellTiter-Glo according to manufacturers' instructions. Immunoblot was performed to confirm knockout of the fusion-oncoproteins using anti-*FOXR1* (21942-1-AP; Thermo Fisher Scientific). Anti-vinculin (13901; Cell Signaling Technology) was used as a loading control. For crystal violet staining, cells were fixed and stained using crystal violet solution (1% crystal violet, 20% methanol) for 20 minutes, washed, and imaged.

Analysis of drug data

AUC values representing sensitivity to compounds and associated metadata were utilized from the Cancer Therapeutics Response Portal and available through DepMap (32–34). The AUC for venetoclax for the B-ALL cell line with the *BCL2-IgH* fusion was compared with other B-ALL cell lines without this fusion, as well as all other cell lines. We took an unbiased approach to screening for compound sensitivity for multiple myeloma cell lines, stratifying by the presence or absence of the *IgH-NSD2* fusion, and using a two-sample t test with the assumption of equal variance. We corrected for multiple-hypothesis testing using the Benjamini–Hochberg method to arrive at Q values. We annotated compounds for the inhibition of kinases to identify compounds of interest.

Analysis of histone ChIP-seq and DNASE-seq data

We downloaded BigWig files from ENCODE for the histone marks of interest for the KMS11 cell line with *IgH-NSD2* fusion, NCIH929 cell line with *IgH-NSD2* fusion, MM1S cell line without *IgH-NSD2* fusion, and peripheral blood mononuclear cell lines. We additionally utilized BigWig files from ENCODE for DNASE-seq analysis of the NCIH929 cell line with *IgH-NSD2* fusion, RPMI8226 cell line without *IgH-NSD2* fusion, and normal B cell lines (Supplementary Table S5). Epigenetic data were visualized using the integrative genomics viewer (IGV): version 2.8.2.

Calculation of phenotypic kill scores in CRISPR spheroid models

We normalized CRISPR guide dropout for genes of interest to nontargeting guides to calculate phenotypic kill scores as described previously (35). CRISPR sgRNA data from three spheroid models derived from NCIH23, NCIH1975, and NCIH2009 cell lines was

analyzed as follows: for each replicate, the count of sgRNAs targeting coding genes and nontargeting sgRNAs was normalized for two replicates at day 21 relative to day 0, and \log_2 -fold change values were calculated. The median and SD of \log_2 -fold change of nontargeting guides for each replicate was calculated, and the \log_2 -fold change values for targeting sgRNAs were normalized using these values to yield phenotypic Z (phenotypic kill) scores as described previously (35). The distribution of phenotypic kill scores for all guides targeting *EML4* in the NCIH23 spheroid model (containing the *THADA-MTA3* fusion) was compared with the distribution of phenotypic kill scores for all guides targeting *EML4* in the NCIH1975 and NCIH2009 spheroid models. This analysis was repeated and the mean of phenotypic kill scores was calculated for all previously defined nonessential genes, available through DepMap.

Results

Across all SVs, partner and collateral genes demonstrate the greatest enrichment among dependencies in the context of fusions

We first aimed to demonstrate that fusion partner genes and collateral genes, respectively and independently, were significantly enriched among functional cancer dependencies. Through analysis of whole-genome sequencing (WGS), RNA-seq-based fusion calls, and genome-scale dependency data for 209 cancer cell lines, we identified >26,000 SVs (translocations, inversions, deletions, duplications, and gene fusions) to assess for enrichment of partner genes (those directly involved in the SV of interest) and collateral genes (those in the same TAD as the SV of interest) among absolute dependencies (Supplementary Table S6, Materials and Methods). We found that there was a significant enrichment of partner genes among dependencies in the context of fusions, and either no enrichment or depletion of partner genes among dependencies in the context of all other SV groups (Fig. 1B; Supplementary Fig. S1A). There was significant enrichment of collateral genes among dependencies across all SV groups, but this enrichment was greatest for fusions as compared with all other SVs (Supplementary Figs. S1B and S1C). For fusions, we additionally carried out iterative enrichment analyses removing a single cell line at a time (Supplementary Materials and Methods; Supplementary Tables S7 and S8), as well as multivariate logistic regression analyses (Supplementary Tables S9 and S10), and confirmed that these observations were not driven by any single cell line or disease category (Supplementary Materials and Methods). These results were robust to variations in TAD size and definitions (Supplementary Fig. S2A–S2C; Supplementary Materials and Methods). Thus, the strongest enrichment of partner genes and collateral genes among dependencies occurred in the context of fusions.

Fusion-associated differential dependencies include partners and collateral genes, occurring more than would be expected by chance

We then developed a statistical framework to identify and nominate biologically relevant fusion-associated partner and collateral dependencies (Materials and Methods). Specifically, we performed dedicated analyses to (i) assess for differential dependencies in the context of fusions, (ii) validate the presence of fusions and associated dependencies through multiple approaches, and (iii) evaluate the relationship of fusions with simple or complex SVs. First, we assessed 3,277 fusions present in 645 cell lines with corresponding genome-scale dependency data (Materials and Methods, Supplementary Fig. S3A, range of mean fusions per cell line: 1–20, Supplementary Fig. S3B;

Supplementary Table S11). For each fusion, we carried out a statistical genome-scale screen to identify genes that had differential dependencies leading to increased fitness in the cell line(s) containing the fusion (Materials and Methods). On the basis of our preceding enrichment analysis, we hypothesized that a subset of fusions would lead to expression changes or activation of proto-oncogenes, creating differential gene dependencies resulting from the rearrangements themselves. Thus, within each fusion-dependency relationship, genes that were identified as being differential dependencies selectively in cell lines with the fusion (range 0–260 genes, mean 37 genes) were evaluated to identify fusion partner dependencies and collateral dependencies, independently (referred to as fusion-associated dependencies hereafter; Fig. 1C, Materials and Methods).

Across all fusions, 363 (11%) had at least one fusion-associated dependency. Fusion-associated dependencies were observed in 223 cell lines (35%) and occurred in greater than half of leukemia, breast cancer, multiple myeloma, bone cancer, liposarcoma, and other sarcoma cell lines (Fig. 1D). We identified 659 unique fusion-dependency pairings in total (112 partner, 547 collateral; Supplementary Table S12); accounting for complex rearrangements (by removing instances of dependencies associated with multiple fusions in the same cell line), we observed 483 fusion-dependency pairings (100 partner, 383 collateral; Fig. 2A, Materials and Methods). Of 223 cell lines with fusion-associated dependencies, 207 (93%) had at least one hotspot driver mutation (range 1–25 driver mutations, mean 2.5 driver mutations; Supplementary Materials and Methods; Supplementary Table S13). Cell lines without fusion-associated dependencies had a significantly increased number of hotspot driver mutations (range 1–62 driver mutations, mean 3.4 driver mutations; Supplementary Fig. S4A, $P = 0.003$, two-sided t test), although the proportion of cell lines with hotspot driver mutations was comparable (383 of 422 cell lines, 91%, $P = 0.456$, Chi-squared test). Fusion-associated dependencies contributed to cancer cell fitness uniquely, even in the presence of other hotspot driver mutations.

Next, to demonstrate that this phenomenon of fusion-associated dependencies was occurring more than would be expected by chance, we carried out fusion-label permutation testing (breaking the link between fusions and dependency scores to create a null distribution, Fig. 2B) and gene-label permutation testing (breaking the link between genes and dependency scores to create a null distribution, Supplementary Figs. S4B and S4C) for partner and collateral dependencies, respectively and independently. Our observed counts of partner and collateral fusion-dependency pairings were significantly greater than those expected by chance ($P < 0.001$, Supplementary Materials and Methods).

Because of the non-Gaussian distribution of dependency scores and the small numbers of cell lines with any given fusion, we performed additional cell line permutation-based FDR estimation as an approach to fusion-associated dependency discovery, and found that 459 fusion-dependency pairings (70%) identified by our genome-scale screen met the FDR threshold < 0.05 by this approach as well (Fig. 2C, Supplementary Materials and Methods, Supplementary Fig. S5, Supplementary Table S14). The fusion-dependency pairings identified by both approaches were most likely to have biological relevance, and we therefore prioritized these pairings for further study.

Furthermore, for the subset of cell lines with corresponding WGS (Fig. 2D), there was evidence for the presence of a correlated structural variant in 86 of 103 fusion-cell contexts (83%, Materials and Methods), comparable with an evaluation of WGS correlates to fusions detected by RNA-seq in clinical samples (10). This proportion was not

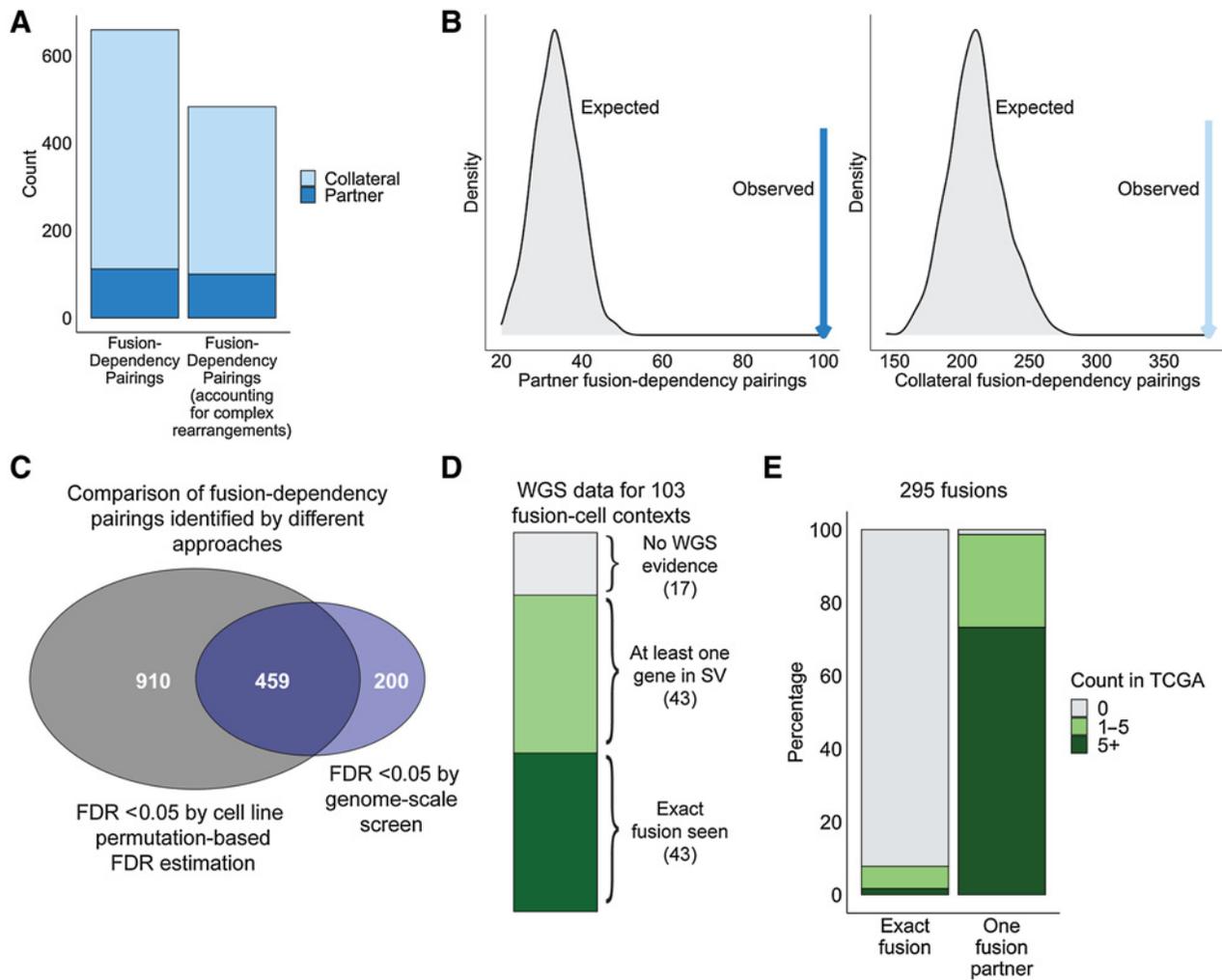


Figure 2.

Fusion-dependency pairings, permutation testing, and supporting data. **A**, Summary of unique fusion-dependency pairings across DepMap. Left bar, all fusion-dependency pairings excluding reciprocal transcripts. Right bar, fusion-dependency pairings with instances of dependencies associated with multiple fusions in the same cell line removed (conservative estimate). **B**, Observed conservative count of total fusion-dependency pairings compared with the null distribution of expected fusion-dependency pairings obtained by 1,000 fusion-label permutations for each of 3,277 fusion-dependency relationships. Left, partner fusion-dependency pairings ($P < 0.001$). Right, collateral fusion-dependency pairings ($P < 0.001$). **C**, Comparison of fusion-dependency pairings identified by cell line permutation-based FDR estimation (gray) and genome-scale screen (blue). **D**, For 103 fusion-cell contexts that had associated dependencies and structural variant calls from WGS data, the proportion for which there was orthogonal supporting evidence. Exact fusion seen = a single structural variant was identified to involve the two fusion partners in WGS data. At least one gene in SV = either both partner genes, the left partner gene, or the right partner gene were independently involved in structural variants. **E**, Of the 295 fusions that had associated dependencies and corresponding diseases in the TCGA, the proportion with representation in this clinical data set. Left bar, the proportion of fusions for which an exact match was seen. Right bar, the proportion of fusions for which at least one partner was seen (counted by the most recurrent partner).

significantly different between fusions with associated dependencies and those without (83% vs. 76%, $P = 0.11$, Fisher exact test; Supplementary Fig. S6A). We also evaluated whether the fusions with associated dependencies from cell lines were identified in a prior study of clinical samples from The Cancer Genome Atlas (TCGA; ref. 29). Of 363 fusions with associated dependencies, 295 were in cell lines with the same histology as tumors in the TCGA (Fig. 2E; Supplementary Fig. S6B; Supplementary Table S15). Of those, 23 (8%) had an exact fusion match in the TCGA, though 291 fusions (99%) had a partner that was seen as part of a fusion in the TCGA, supporting the relevance of preclinical dependency analysis to patient cohorts (Supplementary Materials and Methods). These analyses

provided additional orthogonal validation of the fusions identified for study.

Finally, we also sought to determine when gene fusions with associated dependencies were part of simple or larger complex SVs. Gene fusions varied with regard to their association with other SVs in close proximity: 37% of fusions had no additional SVs in the same TAD beyond those involved in the fusion, 41% had 1 to 5 other same-TAD SVs, 13% had 6–10 other same-TAD SVs, and 9% had >10 other same-TAD SVs (Materials and Methods, Supplementary Fig. S6C). In some cases of collateral dependencies, fusions could be more directly linked to the dependency in question (Supplementary Figs. S7A–S7C), whereas in other cases, they were

proxies for larger complex SVs collectively contributing to cis-regulatory element rearrangement and copy number change (Supplementary Figs. S7D and S8A–S8B). Thus, for fusions with partner and collateral dependencies in cancer cell lines, there were multiple lines of evidence for their enrichment and clinical relevance to support further investigation.

Copy number, mutational, and transcriptional landscapes intersect with fusion-associated dependencies

We next evaluated copy number alterations and somatic mutations for their potential to contribute to the development of fusion-associated dependencies (Materials and Methods). Partner dependency genes were amplified in 65 of 212 fusion-dependency-cell contexts (31%), whereas collateral dependency genes were amplified in 405 of 588 fusion-dependency-cell contexts (69%, Fig. 3A). This high rate of copy number amplification of fusion-associated dependency genes aligned with prior work showing rearrangements and copy number alterations are highly inter-related (36). However, mutations involving fusion-associated dependency genes were relatively infrequent, with partner dependency genes harboring mutations in 16 of 217 fusion-dependency-cell contexts (7%), and collateral dependency genes harboring mutations in 27 of 589 fusion-dependency-cell contexts (5%, Fig. 3B).

Prior studies demonstrated that fusions can contribute to transcriptional dysregulation in patient samples (6, 7, 9). Thus, we next evaluated RNA expression data for all fusions with orthogonal dependency data to determine the degree of overlap between genome-scale pan-cancer unbiased overexpression and dependencies for fusion-associated genes (Fig. 3C; Supplementary Table S16; Supplementary Materials and Methods). Of 631 fusion-associated overexpressed partner genes, 40 were also dependencies (6%). Similarly, of 1,400 fusion-associated overexpressed collateral genes, 70 were dependencies (5%). Although many fusions led to overexpression of associated genes, only a small proportion of these cases were deemed essential to cell survival through this screening modality.

We noted that 40 of 112 (36%) partner dependencies and 70 of 547 (13%) collateral dependencies were differentially overexpressed in an unbiased manner. However, we observed at least a \log_2 -fold change TPM > 1 in 59% of all fusion-associated dependencies without correcting for genome-scale significance (Supplementary Table S12). This was significantly greater than the 5% of fusion-associated dependencies with \log_2 -fold change TPM < -1 (Supplementary Fig. S9A and S9B, $P < 0.001$, Fisher exact test). Certain well-characterized fusion-associated dependencies did not meet these criteria for overexpression. For instance, *BCR* and *ABL1* were strong dependencies associated with the *BCR-ABL1* fusion, but did not meet the threshold of \log_2 -fold change TPM > 1. Similarly, for the *KMT2A* fusions established as childhood myeloid and lymphoid leukemia drivers (37, 38), *KMT2A* was a dependency but did not demonstrate significant overexpression in cell lines harboring these fusions compared with all cell lines without these fusions (Supplementary Fig. S10A–S10D). Thus, essential fusions can induce modest expression changes in associated genes that are context-specific without manifesting as unbiased overexpression, but these events are still important to cell survival and support expression dysregulation as the mechanism leading to fusion-associated dependencies.

COSMIC fusions demonstrate utility of CRISPR-Cas9 for identifying essential genes

Many kinases and COSMIC Cancer Census genes appear among other fusions with associated dependencies. To further explore the

biological relevance of fusion-associated dependencies, we examined whether recurrent biologically established fusions defined by COSMIC (“COSMIC fusions”) could be recovered using genome-scale CRISPR-Cas9 loss-of-function screening (39). Across all high-confidence fusion calls in cell lines with genome-scale dependency data, we identified 35 unique COSMIC fusions: 19 fusions had a partner dependency, 1 was associated with a collateral dependency, and 15 had no associated dependency (Fig. 3D). For fusions such as *BCR-ABL1* and *PAX3-FOXO1*, both partner genes were differential dependencies. For *EWSR1-FLI1*, only *FLI1* was a differential dependency, as *EWSR1* is a common essential gene in many different cell contexts (Fig. 3E; Supplementary Figs. S11A–S11F). For some fusions resulting from unbalanced rearrangements like *EWSR1-ERG*, single-guide-RNA (sgRNA) location could preclude a partner screening as a dependency (Fig. 3F; Supplementary Figs. S12A and S12B; Supplementary Table S17; Supplementary Materials and Methods). Using COSMIC fusions as a positive control for this screening modality, we demonstrated that CRISPR-Cas9 loss-of-function screening could identify fusion-associated dependencies in 20 of 35 (57%) cases where we would have expected them to exist.

We also assessed partner and collateral dependencies that involved either kinases or COSMIC cancer census genes, with the hypothesis that many of these would have biological relevance (Supplementary Fig. S13A). The *BCL2-IgH* fusion has been previously reported in various hematologic malignancies (40, 41), and it was observed concurrently with a *BCL2* missense mutation in a B-ALL cell line (JM1), contributing to a partner dependency on *BCL2*. Cancer Therapeutics Response Portal compound screening data for JM1 showed it was highly sensitive to the *BCL2* inhibitor venetoclax when compared with all other cancer cell lines, with median sensitivity in comparison with other B-ALL cell lines (Supplementary Fig. S13B; refs. 32–34). Genes that were copy number amplified, such as *MDM2* and *ERBB2*, were involved in fusions as well, and sometimes associated with multiple collateral dependencies (Supplementary Fig. S13C; Supplementary Fig. S14A–S14D). A study from the PCAWG consortium showed that the OE33 esophageal cancer cell line was characterized by an inversion around *ERBB2*, disrupting a TAD boundary and leading to the fusion of two TADs (21). We found that the *ERBB2-JUP* fusion in the OE33 cell line was associated with four collateral dependencies in addition to the partner dependency on *JUP*. The known formation of a neo-TAD provides a mechanistic explanation for expression changes and the consequent presence of multiple dependencies in close proximity to *ERBB2* in this cell line. Therefore, through the biological priors of COSMIC fusions and other COSMIC genes involved in known structural variants, we demonstrated that genome-scale CRISPR-Cas9 loss-of-function screening was effective in identifying true fusion-associated dependencies.

Transcription factors are recurrent fusion-associated dependencies

Having established the biological and statistical bases for partner and collateral dependencies, we next assessed statistically significant fusion-dependency pairings (those identified by multiple approaches) for functional relevance (Supplementary Table S12; Materials and Methods). Among recurrent fusion-associated dependencies, several Forkhead-box transcription factors (42, 43) were essential to cancer cell survival. We observed three instances of intrachromosomal *FOXR1* fusions associated with *FOXR1* as a differential dependency, occurring independently in osteosarcoma, lung adenocarcinoma, and bladder carcinoma cell lines (Fig. 4A). All three fusions preserved the

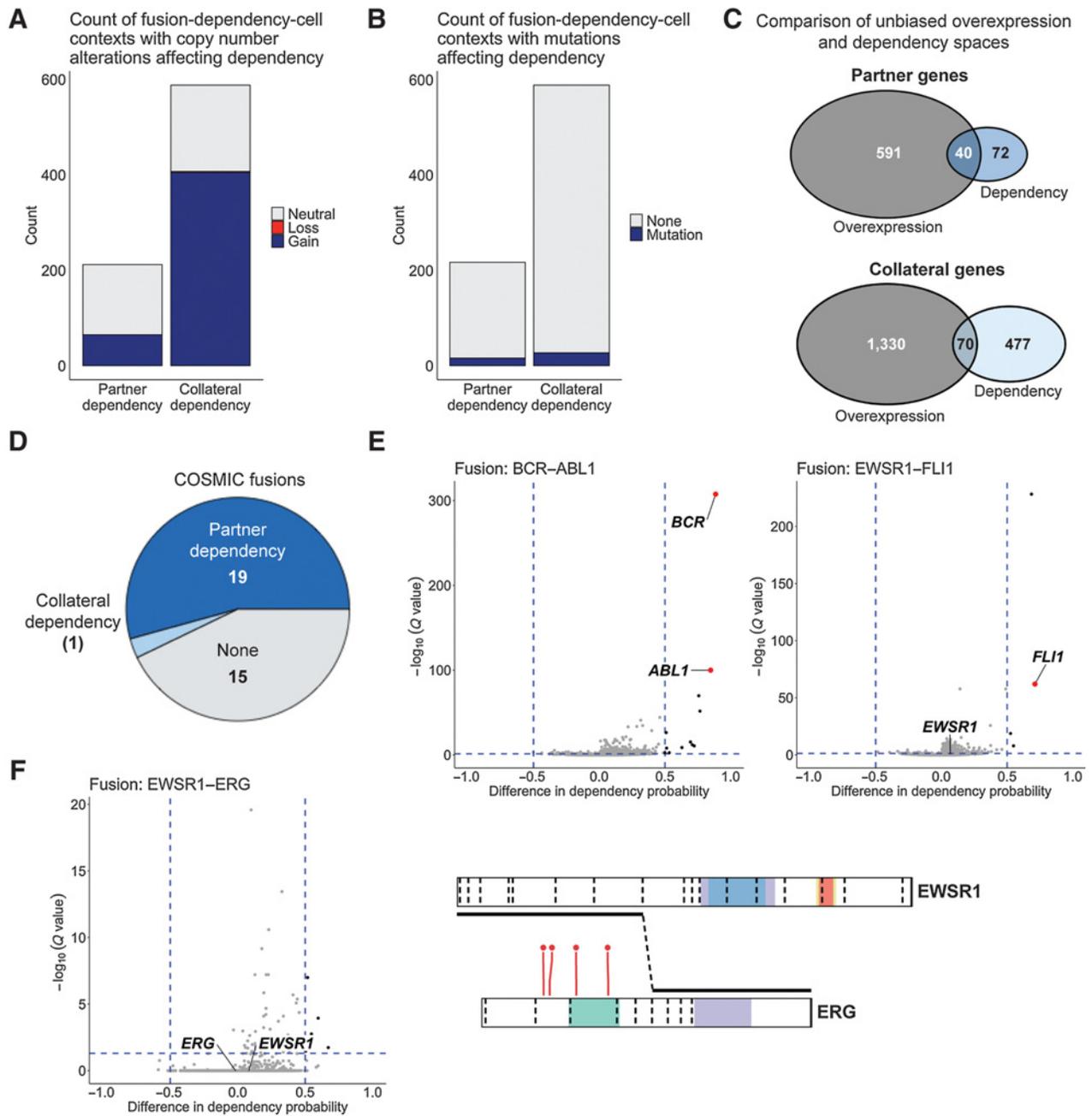
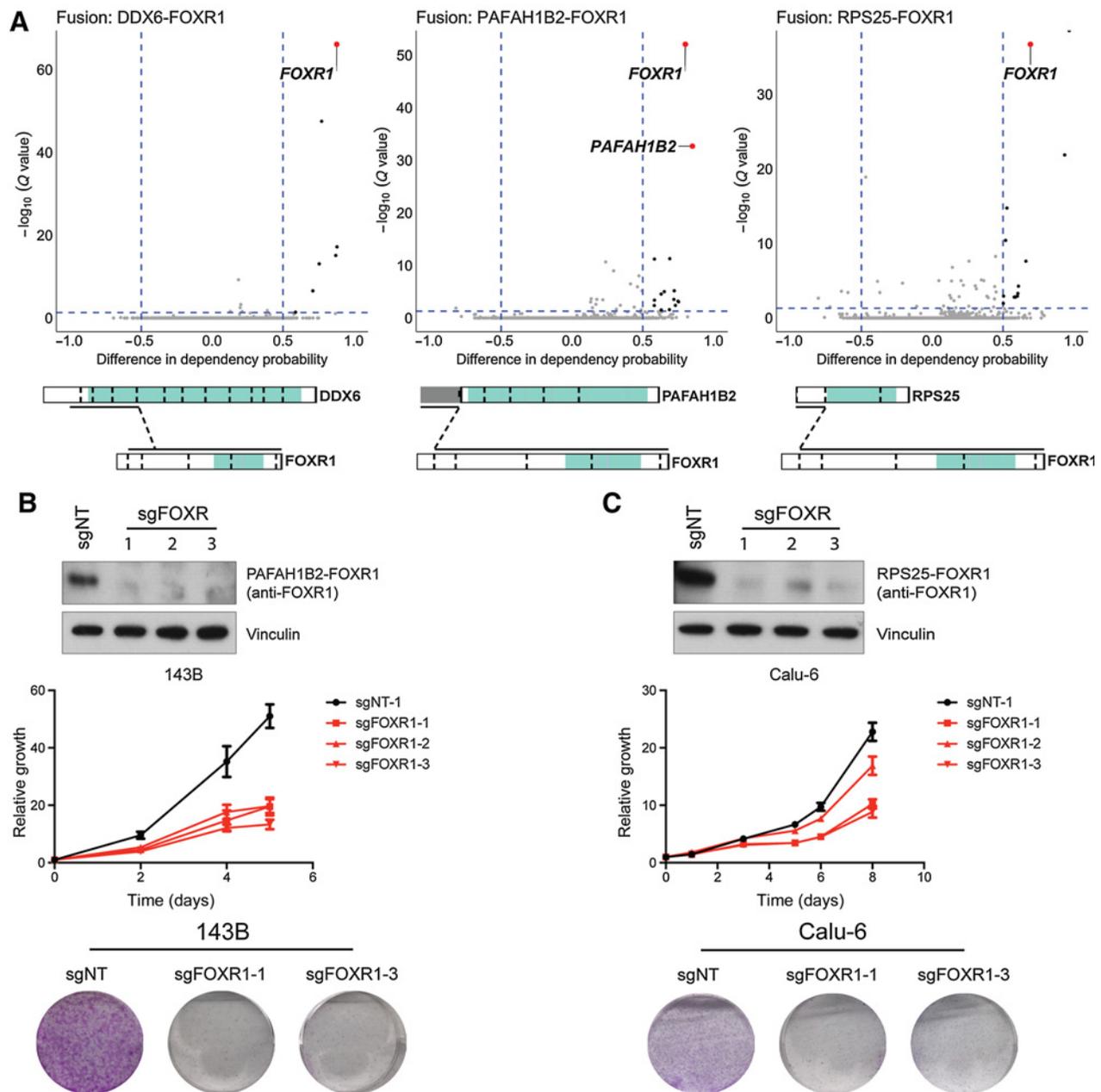


Figure 3.

Copy number and mutation data; comparison of unbiased RNA overexpression and dependencies associated with fusions; COSMIC fusions demonstrate utility and limitations of CRISPR-Cas9 for identifying essential genes. **A**, Count of instances in which a fusion-associated dependency in a cell line is associated with copy number amplification. **B**, Count of instances in which a fusion-associated dependency in a cell line is associated with a concurrent somatic mutation. **C**, Comparison of the count of fusion-overexpressed gene pairings and fusion-dependency pairings. Top, partner genes. Bottom, collateral genes. **D**, Proportion of all COSMIC fusions with corresponding dependency data ($n = 35$) with at least one partner dependency, at least one collateral dependency, or no associated dependency. **E**, Left, dependency space for cell lines with *BCR-ABL1* fusion; both *BCR* and *ABL1* are identified as dependencies. Right, dependency space for cell lines with *EWSR1-FLI1* fusion; *FLI1* is a strong dependency, but *EWSR1* is not a strong selective dependency as it is a common essential gene. **F**, Dependency space for cell lines with *EWSR1-ERG* fusion. *ERG* does not screen as a dependency because of sgRNA location (red lines) off the *EWSR1-ERG* fusion transcript, with breakpoint illustrated.

active Forkhead domain of *FOXRI*. There have been previous reports of intrachromosomal fusions involving *FOXRI*, a member of the Forkhead-box family, in rare cases of neuroblastoma. The *FOXRI* fusions in cell lines were associated with the overexpression of *FOXRI*,

which is normally only seen in embryogenesis (Supplementary Fig. S15A–S15C; refs. 44, 45). Therefore, there was strong statistical evidence for *FOXRI* fusions creating fusion-associated dependencies with implications for oncogenesis.

**Figure 4.**

Fusions involving *FOXR1* result in recurrent associated partner dependencies. **A**, *FOXR1* fusions are associated with dependency on *FOXR1* in three different cell lines. This is supported by fusion transcripts that preserve the functional domain of *FOXR1* (colored regions). *DDX6-FOXR1* is seen in bladder cancer cell line 639V, *PAFAH1B2-FOXR1* is seen in osteosarcoma cell line 143B, and *RPS25-FOXR1* is seen in lung cancer cell line CALU6. **B** and **C**, CRISPR-Cas9 knockout of *FOXR1* validates *FOXR1* fusion dependency in osteosarcoma cell line 143B (**B**) and lung cancer cell line CALU6 (**C**). Top, immunoblot showing *FOXR1* fusion protein expression after CRISPR-mediated knockout. Middle and bottom, cells infected with either nontargeting or *FOXR1* targeting sgRNA were assessed for cell growth using CellTiter-Glo (middle) and crystal violet staining (bottom).

To demonstrate the clinical relevance of *FOXR1* fusions, we used *FOXR1* overexpression in a cohort of >12,000 clinical samples as a preliminary screen for identifying clinical samples that may harbor this fusion (Materials and Methods; Supplementary Fig. S16A). Among clinical samples with the highest degree of *FOXR1* overexpression were four neuroblastoma samples from the TARGET study (Supplementary Fig. S16B; ref. 45). We identified *FOXR1* fusions in all four

neuroblastoma samples, and demonstrated that in the three cases of intrachromosomal fusions, there was associated copy number alteration at the 11q23.3 locus where *FOXR1* resides (Supplementary Table S18).

Having established the clinical relevance of *FOXR1* fusions, we next validated the observed *FOXR1* dependencies in two cell line models. In the osteosarcoma cell line 143B harboring a *PAFAH1B2-FOXR1* fusion

and associated *FOXRI* dependency, we found that CRISPR-mediated knockout of the fusion led to a significant reduction in cell growth (Fig. 4B). Similarly, in the lung cancer cell line CALU6 harboring a *RPS25-FOXRI* fusion and associated *FOXRI* dependency, we demonstrated that CRISPR-mediated knockout of the fusion resulted in decreased cell growth (Fig. 4C). Thus, *FOXRI* fusions created dependency on the fusion partner *FOXRI*, integral to cancer cell survival when present.

FOXA1 is another member of the Forkhead-box family, and contributes to oncogenesis in different cancers, playing a central role in prostate cancer (42). Prior work demonstrated that rearrangement in the *FOXA1* TAD leads to the hijacking of a nearby enhancer known as FOXMIND and contributes to *FOXA1* overexpression in a prostate cancer cell line, VCAP (46). We observed the presence of a *TTC6-MIPOL1* fusion in two prostate cancer (including VCAP) cell lines, as well as one colorectal cancer, breast cancer, and lung cancer cell line. *TTC6* and *MIPOL1* flank the *FOXA1* locus; for the prostate cancer and colorectal cancer cell line with the *TTC6-MIPOL1* fusion and dependency data available, *FOXA1* was a strong collateral dependency (Supplementary Fig. S17A). In other cell lines with the fusion but without dependency data available, *FOXA1* was highly overexpressed. *TTC6-MIPOL1* has been previously reported as a recurrent adjacent gene rearrangement in breast cancers (47). Our results suggest that this previously described rearrangement contributes to oncogenesis through *FOXA1* overexpression not only in prostate cancer, but other cancers as well.

Finally, among other transcription factors, we observed *HNF1A* as a collateral dependency associated with two distinct fusions in gastric cancer cell lines, and it was associated with a mean log₂-fold change of 4 in expression (Supplementary Fig. S17B). The context specificity suggests that rearrangements in close proximity to *HNF1A* may contribute to its overexpression and resulting essentiality in some gastric carcinoma cell lines. In summary, we established that fusions contribute to oncogenesis in several instances by creating partner and collateral dependencies on transcription factors.

Clinical applicability of fusion-associated dependencies

We finally examined whether highly recurrent clinically observed fusions created potential clinically actionable collateral dependencies. Approximately 15% of patients with multiple myeloma have a translocation (4;14), which is associated with poor prognosis (48). In this translocation, the IgH enhancer is juxtaposed with *NSD2* and *FGFR3*, leading to aberrant expression of both genes located in close proximity to each other. Because *FGFR3* overexpression is not universal in t(4;14) cases, there have been different reported conclusions about the gene that is most relevant to oncogenesis in the presence of this rearrangement (49–52). Here, five t(4;14) multiple myeloma cell lines with dependency data were identified as having an *IgH-NSD2* fusion. Compared with multiple myeloma cell lines without this fusion, both *FGFR3* and *NSD2* were overexpressed (Fig. 5A; Supplementary Figs. S18A–S18C). However, only *FGFR3* was a strong dependency in these cell lines (Fig. 5B; Supplementary Table S19). Concurrent *FGFR3* mutations were seen in three of the five cell lines (missense in KMS18 and OPM2, silent in KMS11). *FGFR3* remained a dependency in two of the cell lines without concurrent *FGFR3* missense mutations (KMS26 and KMS34), supporting *IgH-NSD2* as the primary molecular lesion driving this dependency.

A multiple myeloma cell line with the fusion (KMS11) was characterized by increased H3K27ac and H3K9ac, and relatively decreased H3K27me3, at *FGFR3* reflective of an active transcriptional state in the presence of the fusion (Fig. 5C; Supplementary Figs. S19A and S19B; refs. 22, 23). In addition, the top two statistically significant therapies in

this context were cediranib and lenvatinib, which are multikinase inhibitors that also have established anti-*FGFR* activity (Fig. 5D; refs. 53–55). Although not statistically significant for an unbiased screen, *FGFR3* inhibitors AZD4547 and nintedanib also demonstrated increased activity against multiple cell lines with the *IgH-NSD2* fusion (Fig. 5D). Integrating collateral dependencies with matching epigenetic and therapeutic data, we found *FGFR3* to be the targetable dependency in t(4;14) multiple myeloma cell lines.

Finally, given the patient-specific nature of many fusion-associated dependencies, we evaluated whether such events could be translated to spheroid models, which have demonstrated utility for patient-derived prospective precision cancer medicine studies (56). Han and colleagues performed genome-scale CRISPR screening for dependencies in multiple spheroid models, one of which was derived from the NCIH23 lung cancer cell line with a *THADA-MTA3* fusion (35). In addition to being a partner dependency in DepMap cancer cell lines (C10, NCIH3122) with the *EML4-ALK* COSMIC fusion (57), we observed *EML4* to be a collateral dependency in the NCIH23 cell line with the *THADA-MTA3* fusion. There was strong evidence for the presence of the *THADA-MTA3* fusion in the NCIH23 cell line from RNA-seq and WGS data (Fig. 6A).

In evaluating phenotypic kill scores for sgRNAs targeting *EML4*, there was increased dependency on *EML4* in the spheroid model derived from the NCIH23 cell line in comparison with the spheroid models derived from cell lines without the *THADA-MTA3* fusion ($P = 0.013$, two-sided t test; Fig. 6B; Supplementary Fig. S20A). To ensure that this was not the case for all genes, we evaluated mean phenotypic kill scores for sgRNAs targeting nonessential genes as defined by Hart and colleagues, and observed a similar distribution in spheroid models with and without the *THADA-MTA3* fusion (Fig. 6B; Supplementary Fig. S20B; ref. 58). *EML4* as a fusion-associated dependency appeared to be relevant in three-dimensional cancer models, establishing relevance for discovering potentially actionable fusion-associated dependencies from clinical cancer samples.

Discussion

In this study, we demonstrated that many fusions contribute to cancer cell survival by creating partner and collateral dependencies. We also showed that while fusions frequently lead to transcriptional dysregulation, which is the likely intermediate mechanism for creating fusion-associated dependencies when they exist, there is only modest overlap between the unbiased overexpression and dependency spaces. Not all transcriptional dysregulation resulting from structural variation contributes directly to cancer cell survival, and CRISPR-Cas9 dependency provides significantly more insight into essential gene expression changes that often do not manifest as pan-cancer genome-scale overexpression, but still confer a fitness advantage.

We leveraged WGS data to demonstrate that fusions could arise from simple structural variants directly contributing to collateral dependencies, or alternatively be proxies for more complex rearrangements contributing to the development of collateral dependencies. Although in the latter case, the precise role of the fusion in the development of the collateral dependency was more difficult to define, we reasoned that fusions still contributed meaningfully to cancer cell survival in many of these instances because the enrichment of collateral genes among dependencies was greatest for fusions as compared with other SVs.

We showed that specific fusion-associated dependencies had biological and clinical relevance. The *FOXRI* fusions were associated with dependency on *FOXRI* in different cancer cell contexts; we validated these dependencies in cell lines and demonstrated that *FOXRI* fusions

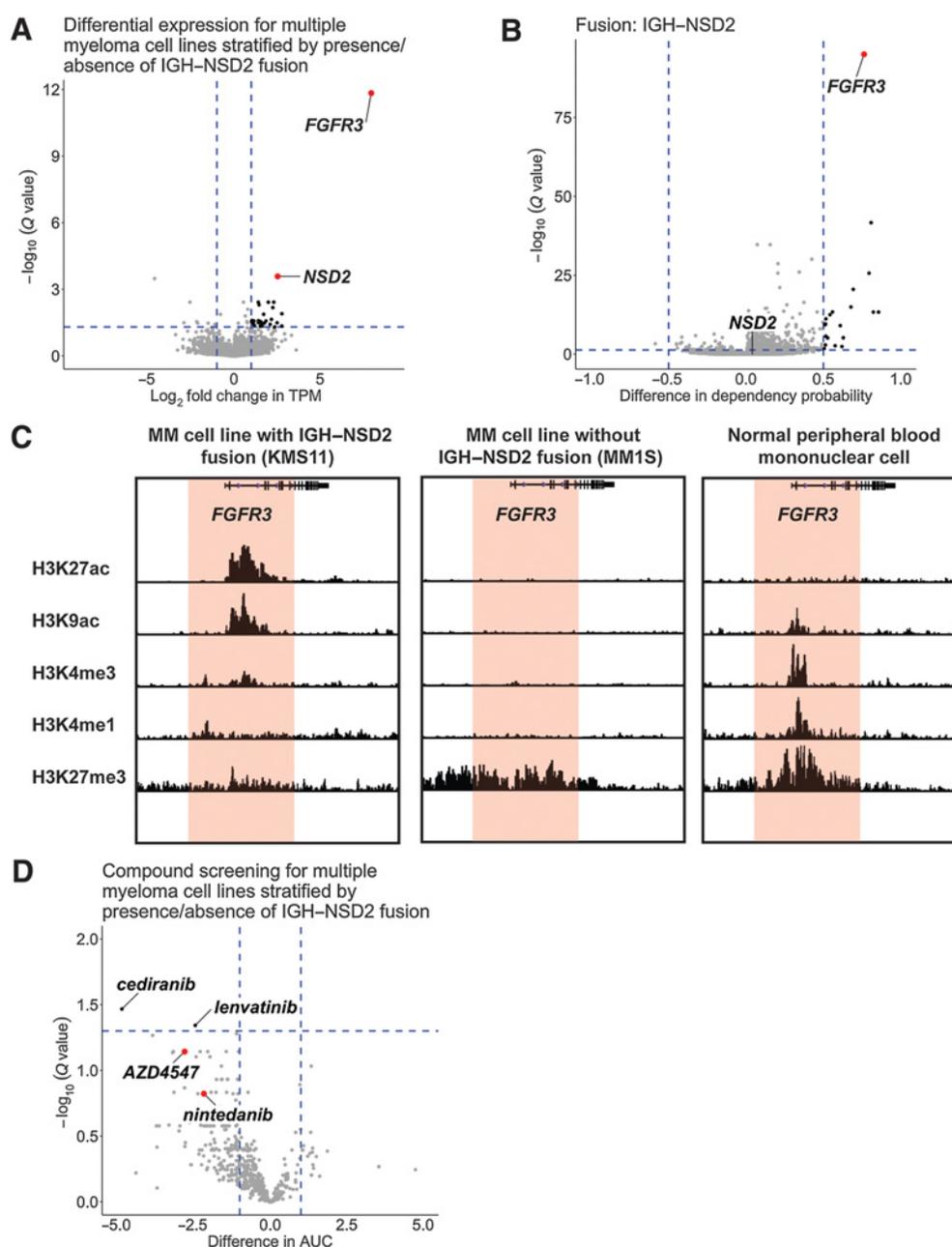


Figure 5.

IgH-NSD2 fusion is associated with *FGFR3* as a collateral dependency in multiple myeloma cell lines. **A**, Differential expression space for multiple myeloma cell lines only, stratified by the presence/absence of *IgH-NSD2* fusion, shows that both *NSD2* and *FGFR3* are overexpressed. **B**, Dependency space for five multiple myeloma cell lines with *IgH-NSD2* fusion (KMS34, KMS11, KMS26, KMS18, OPM2) in which *FGFR3* is the strongest identified dependency. **C**, Histone ChIP-seq data at the *FGFR3* locus. **D**, AUC values from an unbiased compound screen in multiple myeloma cell lines, stratified by the presence/absence of *IgH-NSD2* fusion; negative values for difference in AUC indicate more potent killing for cells with *IgH-NSD2* fusion.

also occur in a subset of clinical samples. Similarly *FGFR3*, a targetable kinase, was the key dependency in t(4;14) multiple myeloma cell lines that harbored an *IgH-NSD2* fusion. We also showed that the implications of fusion-associated dependencies extended beyond two-dimensional cell line space, exemplified by dependency on *EML4* in the context of a *THADA-MTA3* fusion persisting in spheroid models.

There were limitations in our methodology. By focusing our analysis on fusions, applying standardized TAD boundaries to

account for variability across cancer cell lines, and relying on loss-of-function screening, we likely underestimated the total impact of structural variants on cancer cell survival. Our genome-scale screen, which relied on a modified *t* test approach, was inherently limited by the non-Gaussian distribution of dependency probability scores and small numbers of cell lines with any given fusion. We addressed this limitation through an alternative permutation-based identification of fusion-associated dependencies

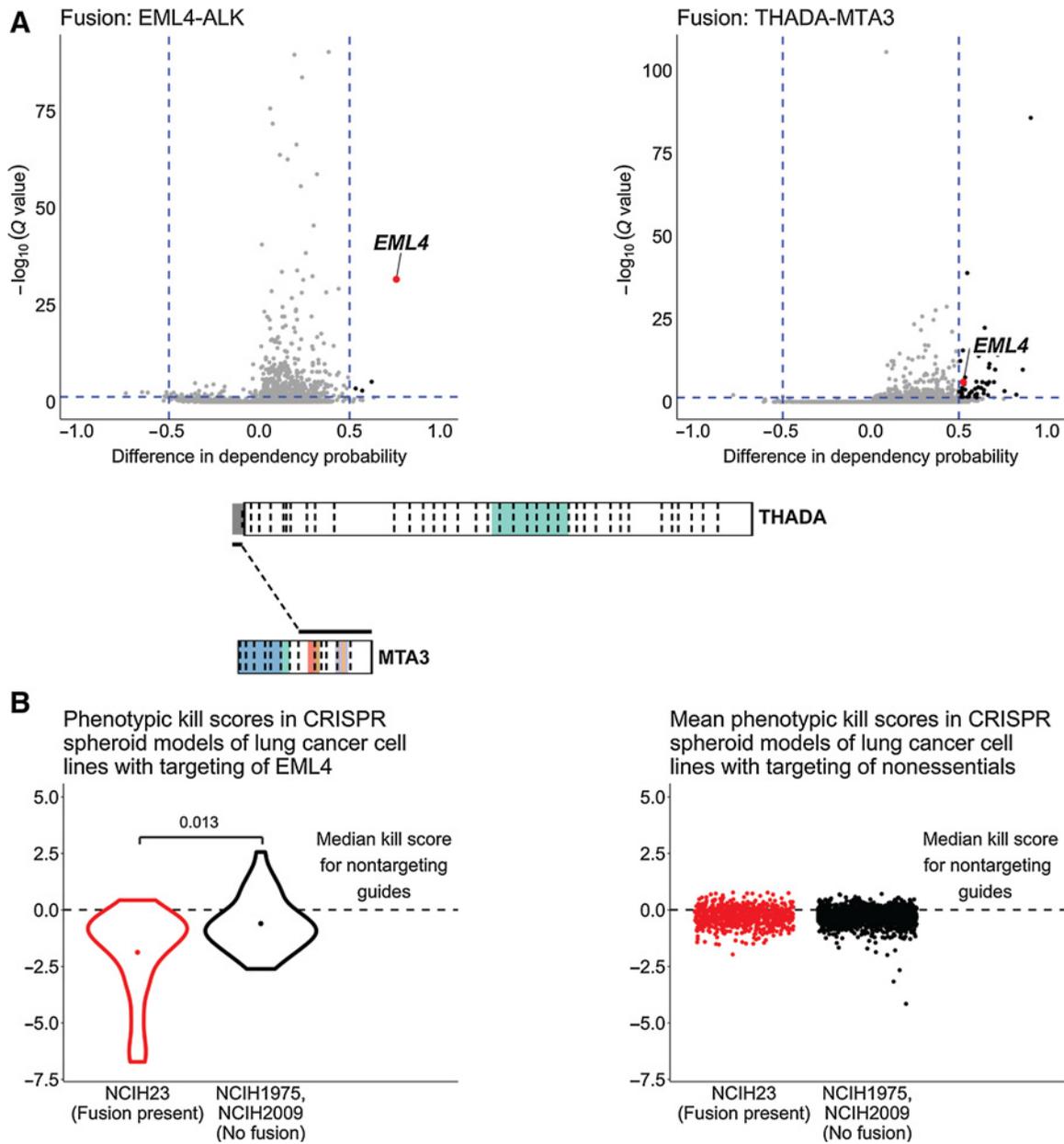


Figure 6. Spheroid models provide an opportunity for further validation of fusion-associated dependencies. **A**, Left, *EML4* is a partner dependency in the context of known COSMIC fusion *EML4-ALK* in two cell lines (C10, NCIH3122). Right, *EML4* is a collateral dependency in the context of a less well-characterized fusion *THADA-MTA3* in one cell line (NCIH23). Bottom, there is good support from RNA-seq and WGS for the presence of the *THADA-MTA3* fusion in the NCIH23 cell line. **B**, Comparison of phenotypic kill scores of CRISPR sgRNAs in spheroid models derived from NCIH23 and two cell lines without *THADA-MTA3* fusion. Left, CRISPR sgRNAs targeting *EML4* (point represents mean of each sgRNA distribution; $P = 0.013$, two-sided *t* test). Right, mean of CRISPR sgRNAs targeting all nonessential genes (each point represents mean of sgRNA distribution for each nonessential gene).

and found substantial overlap in these two approaches, but also differences to suggest that some of our fusion-associated dependencies were more likely to be false positives. Comparison of these approaches also showed that limiting hypotheses to partners or collateral genes increased the discovery of fusion-associated dependencies, suggesting our genome-scale approach may have underestimated how frequently fusions create partner and collateral dependencies.

Regarding some unbalanced rearrangements, sgRNA location relative to fusion breakpoints failed to capture what would likely be true dependency on a partner gene, reducing the sensitivity of CRISPR in identifying fusion partner genes important to cancer cell survival. In other cases, despite sgRNA location off of a fusion transcript, fusion partners still screened as dependencies. We reasoned that these examples may represent an alternate mechanism by which a translocation may contribute to a partner gene becoming a dependency

(interruption of one allele through involvement in a fusion may lead to the contralateral allele becoming essential for cell survival) or that they are balanced translocations and the reciprocal transcripts were simply not detected (Supplementary Table S17; Supplementary Figs. S21A and S21B). We sought to understand whether there were idiosyncratic effects of CRISPR–Cas9 that could lead to the creation of false-positive fusion-associated dependencies. Despite the high rate of copy-number amplification among fusion-associated dependencies, we reasoned that because copy-number correction through the CERES algorithm was incorporated into dependency probability scores, a nonspecific copy number effect was unlikely to be the primary explanation for most fusion-associated dependencies. We also considered whether CRISPR–Cas9 would differentially identify false-positive fusion-associated dependencies at fragile sites and found that only two of 363 fusions with associated dependencies had a partner in a known fragile site and that none of the dependencies themselves were located in fragile sites (59). We concluded that the rate of false-positive fusion-associated dependencies was likely to be low.

Broadly, our research provides new insight into how fusions contribute to fitness in different cancer contexts going beyond their straightforward partner-gene activation events, demonstrating that some of the identified partner and collateral dependencies may have direct implications for clinical care. Future studies are needed for further experimental validation of the regulatory elements involved in the fusion-associated dependencies identified in this work. As WGS of cancer cell lines increases, we can broaden the scope of our approach to more fully characterize the impact of structural variation on cancer cell survival.

Authors' Disclosures

R. Gillani reports grants from NIH during the conduct of the study. B.K.A. Seong reports grants from Department of Defense (CA181249) during the conduct of the study. N.V. Dharia reports grants and other support from St. Baldrick's Foundation during the conduct of the study and other support from Genentech, Inc., outside the submitted work. M.X. He reports grants from NIH and grants from NSF during the conduct of the study and personal fees from Amplify Medicines / Ikena Oncology outside the submitted work. J.S. Boehm reports grants from Cancer Dependency Map Consortium during the conduct of the study. F. Vazquez reports grants from Novo Ventures and grants from Dependency Map Consortium outside the submitted work. J.M. McFarland reports other support from the Dependency Map Consortium during the conduct of the study and other support from the Dependency Map Consortium outside the submitted work. K. Stegmaier reports grants from NIH during the conduct

of the study, grants from Novartis, personal fees from Rigel Pharmaceuticals, Kronos Bio, and AstraZeneca, and other support from Auron Therapeutics outside the submitted work. E.M. Van Allen reports personal fees from Tango Therapeutics, Genome Medical, Invitae, Monte Rosa Therapeutics, Manifold Bio, Illumina, Enara Bio, and Janssen; grants from Novartis and BMS outside the submitted work; and also has institutional patents filed on chromatin mutations and immunotherapy response, and methods for clinical interpretation pending. No disclosures were reported by the other authors.

Authors' Contributions

R. Gillani: Conceptualization, formal analysis, investigation, writing—original draft, writing—review and editing. **B.K.A. Seong:** Validation, writing—original draft. **J. Crowdis:** Conceptualization, visualization, methodology. **J.R. Conway:** Conceptualization, visualization, methodology, writing—review and editing. **N.V. Dharia:** Conceptualization, methodology, writing—review and editing. **S. Alimohamed:** Conceptualization, visualization, methodology. **B.J. Haas:** Conceptualization, methodology. **K. Han:** Resources, data curation, validation. **J. Park:** Conceptualization, writing—review and editing. **F. Dietlein:** Conceptualization, methodology. **M.X. He:** Conceptualization, methodology. **A. Imamovic:** Conceptualization, methodology. **C. Ma:** Conceptualization, methodology. **M.C. Bassik:** Resources, data curation, validation. **J.S. Boehm:** Conceptualization, methodology. **F. Vazquez:** Conceptualization, methodology. **A. Gusev:** Conceptualization, methodology. **D. Liu:** Methodology, writing—original draft. **K.A. Janeway:** Conceptualization, methodology. **J.M. McFarland:** Conceptualization, methodology, writing—review and editing. **K. Stegmaier:** Conceptualization, validation, methodology, writing—review and editing. **E.M. Van Allen:** Conceptualization, supervision, investigation, methodology, writing—original draft, writing—review and editing.

Acknowledgments

This work was supported by a Research Training Grant in Pediatric Oncology NIH T32 CA136432 11 (to R. Gillani), NIH R37 CA222574 (to E.M. Van Allen), R01 CA227388 (to E.M. Van Allen), U01 CA233100 (to E.M. Van Allen), Innovation in Cancer Informatics Award (to E.M. Van Allen), NIH 5R35 CA210030 (to K. Stegmaier), Department of Defense PRCRP Horizon Award CA181249 (to B.K.A. Seong), and Julia's Legacy of Hope St. Baldrick's Foundation Fellowship (to N.V. Dharia). The authors thank K. Salari, M. Meyerson, B. Crompton, and L. Guenther for helpful feedback on analysis and findings.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 10, 2021; revised March 26, 2021; accepted June 4, 2021; published first June 7, 2021.

References

- Wala JA, Bandopadhyay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: Genome-wide detection of structural variants and indels by local assembly. *Genome Res* 2018;28:581–91.
- Garraway LA, Lander ES. Lessons from the cancer genome. *Cell* 2013;153:17–37.
- Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010;463:899–905.
- Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. *Nature* 2018;555:321–7.
- Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* 2018;555:371–6.
- Northcott PA, Lee C, Zichner T, Stütz AM, Erkek S, Kawauchi D, et al. Enhancer hijacking activates GF11 family oncogenes in medulloblastoma. *Nature* 2014; 511:428–34.
- Weischenfeldt J, Dubash T, Drains AP, Mardin BR, Chen Y, Stütz AM, et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017;49:65–74.
- Beroukhi R, Zhang X, Meyerson M. Copy number alterations unmasked as enhancer hijackers. *Nat Genet* 2017;49:5–6.
- Zhang Y, Yang L, Kucherlapati M, Chen F, Hadjipanayis A, Pantazi A, et al. A pan-cancer compendium of genes deregulated by somatic genomic rearrangement across more than 1,400 cases. *Cell Rep* 2018;24:515–27.
- Calabrese C, Davidson NR, Demircioğlu D, Fonseca NA, He Y, Kahles A, et al. Genomic basis for RNA alterations in cancer. *Nature* 2020;578:129–36.
- Zhang Y, Chen F, Fonseca NA, He Y, Fujita M, Nakagawa H, et al. High-coverage whole-genome analysis of 1220 cancers reveals hundreds of genes deregulated by rearrangement-mediated cis-regulatory alterations. *Nat Commun* 2020;11:736.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature*; 2019;569:503–8.
- Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* 2017;49:1779–84.

15. Dempster JM, Rossen J, Kazachkova M, Pan J, Kugener G, Root DE, et al. Extracting biological insights from the Project Achilles genome-scale CRISPR screens in cancer cell lines. *bioRxiv* 2019;720243. DOI: <https://doi.org/10.1101/720243>.
16. Picco G, Chen ED, Alonso LG, Behan FM, Gonçalves E, Bignell G, et al. Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR-Cas9 screening. *Nat Commun* 2019;10:2198.
17. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol* 2019;20:213.
18. Shivram H, Iyer VR. Identification and removal of sequencing artifacts produced by mispriming during reverse transcription in multiple RNA-seq technologies. *RNA* 2018;24:1266–74.
19. Zaphiropoulos PG. Trans-splicing in higher eukaryotes: implications for cancer development. *Front Genet* 2011;2:2–5.
20. Pucker B, Brockington SF. Genome-wide analyses supported by RNA-seq reveal non-canonical splice sites in plant genomes. *BMC Genomics* 2018;19:980.
21. Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* 2020;52:294–305.
22. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
23. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 2018;46:794–801.
24. Lajoie BR, Dekker J, Kaplan N. The hitchhiker's guide to hi-c analysis: practical guidelines. *Methods* 2016;72:65–75.
25. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom R, Raney BJ, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Res* 2018;47:853–8.
26. Durinck S, Moreau Y, Kasprzyk A, Davis S, MBDe, Brazma A, et al. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005;21:3439–40.
27. Cortés-Ciriano I, Lee JJK, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* 2020;52:331–41.
28. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:1–12.
29. Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep* 2018;23:227–38.
30. Morozova O, Newton Y, Cline M, Zhu J, Learned K, Stuart J, et al. Abstract LB-212: treehouse childhood cancer project: a resource for sharing and multiple cohort analysis of pediatric cancer genomics data. *Cancer Res [Internet]* 2015;75:LB-212 LP-LB-212. http://cancerres.aacrjournals.org/content/75/15_Supplement/LB-212.abstract.
31. Nicorici D, Şatalan M, Edgren H, Kangaspeska S, Murumägi A, Kallioniemi O, et al. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv [Internet]*. 2014;11650. <http://biorxiv.org/content/early/2014/11/19/011650.abstract>.
32. Rees MG, Seashore-ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* 2016;12:109–16.
33. Seashore-Ludlow B, Rees MG, Cheah JH, Coko M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov* 2015;5:1210–23.
34. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 2013;154:1151–61.
35. Han K, Pierce SE, Li A, Spees K, Anderson GR, Seoane JA, et al. CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. *Nature* 2020;580:136–41.
36. Stephens PJ, McBride DJ, Lin M, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 2009;462:1005–10.
37. Armstrong SA, Kung AL, Mabon ME, Silverman LB, Stam RW, Den Boer ML, et al. Inhibition of FLT3 in MLL: validation of a therapeutic target identified by gene expression based classification. *Cancer Cell* 2003;3:173–83.
38. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* 2007;7:823–33.
39. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 2019;47:D941–7.
40. Schroeder MP, Bastian L, Eckert C, Gökbuget N, James AR, Sanchez JO, et al. Integrated analysis of relapsed B-cell precursor acute lymphoblastic leukemia identifies subtype-specific cytokine and metabolic signatures. *Sci Rep* 2019;9:1–11.
41. Willis TG, Dyer MJS. The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood* 2000;96:808–22.
42. Zhou S, Hawley JR, Soares F, Grillo G, Teng M, Ali S, et al. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nat Commun* 2020;11:441.
43. Missiaglia E, Williamson D, Chisholm J, Wirapati P, Pierron G, Petel F, et al. PAX3/FOXO1 fusion gene status is the key prognostic molecular marker in rhabdomyosarcoma and significantly improves current risk stratification. *J Clin Oncol* 2012;30:1670–7.
44. Santo EE, Ebus ME, Koster J, Schulte JH, Lakeman A, SP Van, et al. Oncogenic activation of FOXR1 by 11q23 intrachromosomal deletion-fusions in neuroblastoma. *Oncogene* 2012;31:1571–81.
45. Molenaar JJ, Koster J, Zwijnenburg DA, Van Sluis P, Valentijn LJ, Van Der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neurogenesis genes. *Nature* 2012;483:589–93.
46. Parolia A, Cieslik M, Chu SC, Xiao L, Ouchi T, Zhang Y, et al. Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer. *Nature* 2019;571:413–8.
47. Lee S, Hu Y, Kee S, Tan Y, Bhargava R, Lewis MT. Landscape analysis of adjacent gene rearrangements reveals BCL2L14 – ETV6 gene fusions in more aggressive triple-negative breast cancer. *Proc Natl Acad Sci* 2020;117:9912–21.
48. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* 2017;14:100–13.
49. Trudel S, Stewart AK, Rom E, Wei E, Zhi HL, Kotzer S, et al. The inhibitory anti-FGFR3 antibody, PRO-001, is cytotoxic to t(4;14) multiple myeloma cells. *Blood* 2006;107:4039–46.
50. Keats JJ, Maxwell CA, Taylor BJ, Hendzel MJ, Chesi M, Bergsagel PL, et al. Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14) (p16;q32)-positive multiple myeloma patients. *Blood* 2005;105:4060–9.
51. Lauring J, Abukhdeir AM, Konishi H, Garay JP, Gustin JP, Wang Q, et al. The multiple myeloma-associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity. *Blood* 2008;111:856–64.
52. Grand EK, Chase AJ, Heath C, Rahemtulla A, Cross NCP. Targeting FGFR3 in multiple myeloma: inhibition of t(4;14)-positive cells by SU5402 and PD173074. *Leukemia* 2004;18:962–6.
53. Friedman AA, Amzallag A, Pruteanu-Malinici I, Baniya S, Cooper ZA, Piris A, et al. Landscape of targeted anti-cancer drug synergies in melanoma identifies a novel BRAF-VEGFR/PDGFR combination treatment. *PLoS One* 2015;10:1–21.
54. Hussein Z, Mizuo H, Hayato S, Namiki M, Shumaker R. Clinical pharmacokinetic and pharmacodynamic profile of lenvatinib, an orally active, small-molecule, multitargeted tyrosine kinase inhibitor. *Eur J Drug Metab Pharmacokin* 2017;42:903–14.
55. Lim SM, Kim HR, Shim HS, Soo RA, Cho BC. Role of FGF receptors as an emerging therapeutic target in lung squamous cell carcinoma. *Futur Oncol* 2013;9:377–86.
56. Tuveson D, Clevers H. Cancer modeling meets human organoid technology. *Science* 2019;364:952–5.
57. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007;448:561–6.
58. Hart T, Brown KR, Sircoulomb F, Rottapel R, Moffat J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol Syst Biol* 2014;10:733.
59. Li Y, Roberts ND, Weischenfeldt J, Wala JA, Shapira O, Schumacher SE, et al. Patterns of somatic structural variation in human cancer genomes. *Nature* 2020;578:112–21.