

CANCER

A functional genomic approach to actionable gene fusions for precision oncology

Jun Li^{1†}, Hengyu Lu^{2†}, Patrick Kwok-Shing Ng^{3†§}, Angeliki Pantazi^{4†||}, Carman Ka Man Ip⁵, Kang Jin Jeong⁵, Bianca Amador³, Richard Tran³, Yiu Huen Tsang², Lixing Yang⁶, Xingzhi Song^{7,8}, Turgut Dogruluk², Xiaojia Ren⁴, Angela Hadjipanayis⁴, Christopher A. Bristow^{7,8}, Semin Lee⁹, Melanie Kucherlapati⁴, Michael Parfenov^{4¶}, Jiabin Tang^{7,8}, Sahil Seth^{7,8}, Harshad S. Mahadeshwar^{7,8}, Kamalika Mojumdar¹, Dong Zeng^{7,8}, Jianhua Zhang^{7,8}, Alexei Protopopov^{7,8}, Jonathan G. Seidman⁴, Chad J. Creighton¹⁰, Yiling Lu^{3,7}, Nidhi Sahni^{1,11,12}, Kenna R. Shaw³, Funda Meric-Bernstam^{3,13,14}, Andrew Futreal^{3,5,7}, Lynda Chin^{7,8,15}, Kenneth L. Scott^{2,10##}, Raju Kucherlapati^{4*#}, Gordon B. Mills^{16*##}, Han Liang^{1,5,12*#}

Fusion genes represent a class of attractive therapeutic targets. Thousands of fusion genes have been identified in patients with cancer, but the functional consequences and therapeutic implications of most of these remain largely unknown. Here, we develop a functional genomic approach that consists of efficient fusion reconstruction and sensitive cell viability and drug response assays. Applying this approach, we characterize ~100 fusion genes detected in patient samples of The Cancer Genome Atlas, revealing a notable fraction of low-frequency fusions with activating effects on tumor growth. Focusing on those in the RTK-RAS pathway, we identify a number of activating fusions that can markedly affect sensitivity to relevant drugs. Last, we propose an integrated, level-of-evidence classification system to prioritize gene fusions systematically. Our study reiterates the urgent clinical need to incorporate similar functional genomic approaches to characterize gene fusions, thereby maximizing the utility of gene fusions for precision oncology.

INTRODUCTION

Gene fusion in cancer cells represents a class of molecular aberrations, which are mainly caused by genomic translocations, insertions, and deletions or chromosomal inversions. A substantial proportion of fusion genes drive tumorigenesis and/or promote tumor progression, such as *BCR-ABL1* (1), *ETV6-NTRK3* (2), and *TMPRSS2-ETS* (3). Because of their tumor-specific expression and ability to drive tumor pathophysiology, fusion genes that are drivers rather than passengers represent target molecules with tremendous diagnostic and therapeutic potential (4, 5). With advances in next-generation sequencing technology, especially through The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium Pan-Cancer Analysis of Whole Genome efforts, the number of fusion genes detected has increased from several hundreds to >35,000 across the full spectrum of cancer types (6–9). However, the functional

impact and clinical relevance of these cancer fusion genes remain poorly understood, representing a critical knowledge gap for implementing precision cancer medicine.

A number of computational algorithms have been developed to predict the functional impact of somatic point mutations (10). The development and validation of these algorithms have benefited from a “gold standard” set of functionally characterized mutations (11). In contrast, the development and validation of computational algorithms to identify driver gene fusions are limited, at least in part, due to the lack of a gold standard set of functionally characterized fusion genes. Current studies have focused on a small set of highly recurrent gene fusions detected in patient tumors for detailed functional and mechanistic characterization. While this is a reasonable practice to identify targets that can maximize the potential clinical impact, this approach results in many rare gene fusions remaining

¹Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ³Institute for Personalized Cancer Therapy, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁴Department of Genetics, Harvard Medical School, Division of Genetics, Brigham and Women’s Hospital, Boston, MA, USA. ⁵Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁶Ben May Department for Cancer Research and Department of Human Genetics, The University of Chicago, Chicago, IL, USA. ⁷Department of Genomic Medicine, The University of MD Anderson Cancer Center, Houston, TX, USA. ⁸Institute for Applied Cancer Science, The University of MD Anderson Cancer Center, Houston, TX, USA. ⁹Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. ¹⁰Department of Medicine, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA. ¹¹Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX, USA. ¹²Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX, USA. ¹³Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁴Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ¹⁵Dell Medical School, The University of Texas Austin, Austin, TX, USA. ¹⁶Division of Oncologic Sciences, Knight Cancer Institute, Oregon Health Sciences University, Portland, OR, USA.

*Corresponding author. Email: hliang1@mdanderson.org (H. Liang); millsg@ohsu.edu (G.B.M.); rkucherlapati@bwh.harvard.edu (R.K.)

†These authors contributed equally to this work as co-first authors.

‡Present address: The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

§Present address: Department of Pediatrics, University of Connecticut Health Center, Farmington, CT, USA.

||Present address: Southern Alberta Genome Sciences Centre, University of Lethbridge, Lethbridge, AB, Canada.

¶Present address: Invitae Corporation, San Francisco, CA, USA.

#These authors contributed equally to this work as co-senior authors.

**Deceased.

unexplored. Moreover, conventional experimental studies can only characterize a small number of gene fusions in a time-consuming and costly manner, which is insufficient to address the need for informing treatment options in the clinical setting. This is largely due to difficulty in creating, expressing, and characterizing large numbers of specific gene fusions in sensitive and robust functional assays.

In this study, we developed an efficient functional genomics approach to characterize the cellular consequences of gene fusions and applied it to a large number of fusion genes. We further performed an in-depth analysis of the fusions involved in the receptor tyrosine kinase (RTK)–RAS signaling pathway in terms of their functional consequences and potential therapeutic liabilities. Our approach demonstrates a potential way to fill the gap between sequencing-based fusion detection and target-based clinical actions.

RESULTS

Study overview

To identify clinically actionable gene fusions for implementation of precision oncology, we developed an efficient functional genomics

approach to assess the functional impact of gene fusions and applied it to fusion genes detected in TCGA patient samples on a large scale (Fig. 1A). First, we collected ~35,000 gene fusion events from >11,000 TCGA patients across 33 cancer types based on two previous bioinformatics studies (6, 8). We selected 110 gene fusions (mainly from TCGA list and an additional set from MD Anderson patients), with considerations on their related pathways and potential druggability of the involved partner genes. Second, we generated fusion gene constructs using a Gateway multifragment recombination technique, as previously described (12). The quality of the fusion constructs was validated by directly sequencing the corresponding plasmids. We included only those constructs that passed the quality control step for subsequent assessment. Third, for each fusion gene, we evaluated its effect on cell viability through in vitro assays in informer cell lines (i.e., Ba/F3 and/or MCF10A). For selected candidates, we also performed xenograft tumor assays to assess their tumorigenic activity in vivo. Last, for positive hits identified above, we performed drug response assays to determine their therapeutic implications.

In total, we obtained high-quality functional data of 90 fusions (table S1). To illustrate the clinical relevance of fusion partner genes,

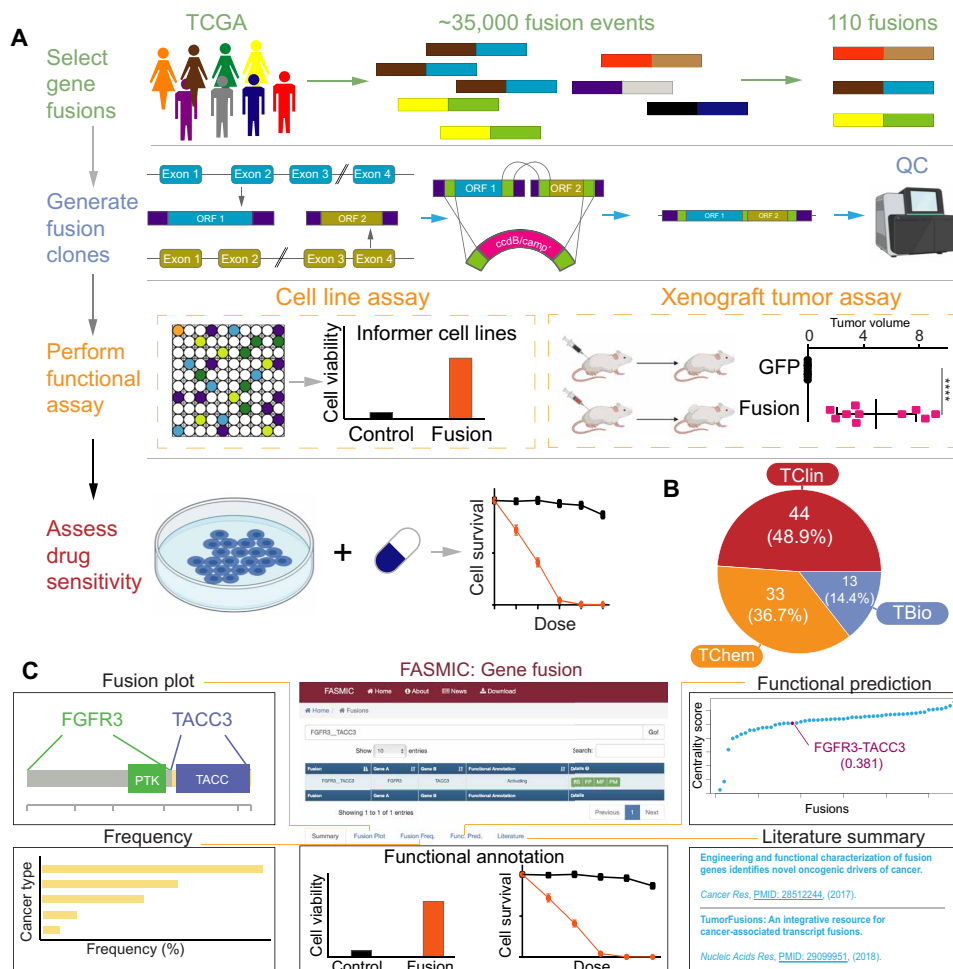


Fig. 1. Overview of our functional genomics approach and the fusion candidates assayed. (A) Schematic representation of the functional genomics platform for fusion genes. ORF 1, open reading frame 1; GFP, green fluorescent protein. *****P* < 0.0001. (B) The Pharos target classification for 90 gene fusions. TClIn, genes with at least one FDA-approved drug; TChem, genes with at least one ChEMBL compound with an activity cutoff of <30 nM; and TBio, genes without known drug or small-molecule activities. (C) Snapshot of the FASMIC: Gene fusion data portal.

we annotated each partner gene using the Pharos database, a knowledge base illuminating the druggable genome (13). Among the 90 fusion genes, 44 (48.9%) fusions had at least one partner categorized as T_{Clin}, which is associated with at least one U.S. Food and Drug Administration (FDA)–approved drug; 33 (36.7%) contained T_{Chem} targets, which had at least one ChEMBL compound with an activity cutoff of <30 nM; and 13 (14.4%) fusions did not have partners associated with known drug or small-molecule activities, annotated as T_{Bio} (Fig. 1B). Notably, none of our fusions have both partners annotated as T_{Dark} about which nothing is known by definition in Pharos.

To facilitate a broad biomedical community to capitalize on our results, we developed a user-friendly web portal for data visualization and exploration (Fig. 1C). The data portal contains five modules: (i) “Functional annotation” provides the experimental data related to a specific gene fusion, including cell viability and drug response assays; (ii) “fusion plot” displays the fusion gene structure in an intuitive graph with annotated position and protein domain information; (iii) “fusion frequency” shows how frequently the fusion occurs in different cancer types; (iv) “functional prediction” shows the predicted functional impact score for the fusion; and (v) “literature summary” shows the functional and clinical evidence about the fusion that has been reported in the literature. Users can easily search for a gene fusion of interest through keywords or download all data for their analysis. The data portal is available at <https://bioinformatics.mdanderson.org/public-software/fasmic/>.

Large-scale functional annotation of fusion genes in cancer

In our functional genomics approach, we mainly used Ba/F3 cells as the “informer” cell line to assess the effect of a specific gene fusion on cell viability. Ba/F3 is a murine pro-B cell line that depends on interleukin-3 (IL-3) for growth and proliferation but can be readily transformed to IL-3 independence in the presence of an oncogenic event, thereby making it widely used for detecting oncogenic events (14). The ability to assay transfer of Ba/F3 addiction from IL-3 to oncogenes has already been used to investigate the activity and therapeutic sensitivity of a variety of oncogenes (15–17). On the basis of whether a gene fusion exhibited activating effects in Ba/F3 cells compared to the negative controls (mCherry and/or luciferase), we classified the 90 gene fusions into two functional categories: 24 activating fusions and 66 fusions with no significant effects. Since the phenotypic effect may depend on cellular contexts, we performed similar cell viability assays for many fusions with no effects in the Ba/F3 cell line using MCF10A, a nontumorigenic breast epithelial cell line that is widely used in anchorage-independent growth assays to quantitate transforming activity. We randomly chose 30 fusions showing no significant effects in the Ba/F3 assay and found that none of them were activating in MCF10A, suggesting that the false-negative rate of our platform is low. In addition, we assessed the functional impact of 132 wild-type genes based on gene overexpression assays in the two model cell lines (Ba/F3 and MCF10A). By overlapping with our fusion annotation, 14 of the 24 (58.3%) activating fusions had at least one partner gene that demonstrated positive (activating) effects on cell viability, whereas only 9 of the 66 (13.6%) no-effect fusions had a partner gene with a positive effect. This pattern indicates that the activating fusions tend to involve genes with oncogenic potential (χ^2 test, $P = 5.7 \times 10^{-5}$; Fig. 2A). Among 140 unique partner genes from the 90 fusion genes, the top four common fusion partner genes were *RET*, *FGFR2*, *BRAF*, and

ALK. For example, four of the five *RET* fusions, three of the five *ALK* fusions, and both *NTRK2* fusions showed activation activity in the reporter cell lines. In terms of fusion recurrence in >11,000 TCGA samples, we found that 25.4% of singleton fusions (those only detected in a single patient sample) showed an activating effect lower than that of recurrent fusions (35.0%). Among recurrent fusions, 16 appeared in multiple cancer types, with *FGFR3-TACC3*, *ETV6-NTRK3*, and *EML4-ALK* being among the top activating hits (Fig. 2B). For recurrent fusions, the frequency in TCGA patients was significantly higher in the activating group than in the no effect group (Wilcoxon test, $P = 2.9 \times 10^{-2}$), consistent with the notion that oncogenic fusions are likely under positive selection. Although highly recurrent gene fusions are more likely to have a functional impact, our results suggest that a considerable proportion of low-frequency or singleton fusions are potential driver events that may be critical for specific tumors and have therapeutic consequences.

To gain insights into the system-level properties of activating fusion genes, we compared our experiment-based functional annotation with the priority scores that were predicted by a computational method based on the properties of fusion partner genes in the biological network (18). Fusions with high centrality scores were significantly enriched in the activating fusions identified by our approach, highlighting the consistency of our experimental approach with the computation predictions (Wilcoxon test, $P = 2.4 \times 10^{-3}$; Fig. 2C). Of the fusion genes assayed, the partner genes of 50 fusions (55.6%) resided in the same chromosome. We performed an enrichment analysis to assess whether the functional effects of the fusions correlated with the distances between their partner genes. Activating fusions tended to have gene partners from different chromosomes (χ^2 , $P = 5.1 \times 10^{-3}$; Fig. 2D). This is probably because cis-fusions (gene partners from the same chromosome or neighbor genes) resulting from transcription read through are more likely to be passenger events.

Gene fusions in the RTK-RAS signaling pathway

The RTK-RAS pathway is frequently disrupted in diverse cancer types, and it is estimated that as many as 40% of all tumors have aberrations in genes of the RAS family (19). Nearly half of the fusions that we assayed (42 fusions) had at least one partner gene from the RTK-RAS pathway, and therefore, we focused on this pathway for an in-depth analysis (our tested gene fusions are highlighted; Fig. 3A). On the basis of TCGA PanCanAtlas dataset, we performed a systematic analysis of the gene fusions of RAS pathway members that include 272 genes, mainly defined by the RAS initiative (20). In the 9105 TCGA tumors, 1289 (14.2%) tumors harbored at least one such fusion (Fig. 3B). The highest rates of RAS pathway fusions were found in sarcoma (on average, 0.8 RAS pathway fusions per tumor), ovarian cancer (0.6), and glioblastoma (0.6), and the lowest rates were found in kidney chromophobe cancer (0.02) and thymoma cancer (0.02). In particular, 297 tumors contained one or more RAS pathway fusions without a mutation in RAS pathway members. These tumors carry potentially driver fusions of known oncogenes, such as *ALK* (*TPM1-ALK*), *FGFR3* (*FGFR3-ELAVL3*), *BRAF* (*MACF1-BRAF*), and *AKT3* (*ACV2RA-AKT3*, *ZEB2-AKT3*) as well as fusions that may affect the function of RAS inhibitors (*RASAL2* and *SPRED2*) and RAS-regulated tumor suppressor genes (*PTEN*, *RBI*, *TP53*, *RAD52*, and *FANCC*). The results suggest that these fusions play a role in tumor initiation and/or maintenance through activation of the RTK-RAS pathway. Among the 992 tumors with co-occurrence of mutations

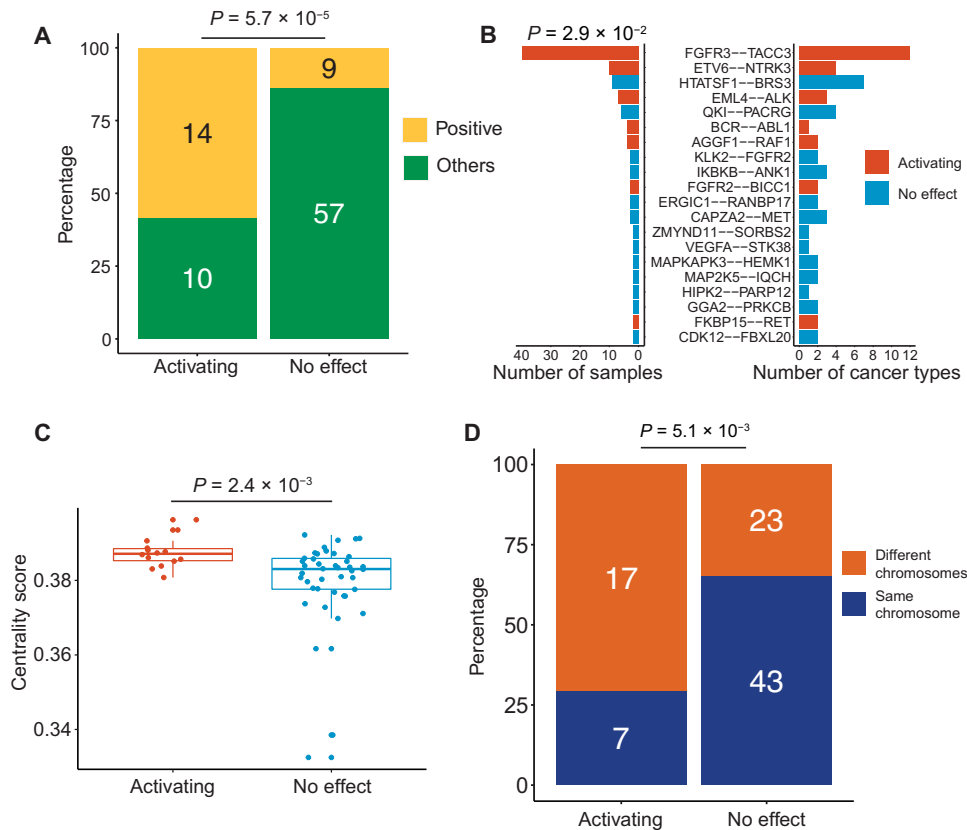


Fig. 2. Summary of our large-scale functional annotation of gene fusions in cancer. (A) Proportions of partner genes exhibiting activating effects in activating fusions versus no-effect fusions. *P* value is based on a chi-square test. (B) Numbers of TCGA samples (left) and cancer types (right) associated with characterized recurrent fusions (detected in ≥ 2 samples). The activating fusions are labeled in red; otherwise, in blue. *P* value is based on a Wilcoxon test. (C) Comparison of centrality scores between activating and no-effect fusions. *P* value is based on a Wilcoxon test. (D) Distribution of partner genes in the same or different chromosomes between activating fusions and no-effect fusions. *P* value is based on a chi-square test.

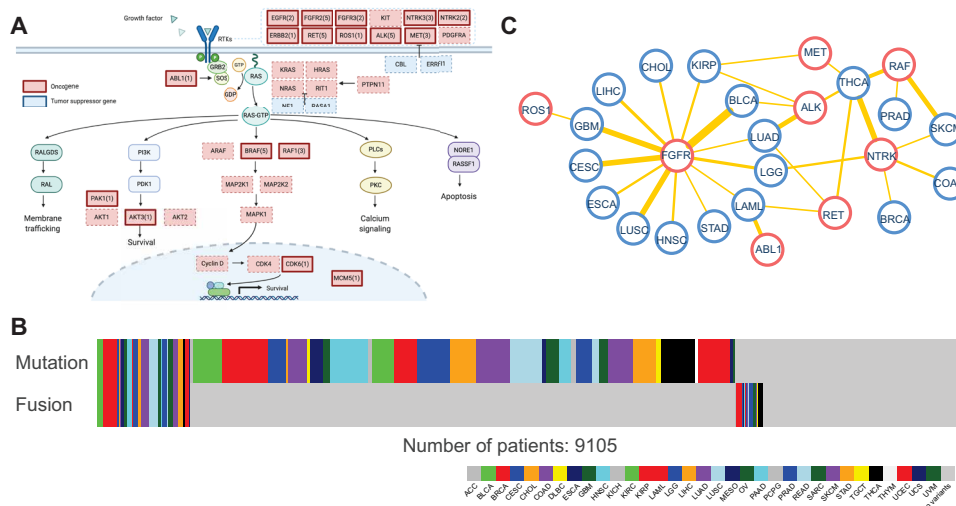


Fig. 3. Summary of assayed fusion genes in the RTK/RAS pathway. (A) All 42 RTK/RAS fusions are highlighted in solid rectangular boxes with the numbers in the brackets indicating the number of corresponding gene fusions assayed. (B) Summary of all tumors across 33 TCGA cancer types, including cases that carry gene fusions and/or mutations in genes of the RTK-RAS pathway ($n = 9105$). The presence of aberration is displayed on the *Y* axis, while all patients are displayed on the *X* axis. Color indicates tumor type. Gray indicates the absence of RTK-RAS aberration. (C) The bipartite network shows the occurrence of each fusion family across cancer types. The width of each edge is proportional to the number of patients harboring the linked fusion.

and fusions in the RTK-RAS pathway, we observed cases of known recurrent activating fusion genes, such as *EML4-ALK* in lung adenocarcinomas and *FGFR3-TACC3* in bladder tumors. This observation implies that fusion genes in these tumors may cooperate with point mutations in RAS pathway members rather than representing passenger events. For instance, we were able to demonstrate the oncogenic capacity of *TRAK1-RAF1* gene fusion in a patient with melanoma who also carried nonsynonymous mutations in *EGFR*, *NF1*, *PAK3*, *PIK3CG*, *PIK3R6*, and *PLCE1*. Therefore, a subset of previously unidentified fusions that co-occur with mutations may also make a notable contribution to activating RTK-RAS signaling.

Figure 3C shows the frequency summary of the 42 gene fusions that we assayed in TCGA patient cohort ($n = 9105$). Specifically, *FGFR* (five *FGFR2* and two *FGFR3* fusions) had the highest fusion frequency ($n = 44$), detected in 13 cancer types, and *NTRK* fusions (two *NTRK2* and three *NTRK3* fusions; $n = 12$) occurred in five cancer types. Among these gene fusions, *ALK*, *BRAF*, *RET*, and *FGFR* were the most frequent partner genes, each with five fusions, and therefore, we discussed them in detail.

ALK gene fusions

ALK [which encodes for the anaplastic lymphoma kinase (ALK) tyrosine kinase receptor] gene fusions have been reported in a wide variety of solid tumors (21). In the case of *EML4* and other *ALK* gene fusions, dimerization of the coiled-coil domain of the 5' partner results in activation of the tyrosine kinase function of *ALK* (22). In addition to the well-described *ALK-EML4* rearrangements in the lung and other types of cancer, *TPM1-ALK* was also identified in bladder cancer. We detected *TPM1-ALK* mature mRNA in-frame transcripts at the RNA level, which was validated by discordant paired-end reads, indicating a reciprocal interchromosomal translocation between the *TPM1* and *ALK* loci. Through this event, most of the coding sequence of *TPM1*, including its coiled-coil domain, is fused to the intracellular C terminus of *ALK* that includes its protein kinase domain (Fig. 4A).

We next assessed the functional effects of the *TPM1-ALK* gene fusion in our Ba/F3 cell viability assays. *TPM1-ALK* strongly enhanced Ba/F3 cell survival and proliferation by 70-fold compared to green fluorescent protein (GFP)-expressing cells ($P < 10^{-4}$) as early as 7 days following the removal of IL-3 (Fig. 4B). We also assessed the transforming activity of full-length *TPM1* and *ALK* in the Ba/F3 assay. Despite quantitative polymerase chain reaction (PCR)-verified expression (Fig. 4C), neither *TPM1* nor *ALK* expression was able to promote Ba/F3 survival and proliferation in contrast to the expression of the fusion gene (Fig. 4D). *TPM1-ALK* promoted cell growth at a level comparable to *EML4-ALK* (Fig. 4B), suggesting their related activities, which is further supported by their similar and robust activation of signaling through signal transducer and activator of transcription 3 (Stat3), as assessed by immunoblot analysis of phospho-Stat3 (Y705) in outgrowth Ba/F3 cell lysates (Fig. 4E). We next performed Ba/F3 dose-response assays to examine the response of *TPM1-ALK* to an ALK inhibitor. Ba/F3 cells expressing *TPM1-ALK* exhibited marked sensitivity to crizotinib [median inhibitory concentration (IC_{50}) = 16.5 nM], an ALK kinase inhibitor used clinically to treat ALK-positive lung cancer patients (23), compared to control cells (Fig. 4F).

To generalize the impact of *ALK* fusions described above on the sensitivity to ALK inhibitors, we compared the public cell line drug sensitivity data of crizotinib and another ALK inhibitor, NVP-TAE684, between cell line samples with or without *ALK* fusions and

found that the cell lines with *ALK* fusions were significantly more sensitive to ALK inhibitors (crizotinib, Wilcoxon test, $P = 1.4 \times 10^{-4}$; Fig. 4G; and NVP-TAE684, Wilcoxon test, $P = 9.0 \times 10^{-5}$; Fig. 4J). We next assessed *ALK* expression across cell lines and observed that cell lines with *ALK* fusions had increased *ALK* expression (Wilcoxon test, $P = 2.2 \times 10^{-6}$; fig. S1A). Although the *ALK* expression level was significantly correlated with drug sensitivity to ALK inhibitors in cell line groups with or without *ALK* fusions, the correlation within *ALK* fusion cell lines [crizotinib, Pearson's correlation coefficient (R) = -0.72 and NVP-TAE684, Pearson's R = -0.74; Fig. 4, H and K] was much stronger than in cell lines without *ALK* fusions (crizotinib, Pearson's R = -0.52 and NVP-TAE684, Pearson's R = -0.50; Fig. 4, I and L, and fig. S1, B and C), suggesting that, in tumors with *ALK* fusions, *ALK* expression is more effective in predicting the response to ALK inhibitors and likely represents greater dependence on *ALK* as a driver. Collectively, these results highlight *ALK* fusions as promising predictive markers for response to ALK inhibitors in general.

RAF1 and BRAF fusions

Previous studies have reported several *BRAF* (B-Raf proto-oncogene, serine/threonine kinase) and *RAF1* (Raf-1 proto-oncogene, serine/threonine kinase) gene fusions with different partners in melanoma, thyroid, and prostate cancer (24). Our analysis identified *BRAF/RAF1* fusions with multiple partners, among which were the known gene fusions *TAX1BP1-BRAF* and *TRAK1-RAF1* in melanoma (7, 25). To study the functional impacts of *BRAF/RAF1* fusions comprehensively, in addition to the above two fusions, we selected another three *BRAF/RAF1* fusions from TCGA cohort that retain an intact protein kinase domain: *CLCN6-RAF1* (TCGA-ER-A19L), *CDC27-BRAF* (TCGA-FS-A1ZU), and *ATG7-BRAF* (TCGA-BF-A5EP) (fig. S2). Our Ba/F3 assays revealed robust driver activity ranging from 30- to 142-fold ($P < 10^{-4}$) in the absence of IL-3 compared to GFP-expressing cells (Fig. 5A). *BRAF* fusion activity in Ba/F3 was similar to cells expressing oncogenic *BRAF*^{V600E}, a well-characterized oncogenic event at hotspot V600 that is found in 46% of melanoma samples (25). Immunoblot analysis of outgrowth Ba/F3 cell lysates verified *RAF1* and *BRAF* fusion expression and enhanced mitogen-activated protein kinase (MAPK) signaling, assessed by phosphorylation of extracellular signal-regulated kinase 1/2 (ERK1/2) (T202/Y204), which was also activated by the expression of *BRAF*^{V600E} but not GFP control (Fig. 5B). Consistent with previous observations, the presence of *BRAF* and *RAF1* gene fusions are mutually exclusive to activating *BRAF/RAS* mutations that together occur in 70% of melanoma cases (Fisher's exact test, $P = 0.018$; Fig. 5C) (7, 25), which further supports that these fusions represent driver events in this disease.

To study the therapeutic utility of these fusions, we performed dose-response inhibitor assays that revealed marked sensitivity to the MAPK kinase (MEK) inhibitor, trametinib (IC_{50} , *CLCN6-RAF1* = 1.223 nM and *TRAK1-RAF1* = 0.9588 nM), in Ba/F3 cells expressing *RAF1* fusions (Fig. 5D), as predicted from the fusion's strong activation of MAPK signaling (Fig. 5B). We next sought to validate the transforming activity of the melanoma-derived fusion genes using primary human melanocytes (HMELs) (26, 27), which are ideal for assessing the tumorigenic activity of melanoma-specific oncogenes. Orthotopic injection of HMEL cells expressing *CLCN6-RAF1* into the dermis of athymic mice revealed its potent oncogenic activity (100% tumor penetrance, $n = 10$; Fig. 5E) compared to GFP-expressing control cells ($n = 0$ of 10). We also observed that cell

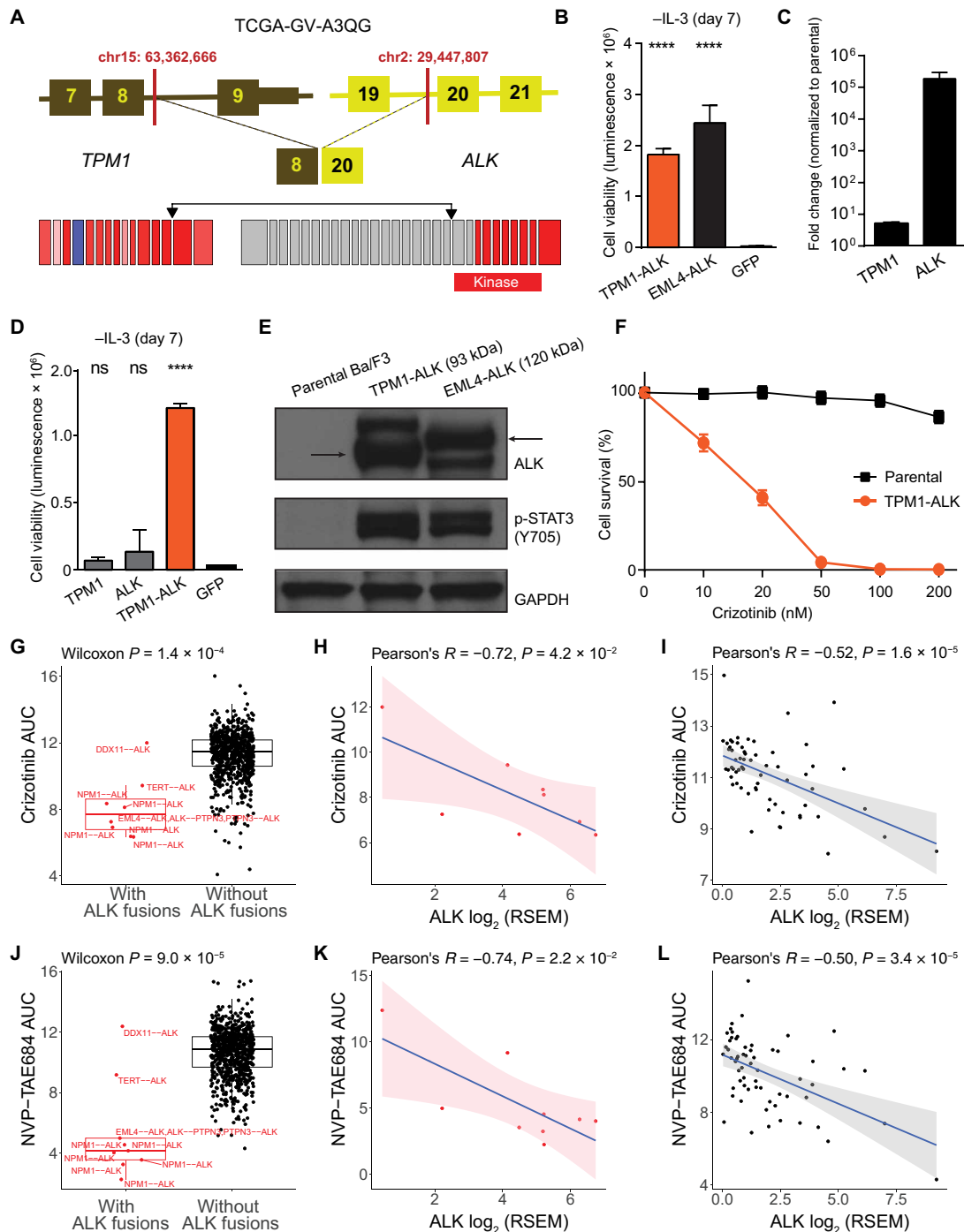


Fig. 4. Effects of ALK fusions on cell viability and drug response. (A) Schematics of the genomic structure and mRNA transcript of the *TPM1-ALK* fusion gene. Breakpoints in both genes are indicated by linked arrows. The mRNA expression levels relative to expression levels in the tumor type at the exon level are illustrated. The red bar indicates exons coding for the tyrosine kinase domains. (B) Ba/F3 cell survival assay for *TPM1-ALK* (mean luminescence, error bars denote SD, $n = 3$) compared to the positive control, *EML4-ALK*. GFP = negative control. (C) Quantitative PCR-based expression of full-length wild-type *TPM1* and *ALK* expressed in Ba/F3 (mean expression fold change normalized to the parental negative control, error bars denote SD, $n = 3$). (D) Ba/F3 cell survival assay for full-length wild-type *TPM1* and *ALK* compared to *TPM1-ALK* (mean luminescence, error bars denote SD, $n = 3$). (E) Immunoblots of *TPM1-ALK* and *EML4-ALK* expression and signaling pathway activation in Ba/F3. Arrows denote the correct sizes of *TPM1-ALK* and *EML4-ALK*. (F) Dose-dependent survival assays of Ba/F3 cells expressing *TPM1-ALK* treated with crizotinib for 72 hours (mean percentage of cell survival, error bars denote SD, $n = 4$). (G) Crizotinib drug sensitivity comparison between cell lines with and without *ALK* fusions. P value is based on a Wilcoxon test. (H and I) Correlation analysis between crizotinib drug sensitivity and *ALK* gene expression in cell lines with *ALK* fusions (H) or without *ALK* fusions (I). AUC, area under the curve. (J) NVP-TAE684 drug sensitivity comparison between cell lines with or without *ALK* fusions. (K and L) Correlation analysis between NVP-TAE684 drug sensitivity and *ALK* gene expression in cell lines with *ALK* fusions (K) or without *ALK* fusions (L). (B and D) All P values are calculated by t test; ns, not significant; **** $P < 10^{-4}$. (G to L) Drug data: CTRPv2, RSEM, RNA-seq by expectation maximization.

Downloaded from https://www.science.org at University of Texas Md Anderson Cancer Center on February 09, 2022

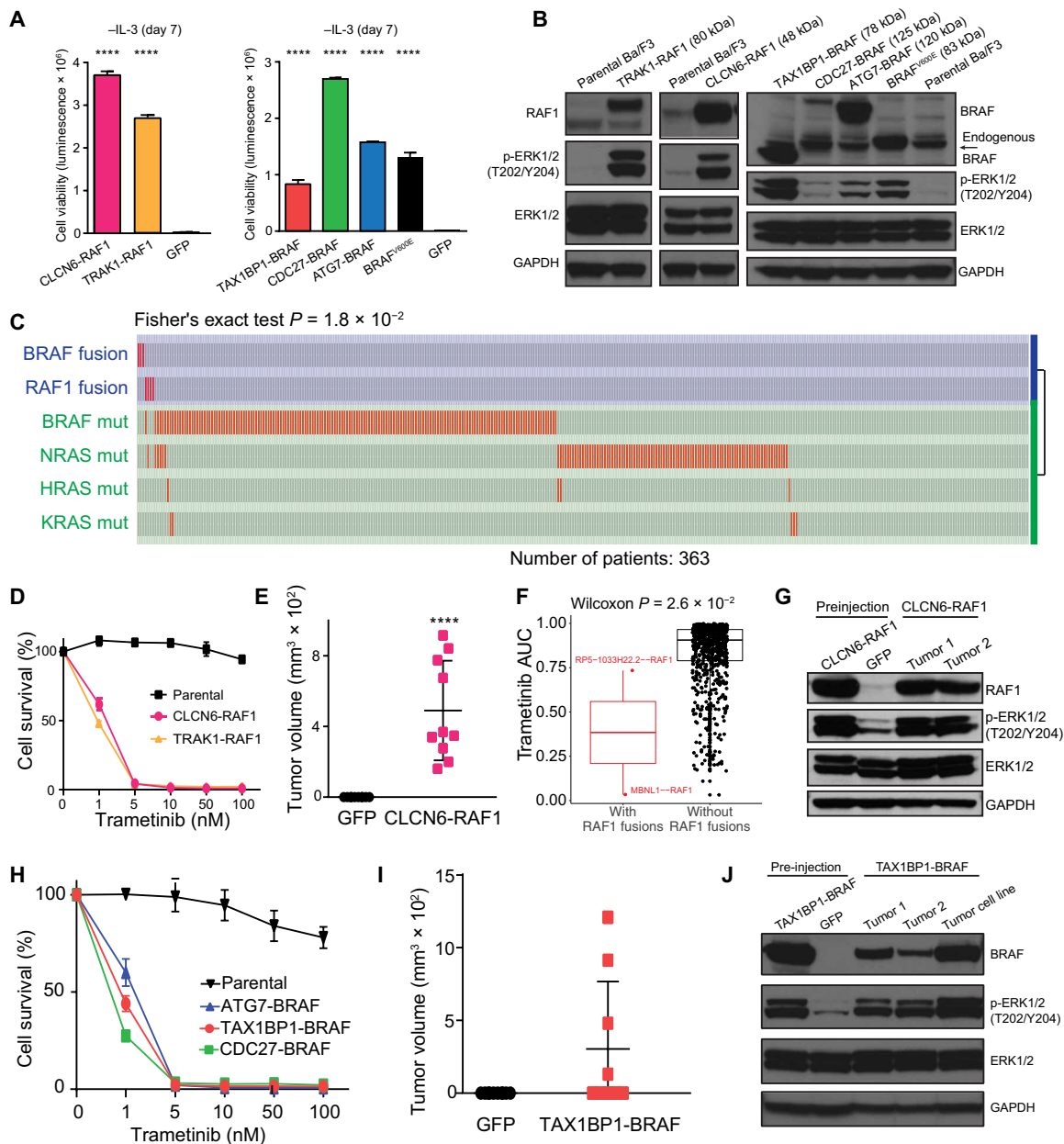


Fig. 5. Effects of *RAF1* and *BRAF* fusions on cell viability and drug response. (A) Ba/F3 cell survival assay for *RAF1* and *BRAF* fusions (mean luminescence, error bars denote SD, $n = 3$ for each group). *BRAF*^{V600E} = positive control and GFP = negative control. (B) Immunoblots of *RAF1* and *BRAF* fusion expression and MAPK signaling activation in Ba/F3. (C) Distribution of the indicated fusion and somatic mutation events across TCGA melanoma cases. Samples with activating *RAS/BRAF* mutations are shown in red. (D) Dose-dependent survival assays of Ba/F3 cells expressing *RAF1* fusions treated with trametinib for 72 hours. (E) End-point volumes (*CLCN6-RAF1*, day 23 after injection) of tumors by HMEC cells expressing *CLCN6-RAF1* ($n = 10$) and GFP control ($n = 10$). (F) Trametinib drug sensitivity comparison between cell lines with or without *RAF1* fusions. P value is based on a Wilcoxon test. Drug data: GDSC. (G) Immunoblot analysis of lysates from preinjected HMEC cells and tumors expressing *CLCN6-RAF1*. (H) *BRAF* fusions treated with trametinib for 72 hours (mean percentage of cell survival, error bars denote SD, $n = 4$ for all groups). (I) End-point volumes (*TAX1BP1-BRAF*, day 43 after injection) of tumors by HMEC cells expressing *TAX1BP1-BRAF* ($n = 10$) and GFP control ($n = 10$). Horizontal bars denote mean volumes; error bars denote SD. (J) Immunoblot analysis of lysates from preinjected HMEC cells, tumors, and tumor-derived cell line expressing *TAX1BP1-BRAF*. (A and E) All P values calculated by t test; **** $P < 10^{-4}$.

lines with endogenous *RAF1* fusions were more sensitive to trametinib using cell line drug sensitivity data (Wilcoxon test, $P = 0.026$; Fig. 5F), consistent with the *RAF1* fusions being drivers of tumor behavior. This also suggested that *RAF1* fusions are generalizable as predictive markers for a possible benefit from trametinib and likely

other *RAS*/MAPK pathway inhibitors. In support of this contention, all three *BRAF* fusions sensitized Ba/F3 cells to trametinib (IC_{50} , *TAX1BP1-BRAF* = 0.9107 nM, *CDC27-BRAF* = 0.6334 nM, and *ATG7-BRAF* = 1.148 nM; parental = 45 nM; Fig. 5H). We observed a similar effect upon injecting HMEC cells stably transduced with

one *BRAF* fusion, *TAX1BP1-BRAF*, which greatly promoted tumor growth (Fig. 5I) in four of the nine injections compared to control cells that did not form tumors ($n = 0$ of 10). Immunoblot analysis of cell and tumor lysates verified *CLCN6-RAF1* and *TAX1BP1-BRAF* expression and activation of MAPK signaling compared to GFP-expressing HME1 cells (Fig. 5, G and J). These results suggest that, despite the notable diversity of the partners, *RAF1* and *BRAF* fusions represent a key class of actionable events in melanoma and potentially in other tumor lineages.

FGFR fusions

Oncogenic rearrangements in fibroblast growth receptor genes *FGFR1*, *FGFR2*, *FGFR3*, and *FGFR4* have been reported previously (28). *FGFR* gene fusions were identified in 20 cancer types in TCGA cohort, including recurrent fusions of *WHSC1L1-FGFR1*, *FGFR2-BICC1*, and *FGFR3-TACC3* (Fig. 6A). Consistent with previous studies (29–31), *FGFR2-BICC1* and *FGFR3-TACC3* showed significant activating effects in our Ba/F3 viability assays. One interesting case is *FGFR3-ELAVL3* fusion detected in low-grade glioma, in which the genomic breakpoint lies within the 3' untranslated region (UTR) of *FGFR3*. Since the fusion transcript *FGFR3-ELAVL3* is expected to retain the stop codon of *FGFR3*, translation of this transcript would not be predicted to yield a fusion protein. However, with alternative mRNA splicing, the last exon with the 3'UTR of *FGFR3* would be spliced out, and the previous exon of *FGFR3* is thus fused to exon 2 of *ELAVL3*, resulting in an in-frame transcript (Fig. 6B). Previous studies have shown that the oncogenic potential of *FGFR3-TACC3* fusion acts through the loss of microRNA control of the 3'UTR of *FGFR3* (32); similarly, the fusion described here would serve to diminish the regulation of *FGFR3* by tumor suppressor microRNAs.

We assessed the driver potential of the *FGFR3-ELAVL3* fusion in MCF10A cells using a colony assay. We found that expression of *FGFR3-ELAVL3* led to a significant increase in colony formation (12-fold, $P = 5 \times 10^{-4}$; Fig. 6C) similar to the oncogenic *PIK3CA*^{H1047R} control (33). Immunoblot analysis (Fig. 6D) of MCF10A cells expressing *FGFR3-ELAVL3* showed elevated phosphorylation of ERK1/2 (T202/Y204) and S6 (S235/236), which is consistent with *FGFR3*'s known role in activating the MAPK pathway (34) and observed for other *FGFR3* gene fusions, such as the well-described *FGFR3-TACC3* event (35) that is present in a variety of cancer types.

An evidence-based classification to prioritize actionable gene fusions for precision oncology

On the basis of the functional genomics approach that we developed and insights obtained from the analyses above, we propose an integrated, level-of-evidence classification system to systematically prioritize gene fusions to communicate the clinical utility of individual fusion events, which consists of four levels (Fig. 7A). Specifically, level 4 (L4): fusion presence evidence, gene fusions identified through tumor sequencing data; L3: correlation-based evidence, fusions showing a significant association with drug sensitivity as a group, e.g., *RAF1* fusions to trametinib (Fig. 5F); L2: functional evidence, fusions exhibiting significant impact on tumor growth or drug response in vitro and/or in vivo, e.g., *TAX1BP1-BRAF1* (Fig. 5, D and H); and L1: clinical evidence, fusions with evidence from patient data that they predict treatment response, e.g., *EML4-ALK* in lung cancer. The lower the level, the more compelling the evidence is for treating patients based on specific gene fusions identified in their tumors.

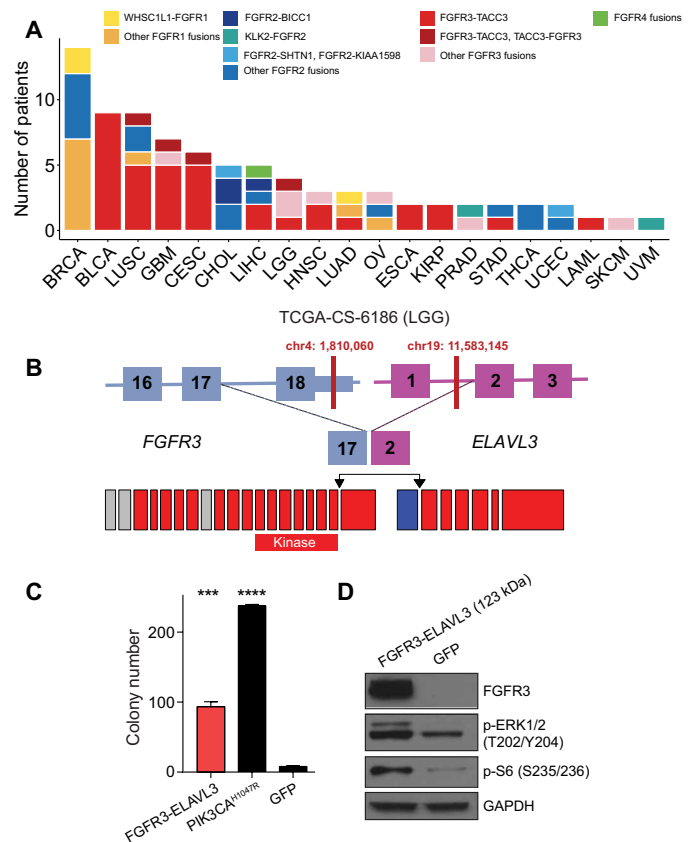


Fig. 6. Frequency and functional effects of *FGFR* fusions. (A) Summary of *FGFR* fusions across TCGA Pan-Cancers. (B) Schematics of the genomic structure and mRNA transcripts of *FGFR3-ELAVL3*. (C) MCF10A anchorage-independent colony formation assays for *FGFR3-ELAVL3* (mean colony count from 10 random areas, error bars denote SD, $n = 3$). *PIK3CA*^{H1047R} = positive control and GFP = negative control. (D) Immunoblots of *FGFR3-ELAVL3* expression and signaling pathway activation in MCF10A. All P values calculated by t test; *** $P < 0.005$ and **** $P < 0.0001$. BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma and endocervical adenocarcinoma; LGG, low-grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PRAD, prostate adenocarcinoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma; and UVM, uveal melanoma.

We next applied the above classification to the complete TCGA patient cohort with a focus on the RTK-RAS pathway, for which we had the most abundant functional evidence. In total, from >11,000 patients, we identified 1653 L4 fusions affecting members of the RTK-RAS pathway in 1450 patients, ranging from 37.2% ($n = 97$) in sarcoma to 0.9% in kidney chromophobe ($n = 1$); 177 L3 fusions in 237 patients; 23 L2 fusions in 86 patients based on the functional evidence generated in this study (e.g., 44 patients carried *FGFR* fusions); and five L1 fusions in 129 patients (Fig. 7B). This analysis elucidates how our functional genomics platform can empower a systematic approach to identify potentially clinically actionable fusion events. To increase the clinical impact, we have optimized each step in our pipeline so that the whole process from identification of

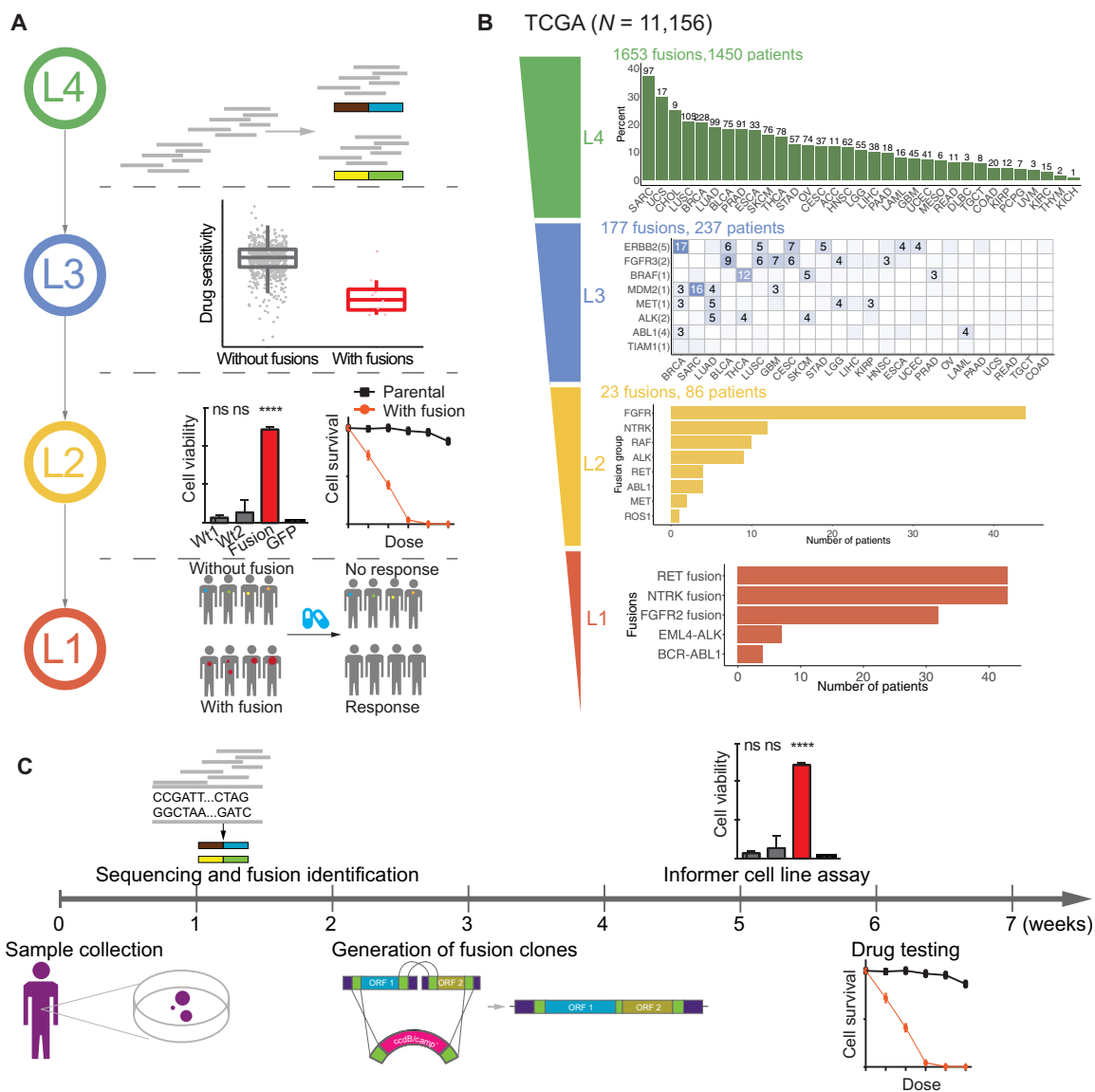


Fig. 7. An evidence-based classification system for prioritizing actionable gene fusions. (A) Overview of the four-level evidence-based classification system to actionable gene fusions for precision oncology. WT, wild type. (B) Summary of fusion classifications in the RTK-RAS pathway based on fusions identified from TCGA patient cohort. L2 fusions only include those that we assessed functionally, and L1 fusions include those with clinical evidence, not necessarily assessed in this study. **** indicates statistical significance. (C) Timeline of our fusion functional assessment pipeline in a clinical setting.

fusion candidates from RNA sequencing data of patient tumors to the completion of in vitro cell viability and drug sensitivity assays can be completed in as short as 7 weeks (Fig. 7C). Moreover, our pipeline is cost-effective, since our recombination engineering approach for gene fusion construction substantially reduces the cost compared to gene synthesis (table S2).

DISCUSSION

In this study, we developed a functional genomics approach to test the transforming potential and drug responsiveness of fusion genes and demonstrated the scalability of the approach, including the ability to perform analysis in a clinically relevant time scale. This approach has a high success rate (>80%) and is time efficient and

cost effective. The gene fusions characterized here provide a gold standard set for evaluating the computational algorithms to predict driver gene fusions in cancer. We found that a considerable fraction of low-frequency (or even singleton) gene fusions had functional effects and potential therapeutic relevance, which has important implications for patient management. Although fusion genes have represented therapeutic targets for decades (e.g., targeting of *BCR-ABL* markedly changed the treatment of chronic myelogenous leukemia), functional characterization and therapeutic efforts have mainly focused on highly recurrent fusion candidates. Our study reiterates the urgent need to consider low-frequency fusions as potential therapeutic targets to realize the full promise of personalized cancer therapy. Since our functional analysis approach can be completed within a relatively short time (e.g., 7 weeks), we propose that

Downloaded from https://www.science.org at University of Texas Md Anderson Cancer Center on February 09, 2022

this approach can be potentially applied in a clinical setting to identify functional fusion genes that likely drive tumor behavior and predict response to specific therapeutic regimens.

We recognize some limitations of our approach. First, we mainly use Ba/F3 as a model cell line to assess the functional effect of gene fusions. Although most known functional fusion genes (e.g., *FGFR3-TACC3* and *ETV6-NTRK3*) were validated, the Ba/F3 system may produce false negatives, as some fusions that are active in a lineage-specific system may not “score” in Ba/F3. For example, *KLK2-FGFR2* has been identified as a potential oncogenic event in prostate adenocarcinoma and uveal melanoma and has been reported to change the cell morphology and promote migration in the NIH 3T3 model (36). However, our parallel assessment using a second informer cell line, MCF10A, suggests that the false-negative rate of our platform is likely low. Second, we mainly used cell viability assays as a readout for phenotypic effects, which does not capture all the functional consequences of oncogenic drivers. However, in multiple systems, this phenotypic assay reports on the response to therapy in patients. As a complement to viability assays, we also explored other functional assays, such as clone formation assays. Third, our approach may identify some fusions as positive in the sensitive Ba/F3 informer cell line that may have more modest effects in a lineage-specific context. Our strategy of assessing drug sensitivity of activating fusions should mitigate this concern. In general, we designed the approach to provide a good balance between experimental efficiency and comprehensiveness of the characterization, with the goal of being able to provide timely support for clinical decisions. Further efforts are required to improve our approach to capture more “context-dependent” phenotypic effects.

We provide a comprehensive view of gene fusions from the RTK-RAS pathway across 33 cancer types. Our results show that, aside from high-frequency activating mutations in genes, such as V600E in *BRAF* and G12D in *KRAS*, gene fusions can also contribute, to a lesser but still compelling extent, to activation of the RTK-RAS pathway. Overall, while approximately one in every seven sequenced tumors carries a fusion gene of an RTK-RAS pathway member, there are remarkable differences in the frequency of fusion genes in the RTK-RAS pathway across different cancer types. This observation is in concordance with previous studies that show a predominance of activating RAS mutations in tumors such as melanoma (25), pancreatic, esophageal (37), and bladder cancer (38). Also, certain solid tumors seem to be more fusion prone compared to others, such as thyroid cancer, which displays a considerable mutational burden of RTK-RAS fusion genes despite being structurally “quiet” (39). Similarly, sarcomas appear to contain frequent fusion genes that are likely to be drivers (40). Cancer types such as melanoma and gastric adenocarcinoma harbor often hyper-rearranged genomic regions that include RTK-RAS pathway genes that have the potential to lead to functionally relevant fusion genes.

Although specific gene fusions are individually rare, they collectively made a substantial contribution to activation of the RTK-RAS pathway. There are a number of therapeutic approaches to inhibit the RTK-RAS pathway members that are available or under investigation. For example, our study identified two driver *MET* fusions (*BAIAP2L1-MET* and *TFG-MET*), and there is at least one ongoing clinical trial targeting *MET* fusion-positive tumors (NCT02978261). Thus, even without FDA-approved drugs, the oncogenic fusions identified by our functional genomics approach could be added to the actionable list of promising therapeutic targets. In addition to

clinically approved recurrent targets such as *BCR-ABL* in leukemia and *ALK* and *ROS* fusions in lung cancer, accumulating data support clinical responses to rare aberrations. For example, a case report presented two patients with metastatic melanoma with *BRAF* fusions that both had symptomatic improvement when treated with the MEK inhibitor trametinib (41). The positive clinical outcome associated with targeting rare fusions supports the consideration of a treatment paradigm, whereby patients are treated on the basis of the presence of such fusions independent of tissue or tumor origin. Functional information such as those provided here, or acquired therapeutically on the specific fusion, would greatly increase the confidence in treating patients based on a particular rare fusion gene. Therefore, an indispensable element of current patient stratification strategies would be the functional characterization for gene fusions, with potential emphasis on druggable targets. Toward that goal, our functional genomic approach and proposed four-level evidence-based classification system represent an initial but critical attempt to improve the clinical impact of targeting rare gene fusion events.

MATERIALS AND METHODS

Mice

Female athymic mice [CrI:NU(NCr)-*Foxn1*tm, 4 to 6 weeks in age; Charles River Laboratories] were used for xenograft tumor studies. In husbandry and housing conditions, all mice were fed a standard chow diet ad libitum and housed in a pathogen-free facility with standard controlled temperature, humidity, and light-dark cycle (12-hour) conditions with no more than five mice per cage under the supervision of veterinarians, in an Association for Assessment and Accreditation of Laboratory Animal Care International-accredited animal facility at Baylor College of Medicine. All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee at Baylor College of Medicine (AN-5428).

Cell lines

293T cells [American Type Culture Collection (ATCC)] were cultured in Dulbecco’s modified Eagle’s medium (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific). Parental Ba/F3 cells were cultured in RPMI 1640 medium (Thermo Fisher Scientific) supplemented with 5% FBS (Thermo Fisher Scientific) and recombinant mouse IL-3 (2.5 ng/ml; R&D Systems). MCF10A cells (ATCC) were cultured as described previously (42). HMEL cells (26) were cultured in RPMI 1640 medium supplemented with 10% FBS, penicillin (100 U/ml), and streptomycin (100 µg/ml). All cell lines were propagated at 37°C and 5% CO₂ in a humidified atmosphere. Cells lines were fingerprinted before use at the MD Anderson Cancer Center Characterized Cell Line Core Facility using a short tandem repeat (STR) testing platform.

Gene fusion selection

To balance the discovery power and assessment bias, we aimed to select 100 fusions from TCGA patients based on two factors, the potential druggability of fusion partner genes and the involvement of RTK-RAS pathway. In terms of the druggability annotation of partner genes, we set 50% for TClin, 35% for TChem, and 15% for TBio; for TClin, we further set 80% for RTK/RAS; and for TChem, we set 20% for RTK/RAS. Then, based on these predefined estimates, we chose the corresponding number of fusion candidates from each

category. In addition, we included a number of gene fusions detected in MD Anderson patient samples. In total, 110 gene fusions were selected.

Gene fusion construction

The fusion gene constructs were generated using the Gateway multifragment recombination technique, as previously described (12). Briefly, fusion gene construction was based on PCR using sequence-verified open reading frames (ORFs) corresponding to each fusion gene fragment obtained from the ORFeome collaboration (www.orfeomecollaboration.org), Mammalian Gene Collection, and commercial ORF sources (Ultimate ORF Clones; Life Technologies) as PCR template. PCR primers were generated to amplify the desired fragments (left and right gene arms of each fusion) with terminal Gateway (Life Technologies) recombination sequences. The left and right arm PCR products were incorporated into pDONR vectors (Life Technologies) and pFUSE-B vector (Addgene, plasmid no. 97185), respectively, through BP recombination (Life Technologies), followed by multifragment recombination into the pFUSE-DEST_R1R4 vector (Addgene, plasmid no. 97186) through LR recombination reaction (Life Technologies) following the manufacturer's recommendations. The reaction mixtures were incubated at room temperature overnight and subsequently transformed into STBL3 (Life Technologies) competent bacteria.

Ba/F3 cell viability assays

Lentivirus production and transduction of Ba/F3 cells are described in Ng *et al.* (17). Briefly, Lenti-X 293T cells were transfected with the fusion gene-containing vector and packaging plasmids (psPAX2 and pMD2.G). Lentivirus was harvested 3 days after transfection and used to transduce Ba/F3 cells by spinoculation at 1000g for 3 hours in the presence of polybrene (final concentration of 8 µg/ml). For Ba/F3 transforming potential assays, transduced Ba/F3 cells were incubated in medium without IL-3 for 7 days, as described earlier (15, 16), and cell viability was determined using CellTiter-Glo (Promega). For inhibitor assays, stable Ba/F3 cells expressing fusions (in IL-3-free medium) and parental Ba/F3 (in regular medium) were seeded in quadruplicates in 96-well plates at 1000 cells per well. Cells were treated with dimethyl sulfoxide or respective inhibitors at different concentrations for 72 hours, and cell viability was determined using CellTiter-Glo (Promega). At least two independent experiments were performed. All inhibitor compounds were purchased from Selleck Chemicals.

MCF10A anchorage-independent growth assays

Fusion genes were transduced into MCF10A cells via lentiviral infection. As previously described (12), soft agar assays were performed in six-well plates in triplicate. First, bottom layers were prepared using 0.8% Noble agar (Affymetrix Inc.) with complete MCF10A growth medium. After solidification, 10,000 cells were mixed with 0.45% agar in complete growth medium and laid on top of the bottom layer. Two milliliters of medium was added in each well after 3 days, and the medium was refreshed every 3 days. Colonies were counted 2 weeks after seeding.

Immunoblotting

Cells and tumor tissues were lysed with radioimmunoprecipitation assay buffer containing protease inhibitor cocktail (Sigma-Aldrich) and phosphatase inhibitor cocktail (Calbiochem). Protein lysates were

separated on 4 to 12% bis-tris gel (Novex) and transferred to polyvinylidene difluoride membranes. The following antibodies were used to detect expression: Raf1 (Abcam), BRAF (B-Raf proto-oncogene, serine/threonine kinase) (Santa Cruz Biotechnology), phospho-Erk1/2 (T202/Y204) (Cell Signaling Technology), Erk1/2 (Cell Signaling Technology), and glyceraldehyde-3-phosphate dehydrogenase (Santa Cruz Biotechnology).

In vivo xenograft tumor assay

All studies using mice were performed in accordance with our Institutional Animal Care and Use Committee-approved animal protocol (AN-5428) at Baylor College of Medicine. One million cells were re-suspended in a 1:1 Hanks' balanced salt solution (Life Technologies) and Matrigel (BD Biosciences) and injected into female athymic mice subcutaneously in bilateral flanks or orthotopically into the dermis (27). Mice were monitored twice a week, tumors were measured, and volumes were calculated as $\text{length} \times \text{width}^2/2$.

Bioinformatics analysis of TCGA gene fusions

The gene fusion events from >11,000 TCGA patients across 33 cancer types were collected on the basis of two previous bioinformatics studies (6, 8). The somatic mutations of genes in the RTK-RAS pathway identified from TCGA tumor samples were obtained from TCGA Pan-Cancer Analysis (19). Centrality scores were obtained from the TumorFusions database (8). The genes involved in the RTK-RAS pathway were obtained and merged on the basis of the RTK-RAS pathway defined by TCGA Pan-Cancer Analysis (19) and the RAS initiative (20). The fusion cancer bipartite network was built using Cytoscape (www.cytoscape.org) based on the fusion frequencies identified in TCGA patient cohort.

Drug sensitivity analysis on ALK and RAF1 gene fusions

The drug sensitivity data were downloaded from Genomics of Drug Sensitivity in Cancer (GDSC) (www.cancerrxgene.org) and Cancer Therapeutics Response Portal (CTRP) v2 (<https://portals.broadinstitute.org/ctrp/>). The gene fusions and the corresponding gene expression data in cell line samples were obtained from Cancer Cell Line Encyclopedia (<https://portals.broadinstitute.org/ccle>). The differential analyses of drug sensitivity (area under the curve) between cell lines with and without ALK or RAF1 fusions were performed by Wilcoxon test. The associations between ALK or RAF1 gene expression and ALK or RAF inhibitors were assessed by Pearson's correlation test.

An evidence-based classification system for prioritizing actionable gene fusions

Using TCGA data, we applied the classification system defined in Fig. 7A to prioritize L4 to L2 actionable gene fusions involved in the RTK-RAS pathway. L4 fusions were filtered from TCGA gene fusion list by selecting those having at least one partner gene in the RTK-RAS pathway. The number of patients harboring RTK-RAS fusions was counted within each cancer to generate the green bar plot in Fig. 7B. To annotate L3 fusions, we performed a differential drug sensitivity analysis for each pair of RTK-RAS gene and its corresponding inhibitors using drug sensitivity data obtained from GDSC and CTRPv2. The significant genes were selected under a false discovery rate of <0.2. The patients with fusions harboring these significant genes were further counted and summarized in the heatmap. The activating fusions annotated by our functional platform were annotated as L2 fusions. We further identified patients harboring

the L2 fusions from TCGA patient cohort and summarized the counts in a bar plot grouped by each gene family (e.g., *FGFR* and *NTRK*).

Data portal development

The functional annotation data, as well as other metadata, were first loaded into CouchDB. The fusion gene frequencies were obtained from two previous studies (6, 8). Literature data were manually curated. The fusion plots were developed on the basis of a command line tool for generating lollipop plots. The bar plots were generated by Highcharts. The interactive tables were generated by DataTables.

Quantification and statistical analysis

Statistical analysis was performed using R (version 3.6.1). To assess the correlation between two continuous variables, the Pearson's correlation test was used; to assess the correlation between two categorical variables, a chi-square test was used; and to compare the mean value between two groups, a Wilcoxon or Student's *t* test was used. Detailed descriptions of statistical tests are provided in Materials and Methods and the respective figure legends.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abm2382>

REFERENCES AND NOTES

- J. D. Rowley, Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
- S. R. Knezevich, D. E. McFadden, W. Tao, J. F. Lim, P. H. Sorensen, A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma. *Nat. Genet.* **18**, 184–187 (1998).
- S. A. Tomlins, D. R. Rhodes, S. Perner, S. M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J. E. Montie, R. B. Shah, K. J. Pienta, M. A. Rubin, A. M. Chinnaiyan, Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).
- B. C. Parker, W. Zhang, Fusion genes in solid tumors: An emerging target for cancer diagnosis and treatment. *Chin. J. Cancer* **32**, 594–603 (2013).
- C. Kumar-Sinha, S. A. Tomlins, A. M. Chinnaiyan, Recurrent gene fusions in prostate cancer. *Nat. Rev. Cancer* **8**, 497–511 (2008).
- Q. Gao, W.-W. Liang, S. M. Foltz, G. Mutharasu, R. G. Jayasinghe, S. Cao, W.-W. Liao, S. M. Reynolds, M. A. Wyczalkowski, L. Yao, L. Yu, S. Q. Sun; Fusion Analysis Working Group; Cancer Genome Atlas Research Network, K. Chen, A. J. Lazar, R. C. Fields, M. C. Wendt, B. A. Van Tine, R. Vij, F. Chen, M. Nykter, I. Shmulevich, L. Ding, Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* **23**, 227–238.e3 (2018).
- K. Yoshihara, Q. Wang, W. Torres-Garcia, S. Zheng, R. Vegesna, H. Kim, R. G. W. Verhaak, The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845–4854 (2015).
- X. Hu, Q. Wang, M. Tang, F. Barthel, S. Amin, K. Yoshihara, F. M. Lang, E. Martinez-Ledesma, S. H. Lee, S. Zheng, R. G. W. Verhaak, TumorFusions: An integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res.* **46**, D1144–D1149 (2018).
- PCAWG Transcriptome Core Group, C. Calabrese, N. R. Davidson, D. Demircioğlu, N. A. Fonseca, Y. He, A. Kahles, K.-V. Lehmann, F. Liu, Y. Shiraiishi, C. M. Soulette, L. Urban, L. Greger, S. Li, D. Liu, M. D. Perry, Q. Xiang, F. Zhang, J. Zhang, P. Bailey, S. Erkek, K. A. Hoadley, Y. Hou, M. R. Huska, H. Kilpinen, J. O. Korbel, M. G. Marin, J. Markovskii, T. Nandi, Q. Pan-Hammarström, C. S. Peadarallu, R. Siebert, S. G. Stark, H. Su, P. Tan, S. M. Waszak, C. Yung, S. Zhu, P. Awadalla, C. J. Creighton, M. Meyerson, B. F. Francis Ouellette, K. Wu, H. Yang; PCAWG Transcriptome Working Group, A. Brazma, A. N. Brooks, J. Göke, G. Rättsch, R. F. Schwarz, O. Stegle, Z. Zhang; PCAWG Consortium, Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136 (2020).
- H. Chen, J. Li, Y. Wang, P. K.-S. Ng, Y. H. Tsang, K. R. Shaw, G. B. Mills, H. Liang, Comprehensive assessment of computational algorithms in predicting cancer driver mutations. *Genome Biol.* **21**, 43 (2020).
- M. H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M. C. Wendt, J. Kim, B. Reardon, P. K.-S. Ng, K. J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E. M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortés-Ciriano, D. C. Zhou, W.-W. Liang, J. M. Hess, V. D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavitai, J. Y. Ko, E. Khurana, P. J. Park, E. M. Van Allen, H. Liang; MC3 Working Group; Cancer Genome Atlas Research Network, M. S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A. J. Lazar, G. B. Mills, R. Karchin, L. Ding, Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e18 (2018).
- H. Lu, N. Villafane, T. Dogruluk, C. L. Grzeskowiak, K. Kong, Y. H. Tsang, O. Zagorodna, A. Pantazi, L. Yang, N. J. Neill, Y. W. Kim, C. J. Creighton, R. G. Verhaak, G. B. Mills, P. J. Park, R. Kucherlapati, K. L. Scott, Engineering and functional characterization of fusion genes identifies novel oncogenic drivers of cancer. *Cancer Res.* **77**, 3502–3512 (2017).
- D. T. Nguyen, S. Mathias, C. Bologna, S. Brunak, N. Fernandez, A. Gaulton, A. Hersey, J. Holmes, L. J. Jensen, A. Karlsson, G. Liu, A. Ma'ayan, G. Mandava, S. Mani, S. Mehta, J. Overington, J. Patel, A. D. Rouillard, S. Schürer, T. Sheils, A. Simeonov, L. A. Sklar, N. Southall, O. Ursu, D. Vidovic, A. Waller, J. Yang, A. Jadhav, T. I. Oprea, R. Guha, Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* **45**, D995–D1002 (2017).
- M. Warmuth, S. Kim, X. J. Gu, G. Xia, F. Adrian, Ba/F3 cells and their use in kinase drug discovery. *Curr. Opin. Oncol.* **19**, 55–60 (2007).
- L. W. T. Cheung, S. Yu, D. Zhang, J. Li, P. K. S. Ng, N. Panupinthu, S. Mitra, Z. Ju, Q. Yu, H. Liang, D. H. Hawke, Y. Lu, R. R. Broaddus, G. B. Mills, Naturally occurring neomorphic PIK3R1 mutations activate the MAPK pathway, dictating therapeutic response to MAPK pathway inhibitors. *Cancer Cell* **26**, 479–494 (2014).
- H. Liang, L. W. T. Cheung, J. Li, Z. Ju, S. Yu, K. Stemke-Hale, T. Dogruluk, Y. Lu, X. Liu, C. Gu, W. Guo, S. E. Scherer, H. Carter, S. N. Westin, M. D. Dyer, R. G. W. Verhaak, F. Zhang, R. Karchin, C.-G. Liu, K. H. Lu, R. R. Broaddus, K. L. Scott, B. T. Hennessy, G. B. Mills, Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer. *Genome Res.* **22**, 2120–2129 (2012).
- P. K.-S. Ng, J. Li, K. J. Jeong, S. Shao, H. Chen, Y. H. Tsang, S. Sengupta, Z. Wang, V. H. Bhavana, R. Tran, S. Soewito, D. C. Minussi, D. Moreno, K. Kong, T. Dogruluk, H. Lu, J. Gao, C. Tokheim, D. C. Zhou, A. M. Johnson, J. Zeng, C. K. M. Ip, Z. Ju, M. Wester, S. Yu, Y. Li, C. P. Vellano, N. Schultz, R. Karchin, L. Ding, Y. Lu, L. W. T. Cheung, K. Chen, K. R. Shaw, F. Meric-Bernstam, K. L. Scott, S. Yi, N. Sahni, H. Liang, G. B. Mills, Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* **33**, 450–462.e10 (2018).
- C. C. Wu, K. Kannan, S. Lin, L. Yen, A. Milosavljevic, Identification of cancer fusion drivers using network fusion centrality. *Bioinformatics* **29**, 1174–1181 (2013).
- F. Sanchez-Vega, M. Mina, J. Armenia, W. K. Chatila, A. Luna, K. C. La, S. Dimitriadou, D. L. Liu, H. S. Kantheti, S. Saghaforina, D. Chakravarty, F. Daian, Q. Gao, M. H. Bailey, W.-W. Liang, S. M. Foltz, I. Shmulevich, L. Ding, Z. Heins, A. Ochoa, B. Gross, J. Gao, H. Zhang, R. Kundra, C. Kandath, I. Bahceci, L. Dervishi, U. Dogrusoz, W. Zhou, H. Shen, P. W. Laird, G. P. Way, C. S. Greene, H. Liang, Y. Xiao, C. Wang, A. Iavarone, A. H. Berger, T. G. Bivona, A. J. Lazar, G. D. Hammer, T. Giordano, L. N. Kwong, G. M. Arthur, C. Huang, A. D. Tward, M. J. Frederick, F. M. Cormick, M. Meyerson; Cancer Genome Atlas Research Network, E. M. Van Allen, A. D. Cherniack, G. Ciriello, C. Sander, N. Schultz, Oncogenic signaling pathways in the Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
- A. G. Stephen, D. Esposito, R. K. Bagni, F. McCormick, Dragging ras back in the ring. *Cancer Cell* **25**, 272–281 (2014).
- H. Mano, ALKoma: A cancer subtype with a shared target. *Cancer Discov.* **2**, 495–502 (2012).
- M. Soda, Y. L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S.-i. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, H. Mano, Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
- A. T. Shaw, D.-W. Kim, K. Nakagawa, T. Seto, L. Crinó, M.-J. Ahn, T. de Pas, B. Besse, B. J. Solomon, F. Blackhall, Y.-L. Wu, M. Thomas, K. J. O'Byrne, D. Moro-Sibilot, D. R. Camidge, T. Mok, V. Hirsh, G. J. Riely, S. Iyer, V. Tassell, A. Polli, K. D. Wilner, P. A. Jänne, Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N. Engl. J. Med.* **368**, 2385–2394 (2013).
- N. Stransky, E. Cerami, S. Schalm, J. L. Kim, C. Lengauer, The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).
- Cancer Genome Atlas Research Network, Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
- K. L. Scott, C. Nogueira, T. P. Heffernan, R. van Doorn, S. Dhakal, J. A. Hanna, C. Min, M. Jaskieloff, Y. Xiao, C.-J. Wu, L. A. Cameron, S. R. Perry, R. Zeid, T. Feinberg, M. Kim, G. Vande Woude, S. R. Granter, M. Bosenberg, G. C. Chu, R. A. DePinho, D. L. Rimm, L. Chin, Proinvasion metastasis drivers in early-stage melanoma are oncogenes. *Cancer Cell* **20**, 92–103 (2011).
- J. Wardwell-Ozgo, T. Dogruluk, A. Gifford, Y. Zhang, T. P. Heffernan, R. van Doorn, C. J. Creighton, L. Chin, K. L. Scott, HOXA1 drives melanoma tumor growth and metastasis and elicits an invasion gene expression signature that prognosticates clinical outcome. *Oncogene* **33**, 1017–1026 (2014).
- Y. M. Wu, F. Su, S. Kalyana-Sundaram, N. Khazanov, B. Ateeq, X. Cao, R. J. Lonigro, P. Vats, R. Wang, S.-F. Lin, A.-J. Cheng, L. P. Kunju, J. Siddiqui, S. A. Tomlins, P. Wyngaard, S. Sadis, S. Roychowdhury, M. H. Hussain, F. Y. Feng, M. M. Zalupski, M. Talpaz, K. J. Pienta, D. R. Rhodes, D. R. Robinson, A. M. Chinnaiyan, Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov.* **3**, 636–647 (2013).

29. Y. Arai, Y. Totoki, F. Hosoda, T. Shiota, N. Hama, H. Nakamura, H. Ojima, K. Furuta, K. Shimada, T. Okusaka, T. Kosuge, T. Shibata, Fibroblast growth factor receptor 2 tyrosine kinase fusions define a unique molecular subtype of cholangiocarcinoma. *Hepatology* **59**, 1427–1434 (2014).
30. J. D. Karkera, G. M. Cardona, K. Bell, D. Gaffney, J. C. Portale, A. Santiago-Walker, C. H. Moy, P. King, M. Sharp, R. Bahleda, F. R. Luo, J. D. Alvarez, M. V. Lorenzi, S. J. Platero, Oncogenic characterization and pharmacologic sensitivity of activating fibroblast growth factor receptor (FGFR) genetic alterations to the selective FGFR inhibitor erdafitinib. *Mol. Cancer Ther.* **16**, 1717–1726 (2017).
31. R. Costa, B. A. Carneiro, T. Taxter, F. A. Tavora, A. Kalyan, S. A. Pai, Y. K. Chae, F. J. Giles, FGFR3-TACC3 fusion in solid tumors: Mini review. *Oncotarget* **7**, 55924–55938 (2016).
32. B. C. Parker, M. J. Annala, D. E. Cogdell, K. J. Granberg, Y. Sun, P. Ji, X. Li, J. Gumin, H. Zheng, L. Hu, O. Yli-Harja, H. Haapasalo, T. Visakorpi, X. Liu, C.-G. Liu, R. Sawaya, G. N. Fuller, K. Chen, F. F. Lang, M. Nykter, W. Zhang, The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. *J. Clin. Invest.* **123**, 855–865 (2013).
33. S. J. Isakoff, J. A. Engelman, H. Y. Irie, J. Luo, S. M. Brachmann, R. V. Pearline, L. C. Cantley, J. S. Brugge, Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Res.* **65**, 10992–11000 (2005).
34. M. Katoh, H. Nakagama, TGF receptors: Cancer biology and therapeutics. *Med. Res. Rev.* **34**, 280–300 (2014).
35. D. Singh, J. M. Chan, P. Zoppoli, F. Niola, R. Sullivan, A. Castano, E. M. Liu, J. Reichel, P. Poratti, S. Pellegatta, K. Qiu, Z. Gao, M. Ceccarelli, R. Riccardi, D. J. Brat, A. Guha, K. Aldape, J. G. Golfinos, D. Zagzag, T. Mikkelsen, G. Finocchiaro, A. Lasorella, R. Rabadan, A. Iavarone, Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**, 1231–1235 (2012).
36. M. A. Krook, H. Barker, H.-Z. Chen, J. W. Reeser, M. R. Wing, D. Martin, A. M. Smith, T. Dao, R. Bonneville, E. Samorodnitsky, J. Miya, A. G. Freud, J. P. Monk, S. K. Clinton, S. Roychowdhury, Characterization of a KLK2-FGFR2 fusion gene in two cases of metastatic prostate cancer. *Prostate Cancer Prostatic Dis.* **22**, 624–632 (2019).
37. A. M. Dulak, P. Stojanov, S. Peng, M. S. Lawrence, C. Fox, C. Stewart, S. Bandla, Y. Imamura, S. E. Schumacher, E. Shefler, A. McKenna, S. L. Carter, K. Cibulskis, A. Sivachenko, G. Saksena, D. Voet, A. H. Ramos, D. Auclair, K. Thompson, C. Sougnez, R. C. Onofrio, C. Guiducci, R. Beroukhi, Z. Zhou, L. Lin, J. Lin, R. Reddy, A. Chang, R. Landrenau, A. Pennathur, S. Ogino, J. D. Luketich, T. R. Golub, S. B. Gabriel, E. S. Lander, D. G. Beer, T. E. Godfrey, G. Getz, A. J. Bass, Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45**, 478–486 (2013).
38. Cancer Genome Atlas Research Network, Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
39. Cancer Genome Atlas Research Network, Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
40. Cancer Genome Atlas Research Network, Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950–965.e28 (2017).
41. A. M. Menzies, I. Yeh, T. Botton, B. C. Bastian, R. A. Scolyer, G. V. Long, Clinical activity of the MEK inhibitor trametinib in metastatic melanoma containing BRAF kinase fusion. *Pigment Cell Melanoma Res.* **28**, 607–610 (2015).
42. J. Debnath, S. K. Muthuswamy, J. S. Brugge, Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**, 256–268 (2003).

Acknowledgments: This study is dedicated to the late K.L.S. **Funding:** This study was supported by the U.S. NIH (CA217842, CA217685, and CA098258 to G.B.M.; CA209851 and CA253472 to H.Li. and G.B.M.; CA144025 to R.K.; CA168394 to K.L.S.; CA016672 to H.Li.; and CA125123 to C.J.C.) and by the Cancer Prevention and Research Institute of Texas (New Investigator grant RR160021 to N.S., RP150535 to F.M.-B., RP120046 to K.L.S. and RP140102 to H. Lu). This study was supported by the National Cancer Institute's Office of Cancer Genomics Cancer Target Discovery and Development (CTD²) initiative. The results published here are based, in whole or in part, on data generated by CTD² Network (<https://ocg.cancer.gov/programs/ctd2/data-portal>) established by the National Cancer Institute's Office of Cancer Genomics. **Author contributions:** K.L.S., R.K., G.B.M., and H.Li. conceived and designed the study. J.L. and A.Pa. led the data analysis. H. Lu and P.K.-S.N. led the experiments. C.K.M.I., K.J.J., B.A., R.T., Y.H.T., L.Y., X.S., T.D., X.R., A.H., C.A.B., S.L., M.K., M.P., J.T., S.S., H.S.M., K.M., D.Z., J.Z., A.Pr., J.G.S., C.J.C., Y.L., N.S., K.R.S., F.M.-B., A.F., L.C., K.L.S., R.K., G.B.M., and H.Li. contributed to experiments and analysis. L.C., K.L.S., R.K., G.B.M., and H.Li. contributed to project management. J.L., H. Lu, P.K.-S.N., A.Pa., R.K., G.B.M., and H.Li. wrote the manuscript with input from the other authors. **Competing interests:** M.P. is a full-time employee of Invitae Corporation. F.M.-B. serves as a consultant to AbbVie, Aduro Biotech Inc., Alkermes, AstraZeneca, Debiopharm, eFFECTOR Therapeutics, F. Hoffmann-La Roche Ltd., Genentech Inc., IBM Watson, Infinity Pharmaceuticals, the Jackson Laboratory, Kolon Life Science, Lengo Therapeutics, Origimed, PACT Pharma, Parexel International, Pfizer Inc., Samsung Bioepis, Seattle Genetics Inc., Tallac Therapeutics, Tyra Biosciences, Xencor, and Zymeworks; is an advisor to Black Diamond, Biovia, Eisai, Immunomedics, Inflection Biosciences, Karyopharm Therapeutics, Loxo Oncology, Mersana Therapeutics, OnCusp Therapeutics, Puma Biotechnology Inc., Seattle Genetics, Silverback Therapeutics, Spectrum Pharmaceuticals, and Zentalis Pharmaceuticals; received grants from Aileron Therapeutics Inc. AstraZeneca, Bayer HealthCare Pharmaceuticals, Calithera Biosciences Inc., Curis Inc., CytomX Therapeutics Inc., Daiichi Sankyo Co. Ltd., Debiopharm International, eFFECTOR Therapeutics, Genentech Inc., Guardant Health Inc., KLUS Pharma, Takeda Pharmaceutical, Novartis, Puma Biotechnology Inc., and Taiho Pharmaceutical Co.; and received honoraria and/or travel-related fees from Chugai Pharmaceutical, Mayo Clinic, and Rutgers Cancer Institute of New Jersey. L.C. is a co-founder and shareholder of Apricity Health LLC. R.K. owns stock in PureTech Health and serves a Board Director for PureTech Health, Gelesis Inc., and KEW Inc. G.B.M. is a Scientific Advisory Board member or consultant for AstraZeneca, Chrysalis Biotechnology, GSK, ImmunoMet, Ionis Pharmaceuticals, Eli Lilly and Company, PDX Pharmaceuticals, Signalchem Lifesciences, Symphogen, Tarveda Therapeutics, Turbine, and Zentalis Pharmaceuticals; has stock/options or financial considerations with Catena Pharmaceuticals, ImmunoMet, Signalchem, and Tarveda Therapeutics; and has licensed HRD assay to Myriad Genetics and DSP patents to NanoString Technologies. H.Li. is a shareholder and advisor for Precision Scientific. The authors declare that they have no other competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Functional data generated in this study are available at the FASMIC data portal (<https://bioinformatics.mdanderson.org/public-software/fasmic/>).

Submitted 3 September 2021

Accepted 16 December 2021

Published 9 February 2022

10.1126/sciadv.abm2382

A functional genomic approach to actionable gene fusions for precision oncology

Jun LiHengyu LuPatrick Kwok-Shing NgAngeliki PantaziCarman Ka Man IpKang Jin JeongBianca AmadorRichard TranYiu Huen TsangLixing YangXingzhi SongTurgut DogrulukXiaoja RenAngela HadjipanayisChristopher A. BristowSemin LeeMelanie KucherlapatiMichael ParfenovJiabin TangSahil SethHarshad S. MahadeshwarKamalika MojumdarDong ZengJianhua ZhangAlexei ProtopopovJonathan G. SeidmanChad J. CreightonYiling LuNidhi SahniKenna R. ShawFunda Meric-BernstamAndrew FutrealLynda ChinKenneth L. ScottRaju KucherlapatiGordon B. MillsHan Liang

Sci. Adv., 8 (6), eabm2382.

View the article online

<https://www.science.org/doi/10.1126/sciadv.abm2382>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of think article is subject to the [Terms of service](#)