A Map of Genetic Interactions in Cancer Cells

Benedikt Rauscher[1], Florian Heigwer[1], Luisa Henkel[1], Thomas Hielscher[2]
and Michael Boutros[1*]

[1]Division of Signaling and Functional Genomics, German Cancer Research Center (DKFZ) and Heidelberg University, 69120 Heidelberg, Germany
[2]Division of Biostatistics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

* Author for correspondence:
Email: m.boutros@dkfz.de
Phone: +49 6221 421950
Fax: +49 6221 421959

ABSTRACT

Cancer genomes often harbor hundreds of molecular aberrations. Such genetic variants can be drivers or passengers of tumorigenesis and, as a side effect, create new vulnerabilities for potential therapeutic exploitation. To systematically identify genotype-dependent vulnerabilities and synthetic lethal interactions, forward genetic screens in different genetic backgrounds have been conducted. Here, we extended genetic interaction network analysis approaches that were established for model organisms to integrate and analyze data from 83 CRISPR/Cas9 screens in human cancer cell lines. We integrated functional data with information on loss-of-function variants to explore the relationships of more than 2.2 million gene-gene combinations. In addition to known gene-gene dependencies, our analysis identified new genotype-specific vulnerabilities of cancer cells. By clustering genes with similar genetic interaction profiles, we drew the largest map of genetic interactions in a cancer cell to date. This approach is scalable and highlights how diverse genetic screens can be integrated to comprehensively build maps of genetic interactions in cancer.

Keywords: Genetic interactions; networks; epistasis; synthetic lethality; cancer

INTRODUCTION

Genes rarely function in isolation to affect phenotypes at the cellular or organismal level. Many studies have described how genes act in complex networks to maintain homeostasis by fine-tuning cellular or organismal reactions to internal or external stimuli (Bergman & Siegal 2003). A loss of genetic buffering can result in the emergence of diseases, such as cancer, in humans (Hartman et al. 2001; Hartwell et al. 1997). Mutations can create genetic vulnerabilities in cancer cells, for example, by deactivating one of two genetically buffered pathways (Nagel et al. 2016; Torti & Trusolino 2011; Luo et al. 2009). Therapeutic approaches attempt to exploit such events by selectively inducing cell death in cancer cells while causing little harm to normal cells (Kaelin 2005; Nijman 2011)

To systematically identify genetic interactions, pairwise gene knockouts or knockdowns experiments can be performed (Mani et al. 2008) In cases where a measured phenotype of the double knockout is stronger than expected based on the the two single knockout phenotypes, it is called aggravating or synthetic lethal interaction (Bridges 1922). In contrast, a buffering (or alleviating) interaction is observed when the measured double phenotype is weaker than expected. Arrayed screens, performed by mating of loss-of-function mutant yeast strains have pioneered combinatorial gene screening (Tong et al. 2001; Davierwala et al. 2005; Baryshnikova et al. 2010; Costanzo et al. 2010) Methods of pairwise gene perturbation were later extended using combinatorial RNAi to map genetic interactions in cultured metazoan cells (Horn et al. 2011; Laufer et al. 2013; Fischer et al. 2015; Martin et al. 2015; Byrne et al. 2007; Snijder et al. 2013; Srivas et al. 2016) However, the screening of all pairwise gene combinations scales poorly with increasing genome size.

Genome-scale perturbation screens can now be efficiently performed in many cell lines using single-guide (sg)RNA libraries, whereby the CRISPR/Cas9 is used for the targeted inactivation of genes (Doudna & Charpentier 2014; Barrangou 2014; Shalem et al. 2014; Wang et al. 2014; Horlbeck et al. 2016). Since each cell line has a different genetic background, this enables the investigation of genotype-specific

vulnerabilities (Hart et al. 2015; Steinhart et al. 2017; Wang et al. 2017; Tzelepis et al. 2016; Garnett et al. 2012; Iorio et al. 2016), thus providing an alternative approach to comprehensively map gene-gene interactions.

Here, we present an integrated analysis based on a curated and reanalyzed data set of 83 genome-scale CRISPR/Cas9 screens in 55 different human cancer cell lines to score gene-gene combinations for genetic interactions. Filtering all annotated mutations for loss-of-function mutations enabled us to create a global profile of loss-of-function perturbations in each cancer cell line. For each cell line included in our dataset, we combined the intrinsic gene loss-of-function profile with gene-level fitness scores derived from single-perturbation CRISPR-screens. By applying a linear mixed-effects model, we tested 2.2 million pairwise gene combinations by comparing wild-type against loss-of-function alleles in cell lines. From genetic interaction profiles, we generated a map of genetic interactions in cancer cells by connecting genes with similar profiles and identified network modules with similar functional characteristics.

RESULTS

Integration of somatic variation and copy number data creates loss-of-function profiles of cancer cells

We devised a workflow to create an aggregated dataset containing relative fitness scores for each gene perturbation across all genes that were analyzed (Figure 1A-D). First, we assembled cell line-specific information about gene copy-numbers and somatic mutations from the Cancer Cell Line Encyclopedia (CCLE; Barretina et al. 2012) and COSMIC (Forbes et al. 2017) Loss of both genetic alleles and disruptive mutations such as frameshift or nonsense mutations on at least one allele were considered as loss-of-function events. To correct for sequencing errors and variants that are exclusive to the cell line model, loss-of-function events were compared to variants observed in primary tumors, as reported in COSMIC, and to mutations listed in dbSNP (Sherry et al. 2001) Overall, our analysis revealed 54,064 loss-of-function events across 1,412 cancer cell lines (Supplementary Table 1). We identified a variety

of tumor suppressor genes as being commonly affected by loss-of-function events, with *CDKN2A* as the most frequently mutated gene (Figure S2C) (p16, Liggett & Sidransky 1998) p16, which is encoded by *CDKN2A,* has recently been reported to directly interact with p53 and to enhance its transcriptional and apoptotic functions (Al-Khalaf et al. 2017). Similarly, loss of *CSMD1,* which ranked second in our analysis of loss-of-function events in cancer cell lines, has been linked to the development of a variety of cancers (Escudero-Esparza et al. 2016; Zhu et al. 2016; Tang et al. 2012; Zhang & Song 2014)

We further compared the frequency of loss-of-function events across tissues. In our dataset, most gene deletions can be found in colorectal, prostate and endometrial cancer cell lines (Figure 2B). On average, we identified 38 loss-of-function events per cell line, while few cell lines harbored more than 400 loss-of-function mutations (Figure 2C). Next, we examined whether cell lines could be clustered by tissue types on their loss-of-function genotypes. For example, our analysis allowed for colorectal cell lines to be distinguished from pancreatic and skin cancer cell lines based on the presence of loss-of-function mutations in nine characteristic genes. Pancreatic and skin cancer cell lines commonly share loss-of-function of *CDKN2A, MTAP, DKN2B* and *DMRTA1* loss-of-function*,* all of which are located on region p21.3 of chromosome 9, indicating a mutual mutation or deletion event. On the other hand, *APC* loss-of-function is characteristic of colorectal cancer cell lines (Figure 2D).

Integrated analysis of CRISPR-Cas9 screens identifies genotype-specific vulnerabilities

To examine whether the genotypic makeup of a cell can influence the effect of CRISPR/Cas9-mediated gene perturbation, we analyzed all viability screens that were available in the GenomeCRISPR database (Supplementary Table 2, Figure S2A). These screens are pooled loss-of-function screens using lentiviral single-guide RNA (sgRNA) libraries. Using BAGEL, the degree of essentiality of a gene can be quantified by deriving a fitness score (Hart & Moffat 2016). We individually performed BAGEL for each screen individually (Figure S2C).

To assess the quality of each screen we created precision-recall curves to assess how well gold-standard sets of core-essential and non-essential genes could be identified. Based on the results, we excluded seven screens from downstream analysis (Figure S2B). We then integrated the results by performing quantile normalization (Figure S2D-E). Clustering analysis revealed strong batch effects, which correlate with the fact that the analyzed screens were derived from different laboratories using different protocols and sgRNA libraries (Figure S2F). Therefore, we performed a batch-correction using ComBat (Leek et al. 2012) to remove confounding effects from the data (Figure S2G; Supplementary Table 3; Figure 3A). After batch-correction we observed that screens in similar cell lines and tissues clustered together (Figure S2G). We robustly observed low fitness scores for known core-essential genes (Figure 3B) and high fitness scores for non-essential genes (Figure 3C).

To investigate how the response to CRISPR/Cas9-mediated gene disruption can vary across tissue, we performed one-way ANOVA analysis to identify genes that display significantly altered fitness scores in different tissues. As an example, we found that colorectal cancer cells are more sensitive to perturbation of β-catenin compared to cell lines that originated from other tissues (*CTNNB1*; Figure 3D). Especially in colorectal cancer, Wnt/β-catenin signaling has been frequently shown to be aberrantly upregulated (Zhan et al. 2017). Furthermore, we identified a strong dependency of hematopoietic and lymphoid cancer cells on *CBFB*, *CCND3* and *MYB*, all of which have been identified as essential regulators of hematopoiesis (C. Q. Wang et al. 2015; Cooper et al. 2006; Greig et al. 2008). Moreover, we analyzed genes frequently mutated in various types of cancer. As a prominent example, *MYC* is dysregulated in more than half of all human cancers (Meyer & Penn 2008) and is a well-known driver of cancer development (Little et al. 1983) From our dataset, we selected all cancer cell lines harboring *MYC* mutations. Based on these data, we performed a gene set enrichment test to identify pathways that either confer a fitness advantage or disadvantage when perturbed by loss-of-function mutations using CRISPR. Interestingly, we observed that the knockout of several genes that are either part of the SWI/SNF chromatin remodeling complex or EGFR signaling significantly reduces the

fitness of *MYC*-mutated cancer cell lines. Vice versa, loss-of-function of genes implicated in *FGFR3* signaling or the Unfolded Protein Response (UPR) significantly enhances the cellular fitness of the aforementioned cell lines (Fig. 3D). Interactions and the mutual regulation of gene expression between several parts of the SWI/SNF chromatin remodeling complex and *MYC* have been previously described (Nagl et al., 2006 16452181, Dingar et al., 2015 25452129 (Stojanova et al. 2016). We speculate that mutations within the SWI/SNF complex could perturb the transcriptional regulation of target genes that are mutually regulated by SWI/SNF and *MYC*, thus resulting in reduced cellular fitness.

Furthermore, it has been shown that increased *MYC* function can lead to an increased cellular dependency on glucose and glutamine in cancer cells. Glutamine uptake in glucose-deprived cells triggers UPR, which results in the inhibition of cell growth by JNK activation in most cases (Shajahan-Haq et al. 2014). This could explain the observed increase in mean fitness upon disruption of UPR in *MYC* mutated cell lines. Interestingly, there are observed different cellular responses upon either targeting components of *EGFR* or *FGFR3* signaling. Recently, it has been reported that in $Kras^{G12D}$-driven cancers, *MYC* expression, phosphorylation and *MYC* target gene expression are induced in an EGFR-dependent manner and that *EGFR* inhibition prevents $Kras^{G12D}$-induced Myc protein expression (Diersch et al. 2016). In *FGFR*-mutated cancers, c-*MYC* has been described as a key downstream effector that precedes FGFR-MEK/ERK signaling, and ectopic expression of non-degradable c-Myc conferred resistance to *FGFR* inhibition (Liu et al. 2017).

We next hypothesized that genes that play a role in similar biological pathways could be identified based on similar response to perturbation following specific driver mutations in cancer cells. Again, we selected three genes (*NRAS, PIK3CA and FGFR3*) that are frequently mutated across multiple cancer types. Using a one-way ANOVA, we identified genes with significantly altered fitness scores based on the mutation status of the selected cancer drivers. We observed, that NRAS-mutated cells, along with *NRAS* itself, are dependent on *TFAP4, DNAJC9* and *STK11* (Figure 3F). These findings suggest that these genes might contribute to NRAS-mediated carcinogenesis.

Furthermore, we observed that cells with mutations in *FGFR3* or *PIK3CA* share many genetic dependencies (e.g. *KIF18A*, *NDE1*, Figure 3F). Simultaneous mutations in *PIK3CA* and *FGFR3* have commonly been described in bladder tumors (Kompier et al. 2010; López-Knowles et al. 2006) which might indicate common dependencies that arise upon perturbation of either pathway.

A linear mixed-effects model identifies synthetic interactions

To systematically identify synthetic genetic dependencies, we formed all pairwise combinations between genes that were perturbed in screening experiments (target genes) and genes that are frequently lost in cancer cell lines (query genes). We then used a linear mixed-effects model (LME) to test each of these combinations for differential fitness effects between the mutated and the wild-type cell lines. An advantage of using an LME in the context of our data is that it allows for the consideration of potential batch effects in the form of random effect modeling. Among the ~2.2 million gene-gene-combinations that were tested, we found 2,365 significant interactions at a 20% false discovery rate (FDR; Figure 4A; Supplementary Table 4). Globally, we observed an increase in fitness sensitivity more frequently than an increase in resistance.

By this approach, we found that *SMAD4*-depleted cells showed a higher sensitivity to *KLF5* perturbation (Figure 4B). It has been previously reported that *KLF5* can prevent *SOX4*-induced apoptosis in *SMAD4*-deficient cells (David et al. 2016), suggesting *KLF5* as potentially interesting target for drug therapy in *SMAD4*-negative cancers. Similarly, we observed an increased sensitivity to *ERBB2* mutation depending on the mutation status of *SMAD4*, which we could also corroborate with drug sensitivity data from Iorio et al. (Fig. 4C, Iorio et al. 2016). Deleterious *SMAD4* mutations were previously shown to elevate levels of *ERBB2* in lung and pancreatic tumors and thereby create a dependency on *ERBB2* that can be exploited using drug treatment (Liu et al. 2015; Zhao et al. 2010). Extending this approach to other gene pairs, we also identified an induced vulnerability of *BCOR* mutant cancer cell lines to *MET* mutation or inhibition by Crizotinib (Fig. 4D). We hypothesized that the loss of *BCOR* sensitizes cells to

apoptosis by up-regulating *BCL6,* thereby creating a vulnerability to anti-proliferative treatment (Grossmann et al. 2011; Huynh et al. 2000; Tortora et al. 2003).

Finally, we noticed that many cell lines with disruptions of the tumor suppressor gene *FHIT* (Waters et al. 2014) showed high sensitivity to perturbation of β-catenin (*CTNNB1*; Figure 4D). Previous studies have shown that *FHIT* can repress the transcriptional activity of β-catenin (Weiske et al. 2007) suggesting that loss of *FHIT* could mediate Wnt signaling-driven oncogenesis. However, within the panel of *FHIT*-mutant cell lines, we observed a cell line-specific dependency on the function of *CTNNB1*, indicating that additional genetic alterations might explain the differences that were observed. Therefore, we searched for mutation patterns that separated separating the β-catenin sensitive from the resistant *FHIT*-deficient cell lines. We observed that all resistant cell lines share a 3-base-pair deletion in the tumor suppressor gene *GUCY2C* (Lin et al. 2010), whereas it remains unaltered in all sensitive cell lines. Other genes with different mutation patterns were *CAMTA1*, *C10orf11*, *CWF19L2*, and *PRSS2* (loss-of-function in *CTNNB1*-LoF-sensitive cell lines) as well as *TMEM123*, *CSMD1*, *NPR1*, *CLTCL1* and *EPHA6* (loss-of-function in *CTNNB1*-LoF-resistant cell lines). Taken together, these mutations could provide a differential genetic background, which could explain why some cell lines exhibit *CTNNB1* interactions with *FHIT* while others do not.

Π-scores quantify genetic interaction strength

To quantify genetic interactions, we have previously described the π-score to describe the difference between the observed fitness score and a prediction based on a non-interacting model (Laufer et al. 2013; Horn et al. 2011) We computed π-scores for all pairwise combinations of perturbed genes and loss-of-function mutations (Figure 5A; Supplementary Table 5). We observed a tendency towards positive interactions among genes with low interaction counts, while genes with intermediate interaction counts displayed more negative (synthetic lethal) interactions at an absolute π-score cutoff of 0.3 (Figure 5B). These findings agree with results previously reported in yeast (Costanzo et al. 2010). We examined interaction scores of synthetic interactions

described above (Figure 5C-E). For all three interactions, the measured fitness strongly differed from the median fitness scores that were individually observed for the target and query across all data points. We noticed, that the query main effect is very small throughout all query genes. This could be explained by the fact, that cancer cells are able to acquire loss-of-function mutations in these genes without suffering selective disadvantage. For our analysis, this implies that the difference between target main effect and the measured effect is the main determinant of genetic interaction strength.

Previous studies have shown that genes sharing common functions can be clustered based on the similarity of their interaction profiles (Costanzo et al. 2010; Baryshnikova et al. 2010; Horn et al. 2011; Laufer et al. 2013; Fischer et al. 2015; Costanzo et al. 2016) To investigate whether such a clustering could be achieved based on interaction scores derived from CRISPR/Cas9 mediated knockout screens, we selected members from the NADH-ubiquinone oxidoreductase and mediator complexes and clustered them by the similarity of their interaction profiles with query gene/loss-of-function mutations. Members of both complexes could be separated, forming distinct clusters based on Ward's hierarchical clustering method. As a negative control, we added four ribosomal proteins to the clustering. As these proteins constitute essential housekeeping genes, we did not expect them to show strong interactions with any query gene. This assumption could be further confirmed by the clustering analysis (Figure 5F).

Interaction profile similarities map a network of genetic interactions in cancer cells

To generate a network of genetic interactions in cancer cells, we defined genetic interactions based on a combined threshold (absolute π-score > 0.2, p-value < 0.01) and calculated the Pearson correlation coefficient between pairs of interaction profiles among all target genes. We selected all target gene pairs whose profiles similarity exceeded a correlation of 0.6 and connected them to a network, applying a force-directed spring-embedded layout that can position highly correlating genes proximal to each other (Figure 6A). We next used Spatial Analysis of Functional Enrichment (SAFE;

Baryshnikova 2016) to identify regions in the network enriched for specific biological processes as annotated by Gene Ontology (GO; Ashburner et al. 2000; Figure 6B; Supplementary File 1; Supplementary Table 6). SAFE analysis revealed clustering of 16 network regions, which were associated with 66 different GO terms and comprised in total 601 genes. Many of the enriched modules display functional profiles in metabolism, signaling and developmental processes and are as such known to be aberrant in cancer cells. During cancer development and based on the accumulation of cancer-driving mutations, cancer cells commonly acquire changes in signaling pathways that affect proliferation, motility and survival (Martin 2003). These alterations constitute main drivers of malignant transformation. In accordance with this, we could observe a high number of genetic interactions focusing on “tissue development and signal transduction” in cancer cells (Fig. 6B). Other related clusters were “cellular locomotion” and “cell adhesion and immune response”. “Cellular respiration” was among the largest clusters obtained in our analysis with 21 GO terms and 46 genes related to mitochondrial oxidative phosphorylation and metabolic processes. To achieve a high proliferation rate and growth, cancer cells require a large amount of energy in form of ATP as well as metabolite building blocks for the synthesis of e.g. nucleic acids, proteins (Dell’ Antone 2012). In line with this, we obtained a second cluster centering on the conversion of pyruvate to acetyl-CoA, the latter one being the main metabolite required for driving the production of ATP in the citric acid cycle (“Pyruvate metabolism”, Fig. 6B). The high need for building blocks could further explain the strong epistatic interactions of genes implicated in “molecule transport” and “macromolecule biosynthesis”, two additional clusters we obtained with our analysis of genetic interactions in cancer cells. Surprisingly, cancer cells commonly switch from the highly-efficient process of mitochondrial oxidative phosphorylation to the, in terms of ATP production, less efficient degradation of pyruvate into lactate – a metabolic change that is thought to provide the increased amount of nutrients needed to maintain high proliferation rates (Van der Heiden et al. 2009). However, a recent report supports the notion that mitochondrial respiration nevertheless provides an advantage in tumor

development and progression. Thus re-establishment of lost mitochondrial function is also favored in tumor cells (Tan et al. 2015).

In addition, we found processes that are less widely studied and not directly linked to cancer. Protein UFMylation, for example, is a post-translational protein Ubiquitin-like modification, that comprises the conjugation of ubiquitin fold modifier 1 (UFM1) to target proteins. Aberrations in the UFMylation pathways have been linked to several diseases, including cancer (Wei & Xu 2016; Yoo et al. 2014). Interestingly, genes implicated in protein ufmylation cluster with genes regulating intracellular estrogen receptor signaling in our analysis. One major target protein of UFM1 conjugation is activating signal cointegrator 1 (ASC1). Polyufmylation of ASC1 has been shown to promote breast cancer development by transcriptional regulation of estrogen receptor α signaling (Yoo et al., 2014). Similar genetic interactions of genes involved in protein ufmylation and estrogen receptor signaling support the notion of ufmylation as an modulator of estrogen receptor signaling.

DISCUSSION

To identify novel functions of known genes or to assign cellular function to unknown genes, forward genetic screens have been conducted in many model systems ranging from bacteria to human cells (Boutros & Ahringer 2008). Combining high-throughput screening methods with the ability to reliably knock out every gene in the human genome by programmable nucleases now opens up the possibility of studying the consequences of complete or partial loss-of-function mutations with unprecedented accuracy in various mutational backgrounds. Genome-wide screens, predominantly for gene essentiality, have been performed and have identified a large number of known, new and context-specific essential genes (Hart et al. 2015; Wang et al. 2015; Wang et al. 2014; Evers et al. 2016; Morgens et al. 2016; Rauscher et al. 2017; Zhan & Boutros 2016) We developed an analytical approach to integrate dozens of high-throughput CRISPR/Cas9 viability screens independent of screen size, library, Cas9 type and screening protocol. Because, compared to other techniques, CRISPR/Cas9 screens have shown to be a more sensitive method by which perturbation-induced phenotypes

can be discovered in human cells (Wang et al. 2015; Hart et al. 2015), such an approach shows great promise for the systematic discovery of cancer vulnerabilities. We integrated data from 83 screens in human cancer cell lines and analyzed the viability effects of CRISPR/Cas9 perturbations in the context of the cell lines' genetic backgrounds. Using this data, we uncovered novel genetic interactions among genes implicated in tumorigenesis and the resistance to therapy. We further show that genetic interactions differ in cancer cells derived from different tissues, which can shed light on pathways that play important roles in specific cancer types, e.g. *CTNNB1*-dependent Wnt signaling. Furthermore, differential response to gene perturbation can be analyzed in the context of specific driver mutations. Beyond this, genes can be clustered by similarities in their genotype-dependent perturbation response, which can also associate them with joint biological processes but limited currently available data still poses challenges on the analysis.

In our analysis, we grouped mutations into gain-of-function and loss-of-function mutations. However, specific mutations can have distinct functional implications, which could confound the analysis. Moreover, in our analysis we have only been able to look at a limited number of mutated genes. We believe that as more data become available, it will be possible to analyze the mutations at a higher resolution, which could then highlight very specific interactions that could not be discovered previously. Moreover, richer data will allow for a more in-depth investigation of driver genes.

Many efforts have been dedicated to the identification of synthetic lethal interactions and findings have been translated into clinical successes (Lord et al. 2015) We systematically evaluated 2.2 million combinations of genes and found many candidates for synthetic genetic interactions. Among these, we found known interactions, suggesting that our approach could identify known biological pathways. Furthermore, we identified new dependencies that could be interesting targets for drug therapy. One potentially actionable example is a synthetic lethal interaction between *MET* and the *BCL6* co-repressor *BCOR*. Crizotinib, an FDA-approved *MET* inhibitor, might be an interesting candidate for the treatment of *BCOR*-deficient cancers, including stomach, head and neck or esophageal cancer (Gao et al. 2013).

It has previously been demonstrated that profiles of synthetic genetic interactions can group functionally related genes through “guilt by association”. Studies in human cells have formerly relied on RNA interference. However, it has been shown that this method has limitations, such as off-targeting and dosage compensation effects, that can be overcome by CRISPR/Cas9. Our approaches allowed us to analyze interaction profiles using data from many high throughput CRISPR/Cas9 experiments. We created a network that groups genes into clusters with enriched functional profiles. Findings from this analysis may be important for two reasons: first, hypotheses about the function of weakly characterized genes that are frequently deleted in cancer cells can be generated by looking at the common interaction partners members within functional network modules; and second, such a network may serve as a powerful tool to infer the function of entirely uncharacterized genes based on the function of connected genes. For example, over 10% of the genes in our network are not annotated with GO biological processes.

We noticed that our network does not enrich for modules corresponding to housekeeping processes even though this is frequently observed in analogous analyses in other organisms. This is a consequence of the fact that in our data query main effects are consistently very small causing the strength of an interaction to be determined mainly by the difference in target main effect and measured effect. We reason that small query effects are explained by the fact that cancer cells can acquire loss-of-function variants in these genes with suffering selective disadvantage. Consequently, as the knockout of core-essential genes tends to be lethal irrespective of the genetic background of the cell, a genetic interaction is here rarely observed. Our approach therefore specifically maps interactions in pathways that are, in one way or another, relevant in the context of cancer. We believe, however, that screening in isogenic cell line models, where loss-of-function variants that do not naturally occur are introduced artificially, could extend this context.

In its current state, a limiting factor of this type of analysis is the amount of available data. At present, there are approximately 200 genes that have been found to be frequently deleted in the cell lines included in our data and for which synthetic

genetic interactions can be tested. Therefore, only genes that interact with these frequently deleted genes can currently be examined. Nevertheless, this number will improve rapidly as new data are published, which will then allow for the creation of increasingly complex interaction networks. Our approach is scalable and can easily be expanded as new data becomes available. All in all, we believe that the presented approach can be a powerful way to systematically discover synthetic genetic interactions that may be of clinical interest. Furthermore, we believe that it can serve as an important asset to the quest towards more complete understanding of how human genes function. The presented workflow scales well as increasing amounts of data are becoming available.

We expect many more CRISPR/Cas9 screens in various cell lines to be carried out in future. We will expand our analysis once these data become available to improve and diversify our findings. Finally, we aim to extend our analysis to also include data from other experiment types such as physical interactions derived from protein-protein interaction studies. Most synthetic genetic interactions, for example, do not link genes that are members of the same pathways but instead they connect members of two interacting pathways (Kelley & Ideker 2005). Therefore, integrating synthetic interactions and physical interactions derived from protein-protein-interaction experiments might provide important new insights into how biological pathways interact with each other.

METHODS

Genetic profiles of cancer cell lines

To generate a profile of all loss-of-function genes in cancer cell lines all somatic mutation data was downloaded from the COSMIC database (Forbes et al. 2017). Mutations of type ‘Unknown’, ‘Substitution - coding silent’, ‘Complex - deletion inframe’, ‘Complex - insertion inframe’, ‘Deletion - In frame’, ‘Insertion - In frame’, ‘Nonstop extension’ and ‘Substitution - Missense’ were excluded from the data. Additional mutation data was downloaded from the Cancer Cell Line Encyclopaedia (Barretina et

al. 2012) All mutations classified as one of 'Frame_Shift_Del' (frameshift deletion), 'Frame_Shift_Ins' (frameshift insertion), 'Nonsense_Mutation', 'Stop_Codon_del' (deletion of start codon) and 'Stop_Codon_Ins' (insertion of a stop codon) were retained for further analysis. Furthermore, copy number data was downloaded from CCLE (Barretina et al. 2012). These data contain $\log_2$-transformed copy number fold changes between healthy samples and cancer cell lines at the gene level. The absolute copy number of each gene per cell line was inferred from the fold change data as

$$C = \lfloor 2^x \times 2 \rfloor$$

Where C is the absolute copy number and x is the $\log_2$ fold change between cell line and healthy sample. In order to assess whether this provides a realistic estimate of the total copy number we analyzed the derived copy number for all Y-chromosome genes in all 378 female cell lines (Figure S1) where copy numbers of 0 were robustly estimated. Finally, we downloaded pre-processed gene-level copy number data from COSMIC. All genes where a copy number of 0 was estimated in a cell line were marked as loss-of-function genes, excluding Y-chromosome genes. Loss-of-function data from COSMIC and CCLE was aggregated into one matrix which was used as the final profile of cancer cell line loss-of-function. To eliminate sequencing errors and variants specific to the cell line model, we then compared identified loss-of-function events to primary tumor data. We used primary tumor data reported in COSMIC, which include amongst others information from The Cancer Genome Atlas (TCGA; Cancer Genome Atlas Research Network et al. 2013). We removed all loss-of-function events that could not be found in any primary tumor. We furthermore excluded somatic variants for which no entry could be found in dbSNP (Sherry et al. 2001).

Analysis of CRISPR-Cas9 screens

To compare fitness phenotypes of high-throughput CRISPR-Cas9 screens, all screens were reanalyzed uniformly using the Bayesian Analysis of Gene Essentiality (BAGEL; Hart & Moffat 2016). As a first step, all negative selection screens for cell viability were downloaded from the GenomeCRISPR database (Rauscher et al. 2017). Screens in cell

lines that could not be mapped to any of the cell lines in the loss-of-function profile were removed. Moreover, screens in the HeLa cell line were excluded from further analysis, as this cell line is known to be very heterogeneous (Masters 2002). Essentiality scores for all remaining screens were calculated for each screen individually using BAGEL (Hart & Moffat 2016). BAGEL takes as input a table containing $\log_2$-transformed fold changes of single guide RNA (sgRNA) read counts. To generate the input, raw read count data provided by the original authors of the experiment were scaled to a median of 1 and fold changes were calculated for each replicate individually as

$$fc_{sgRNA} = \log_2\left(\frac{rc_{sample}}{rc_{T0}}\right)$$

where rc_{sample} is the normalized read-count measured in the sample cell population and rc_{T0} is the normalized read count measured at time point 0. In some cases, the read-count abundance in the plasmid DNA pool was given instead of time point 0 sequencing data of cells. Thus, the plasmid DNA read-counts were used to calculate the fold changes for all sample replicates of those screens. Furthermore, in 2 cases (Doench et al. 2016; Munoz et al. 2016) no read count data was available. Here we used the fold change values provided by the authors of the original experiment as input for BAGEL. In each screen, all sgRNAs with a measured read count of < 20 at time point 0 were removed from the analysis. Furthermore we excluded all sgRNAs in the GeCKOv2 library (Sanjana et al. 2014) that were flagged as 'isUsed = FALSE' in the 'Achilles_v3.3.8.reagent.table.txt' (https://portals.broadinstitute.org/achilles/datasets/7/download) on the Project Achilles (Aguirre et al. 2016) website. To estimate essentiality scores for genes in a screen, BAGEL relies on a training set of core-essential genes and non-essential genes. These are lists of genes that are known to be essential or non-essential, respectively, in almost every cell line (Hart et al. 2014). For our analysis, we used the gene lists provided on the BAGEL SourceForge webpage (https://sourceforge.net/projects/bagel-for-knockout-screens/files/?source=navbar). Overall, 83 screens in 55 different cell lines were analyzed.

Quality control of screening data

To assess the quality of each screen after BAGEL re-analysis we determined if reference core- and non-essential genes could be identified successfully. We used the ROCR R package (Sing et al. 2005) to generate precision-recall-curves for each experiment (Figure S3). Here the Bayes Factors in the BAGEL output calculated for these core- and non-essential reference genes were considered as predictions. Seven screens achieved an area under the precision-recall-curve of less than 0.85. They were excluded from downstream analysis (Figure S3).

Data normalization

For each gene perturbation in a screen BAGEL provides a Bayes Factor (BF) that can serve as a quantitative measure of essentiality of the gene in that screen (Steinhart et al. 2017) where higher Bayes Factors indicate more essentiality (Hart et al. 2015). In the following we refer to negative Bayes Factors as fitness scores. Gene-level fitness scores determined for all screens were aggregated into one matrix where each column represents one screen and each row represents one gene. Fitness scores were quantile-normalized using the 'normalize.quantiles' function in the 'preprocessCore' R/Bioconductor package (Bolstad 2013; Gentleman et al. 2004). Clustering the normalized fitness scores of each screen, we observed clear batch effects. This was unsurprising as most screens were generated using different lab protocols and libraries. For the analysis shown in Figure 3, batch effects were corrected with the 'ComBat' function in the 'sva' R/Bioconductor package (Leek et al. 2012) using the PubMed ID of each screen as batch variable and the tissue of the screened cell lines as biological covariate. The 'ward.D' method as implemented in R was used for hierarchical clustering of screens. For the clustering 1,000 genes with the biggest standard deviation in fitness scores across screens were selected.

Identification of genotype-specific vulnerabilities

To identify *MYC*-dependent fitness effects in pancreatic cell lines, these cell lines were divided into *MYC*-mutated and *MYC*-wild-type based on COSMIC (Forbes et al. 2017)

somatic mutation data, excluding silent mutations and mutations of type ‘unknown’ but considering all other mutation types. To select differential genes, a two-sided Student’s t-test was performed. False discovery rate was controlled at 20 % using the Benjamini-Hochberg method (Benjamini & Hochberg 1995). To find genes with differential fitness effects in pan-cancer cell lines with *NRAS*, *PIK3CA* or *FGFR3* mutations mutated cell lines were selected as described above. One-way ANOVA followed by computation of Tukey's Honestly Significant Differences (Abdi & Williams 2010) was used to identify candidate genes. This approach was also applied to find genes with differential effects across tissues. The ‘ward.D’ hierarchical clustering method as implemented in R was used for clustering.

Combinatorial testing of gene-gene interactions

To test for differences in fitness response based on loss-of-function genotypes, fitness scores for all CRISPR-Cas9 screens in cell lines that are contained in both COSMIC (Forbes et al. 2017) and CCLE (Barretina et al. 2012) databases were selected. From the global loss-of-function profile we selected all genes that were marked as lost in at least 3 distinct cell lines of the remaining screening data as query genes. In total, 169 genes were selected. Consequently, we identified all combinations between these query genes and genes perturbed in screens (target genes). Target genes were selected such that, fitness scores were available for at least 3 distinct cell lines with and without a query loss-of-function. Overall, we identified ~2.2 million such combinations. As input data for the test, we used quantile-normalized fitness scores before batch-correction. We fitted a linear mixed-effects model for each combination, modeling the loss-of-function genotype as fixed effect and the PubMed ID as a random effect. We additionally modeled the cell line as random effect to account for cell-line-specific biases. For modeling, the R package ‘lme4’ (Bates et al. 2014) was used. The R package ‘lmerTest’ (Kuznetsova et al. 2016) was used to calculate an estimation of significance (p-value) for each model. The false discovery rate (FDR) was estimated using the Benjamini-Hochberg method (Benjamini & Hochberg 1995). To separate cell lines based on their mutation profile all copy number and somatic mutation data were

taken from COSMIC and CCLE and mutations of type 'silent' and 'unknown' were removed. Furthermore, copy number observations of 1-3 were removed from the analysis. We created a binary matrix where 1 denotes an alteration and 0 implies no alteration. We used Fisher's exact test to identify genes with different mutation patterns across sensitive and insensitive cell lines.

Quantification of genetic interactions

Interactions between genes were quantified using the π-score statistic (Horn et al. 2011). π-scores were calculated using the 'HD2013SGImaineffects' function implemented in the R/Bioconductor package 'HD2013SGI' (Laufer et al. 2013). To generate the input for the 'HD2013SGImaineffects' function, batch corrected fitness scores were entered by subtracting column means and scaled by dividing columns by their standard deviation. To cluster genes from the NADH-ubiquinone oxidoreductase and Mediator complexes by their π-score profiles, complex member showing low interaction frequency with the query genes, determined by a π-score standard deviation threshold of 0.4, were removed. The remaining genes were clustered using the 'ward.D' method for hierarchical clustering as implemented in R.

Network modeling

To create a map of genetic interactions we used the matrix of π-scores as calculated in the previous step and applied a double cutoff to select interactions. We considered every target-query combination as an interaction if its absolute π-score was greater than 0.2 and if it achieved a p-value of less than 0.01 in the combinatorial tests. Positive interactions (positive π-score > 0) were assigned the value 1, negative interactions (negative π-score) were assigned the value -1 and non-interacting pairs were given the value 0. Based on the resulting ternary matrix we calculated the Pearson-correlation between the interaction profiles of all possible target gene combinations. Correlations between genes > 0.6 were selected as edges for the network and were rescaled to a range of 0-1. The network was visualized using Cytoscape (Shannon et al. 2003). A

force-directed spring-embedded layout was used to position the nodes of the network, using the rescaled correlation values as edge-weights. The visual representation of the network was inspired by previous studies in yeast (Baryshnikova 2016a). The Spatial Analysis of Functional Enrichment (SAFE; Baryshnikova 2016b) Cytoscape-plugin was used to identify functional modules in the network. For SAFE analysis, the map-based distance-metric was chosen with a maximum distance threshold of 0.6 (percentile). To build the composite map, a minimal landscape size of 10 was chosen and the Jaccard distance was used as a similarity metric for group attributes with a similarity threshold of 0.75. As background for the enrichment, all nodes in the annotation standard were chosen (Baryshnikova 2016b). In SAFE the annotation standard is a binary matrix of genes (rows) and annotation terms (columns). A value of 1 indicates that a gene is annotated with a specific annotation term. For our analysis, we generated such an annotation standard containing Gene Ontology (GO; Ashburner et al. 2000) Biological Process annotations for all target genes tested. GO annotations were downloaded from the example data section of the SAFE algorithm's GitHub page (https://github.com/baryshnikova-lab/safe-data/blob/master/attributes/go_Hs_P_160509.txt.gz) and filtered to contain only genes tested in our interaction analysis.

Availability of source code

All R code that was written for the analyses described in this study is available as a R markdown file (Supplementary Code). Furthermore, Cytoscape session files related to Figures 6 and S6 are available (Supplementary Files 1 and 2).

Acknowledgements

We thank Niklas Rindtorff, Tianzuo Zhan, Johannes Betge and Christian Scheeder for critical comments on the manuscript. We would like to thank the Boutros lab for helpful discussions.

Author contributions

BR, FH, and MB designed the study. BR wrote the analysis code. TH consulted on statistical analysis. All authors discussed and analyzed results. BR, FH, LH and MB wrote the manuscript. All authors read and approved the final manuscript.

Conflict of interest

We declare no conflict of interest.

Funding

B.R. was supported by the BMBF-funded Heidelberg Center for Human Bioinformatics (HD-HuB) within the German Network for Bioinformatics Infrastructure (de.NBI) (Grant #031A537A). Work in the Boutros lab is supported in part by an ERC Advanced grant.

REFERENCES

Abdi, H. & Williams, L.J., 2010. Tukey's Honestly Significant Difference (HSD) Test.

Aguirre, A.J. et al., 2016. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discovery*, 6(8).

Al-Khalaf, H.H. et al., 2017. p16(INK4A) enhances the transcriptional and the apoptotic functions of p53 through DNA-dependent interaction. *Molecular carcinogenesis*.

Ashburner, M. et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), pp.25–9.

Barrangou, R., 2014. RNA events. Cas9 targeting and the CRISPR revolution. *Science (New York, N.Y.)*, 344(6185), pp.707–8.

Barretina, J. et al., 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391), pp.603–307.

Baryshnikova, A., 2016a. Exploratory Analysis of Biological Networks through Visualization, Clustering, and Functional Annotation in Cytoscape. *Cold Spring Harbor protocols*, 2016(6), p.pdb.prot077644.

Baryshnikova, A. et al., 2010. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods*, 7(12), pp.1017–24.

Baryshnikova, A., 2016b. Systematic Functional Annotation and Visualization of Biological Networks. *Cell systems*, 2(6), pp.412–21.

Bates, D. et al., 2014. Fitting Linear Mixed-Effects Models using lme4.

Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. *Journal of the Royal Statistical Society*, 57(1), pp.289–300.

Bergman, A. & Siegal, M.L., 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424(6948), pp.549–52.

Bolstad, B.M., 2013. preprocessCore: A collection of pre-processing functions.

Boutros, M. & Ahringer, J., 2008. The art and design of genetic screens: RNA interference. *Nature Reviews Genetics*, 9(7), pp.554–566.

Bridges, C.B., 1922. The Origin of Variations in Sexual and Sex-Limited Characters. *The American Naturalist*, 56(642), pp.51–63.

Byrne, A.B. et al., 2007. A global analysis of genetic interactions in Caenorhabditis elegans. *Journal of biology*, 6(3), p.8.

Cancer Genome Atlas Research Network, K. et al., 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*, 45(10), pp.1113–20.

Cooper, A.B. et al., 2006. A unique function for cyclin D3 in early B cell development. *Nature immunology*, 7(5), pp.489–97.

Costanzo, M., Van der Sluis, B., Koch, E.N., et al., 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 353(6306).

Costanzo, M. et al., 2010. The Genetic Landscape of a Cell. *Science*, 327(5964).

David, C.J. et al., 2016. TGF-β Tumor Suppression through a Lethal EMT. *Cell*, 164(5), pp.1015–30.

Davierwala, A.P. et al., 2005. The synthetic genetic interaction spectrum of essential genes. *Nature genetics*, 37(10), pp.1147–52.

Dell' Antone, P., 2012. Energy metabolism in cancer cells: how to explain the Warburg

and Crabtree effects? *Medical hypotheses*, 79(3), pp.388–92.
Diersch, S. et al., 2016. Kras(G12D) induces EGFR-MYC cross signaling in murine primary pancreatic ductal epithelial cells. *Oncogene*, 35(29), pp.3880–6.
Doench, J.G. et al., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*, 34(2), pp.184–91.
Doudna, J.A. & Charpentier, E., 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213), p.1258096.
Escudero-Esparza, A. et al., 2016. Complement inhibitor CSMD1 acts as tumor suppressor in human breast cancer. *Oncotarget*, 7(47), pp.76920–76933.
Evers, B. et al., 2016. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature biotechnology*, 34(6), pp.631–3.
Fischer, B. et al., 2015. A map of directional genetic interactions in a metazoan cell. *eLife*, 4, p.e05464.
Forbes, S.A. et al., 2017. COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*, 45(D1), pp.D777–D783.
Gao, J. et al., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*, 6(269), p.pl1.
Garnett, M.J. et al., 2012. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 483(7391), pp.570–575.
Gentleman, R.C. et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10), p.R80.
Greig, K.T., Carotta, S. & Nutt, S.L., 2008. Critical roles for c-Myb in hematopoietic progenitor cells. *Seminars in immunology*, 20(4), pp.247–56.
Grossmann, V. et al., 2011. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood*, 118(23), pp.6153–63.
Hart, T. et al., 2015. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, 163(6), pp.1515–26.
Hart, T. et al., 2014. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Molecular Systems Biology*, 10(7), pp.733–733.
Hart, T. & Moffat, J., 2016. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC bioinformatics*, 17(1), p.164.
Hartman, J.L., Garvik, B. & Hartwell, L., 2001. Principles for the buffering of genetic variation. *Science (New York, N.Y.)*, 291(5506), pp.1001–4.
Hartwell, L.H. et al., 1997. Integrating genetic approaches into the discovery of anticancer drugs. *Science (New York, N.Y.)*, 278(5340), pp.1064–8.
Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B., 2009. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science (New York, N.Y.)*, 324(5930), pp.1029–33.
Horlbeck, M.A. et al., 2016. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife*, 5.
Horn, T. et al., 2011. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nature Methods*, 8(4), pp.341–346.

Huynh, K.D. et al., 2000. BCoR, a novel corepressor involved in BCL-6 repression. *Genes & development*, 14(14), pp.1810–23.

Iorio, F. et al., 2016. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, 166(3), pp.740–754.

Kaelin, W.G., 2005. The concept of synthetic lethality in the context of anticancer therapy. *Nature reviews. Cancer*, 5(9), pp.689–98.

Kelley, R. & Ideker, T., 2005. Systematic interpretation of genetic interactions using protein networks. *Nature biotechnology*, 23(5), pp.561–6.

Kompier, L.C. et al., 2010. FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. V. Cheriyath, ed. *PloS one*, 5(11), p.e13821.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B., 2016. Title Tests in Linear Mixed Effects Models.

Laufer, C. et al., 2013. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nature Methods*, 10(5), pp.427–431.

Leek, J.T. et al., 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6), pp.882–883.

Liggett, W.H. & Sidransky, D., 1998. Role of the p16 tumor suppressor gene in cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 16(3), pp.1197–206.

Lin, J.E. et al., 2010. The hormone receptor GUCY2C suppresses intestinal tumor formation by inhibiting AKT signaling. *Gastroenterology*, 138(1), pp.241–54.

Little, C.D. et al., 1983. Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature*, 306(5939), pp.194–6.

Liu, H. et al., 2017. c-Myc Alteration Determines the Therapeutic Response to FGFR Inhibitors. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 23(4), pp.974–984.

Liu, J. et al., 2015. ErbB2 Pathway Activation upon Smad4 Loss Promotes Lung Tumor Growth and Metastasis. *Cell reports*, 10(9), pp.1599–1613.

López-Knowles, E. et al., 2006. PIK3CA mutations are an early genetic alteration associated with FGFR3 mutations in superficial papillary bladder tumors. *Cancer research*, 66(15), pp.7401–4.

Lord, C.J., Tutt, A.N.J. & Ashworth, A., 2015. Synthetic Lethality and Cancer Therapy: Lessons Learned from the Development of PARP Inhibitors. *Annual Review of Medicine*, 66(1), pp.455–470.

Luo, J., Solimini, N.L. & Elledge, S.J., 2009. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell*, 136(5), pp.823–37.

Mani, R. et al., 2008. Defining genetic interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), pp.3461–6.

Martin, G.S., 2003. Cell signaling and cancer. *Cancer cell*, 4(3), pp.167–74.

Martin, H. et al., 2015. Differential genetic interactions of yeast stress response MAPK pathways. *Molecular Systems Biology*, 11(4), pp.800–800.

Masters, J.R., 2002. HeLa cells 50 years on: the good, the bad and the ugly. *Nature reviews. Cancer*, 2(4), pp.315–9.

Meyer, N. & Penn, L.Z., 2008. Reflecting on 25 years with MYC. *Nature reviews. Cancer*, 8(12), pp.976–90.

Morgens, D.W. et al., 2016. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nature biotechnology*, 34(6), pp.634–6.

Munoz, D.M. et al., 2016. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discovery*, 6(8).

Nagel, R., Semenova, E.A. & Berns, A., 2016. Drugging the addict: non-oncogene addiction as a target for cancer therapy. *EMBO reports*, 17(11), pp.1516–1531.

Nijman, S.M.B., 2011. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. *FEBS letters*, 585(1), pp.1–6.

Rauscher, B. et al., 2017. GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Research*, 45(D1), pp.D679–D686.

Sanjana, N.E., Shalem, O. & Zhang, F., 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nature methods*, 11(8), pp.783–4.

Shajahan-Haq, A.N. et al., 2014. MYC regulates the unfolded protein response and glucose and glutamine uptake in endocrine resistant breast cancer. *Molecular cancer*, 13(1), p.239.

Shalem, O. et al., 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, 343(6166).

Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498–504.

Sherry, S.T. et al., 2001. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29(1), pp.308–11.

Sing, T. et al., 2005. ROCR: visualizing classifier performance in R. *Bioinformatics (Oxford, England)*, 21(20), pp.3940–1.

Snijder, B. et al., 2013. Predicting functional gene interactions with the hierarchical interaction score. *Nature Methods*, 10(11), pp.1089–1092.

Srivas, R. et al., 2016. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. *Molecular cell*, 63(3), pp.514–25.

Steinhart, Z. et al., 2017. Genome-wide CRISPR screens reveal a Wnt-FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nature medicine*, 23(1), pp.60–68.

Stojanova, A. et al., 2016. MYC interaction with the tumor suppressive SWI/SNF complex member INI1 regulates transcription and cellular transformation. *Cell cycle (Georgetown, Tex.)*, 15(13), pp.1693–705.

Tan, A.S. et al., 2015. Mitochondrial genome acquisition restores respiratory function and tumorigenic potential of cancer cells without mitochondrial DNA. *Cell metabolism*, 21(1), pp.81–94.

Tang, M.-R. et al., 2012. CSMD1 exhibits antitumor activity in A375 melanoma cells through activation of the Smad pathway. *Apoptosis : an international journal on programmed cell death*, 17(9), pp.927–37.

Tong, A.H. et al., 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)*, 294(5550), pp.2364–8.

Torti, D. & Trusolino, L., 2011. Oncogene addiction as a foundational rationale for targeted anti-cancer therapy: promises and perils. *EMBO molecular medicine*, 3(11), pp.623–36.

Tortora, G. et al., 2003. Combined targeted inhibition of bcl-2, bcl-XL, epidermal growth factor receptor, and protein kinase A type I causes potent antitumor, apoptotic, and antiangiogenic activity. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 9(2), pp.866–71.

Tzelepis, K. et al., 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell reports*, 17(4), pp.1193–1205.

Wang, C.Q. et al., 2015. Cbfb deficiency results in differentiation blocks and stem/progenitor cell expansion in hematopoiesis. *Leukemia*, 29(3), pp.753–7.

Wang, T. et al., 2017. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*, 168(5), p.890–903.e15.

Wang, T. et al., 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*, 343(6166).

Wang, T. et al., 2015. Identification and characterization of essential genes in the human genome. *Science (New York, N.Y.)*, 350(6264), pp.1096–101.

Waters, C.E. et al., 2014. The FHIT gene product: tumor suppressor and genome "caretaker". *Cellular and molecular life sciences : CMLS*, 71(23), pp.4577–87.

Wei, Y. & Xu, X., 2016. UFMylation: A Unique & Fashionable Modification for Life. *Genomics, proteomics & bioinformatics*, 14(3), pp.140–6.

Weiske, J., Albring, K.F. & Huber, O., 2007. The tumor suppressor Fhit acts as a repressor of beta-catenin transcriptional activity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), pp.20344–9.

Yoo, H.M. et al., 2014. Modification of ASC1 by UFM1 is crucial for ERα transactivation and breast cancer development. *Molecular cell*, 56(2), pp.261–74.

Zhan, T. & Boutros, M., 2016. Towards a compendium of essential genes - From model organisms to synthetic lethality in cancer cells. *Critical reviews in biochemistry and molecular biology*, 51(2), pp.74–85.

Zhan, T., Rindtorff, N. & Boutros, M., 2017. Wnt signaling in cancer. *Oncogene*, 36(11), pp.1461–1473.

Zhang, R. & Song, C., 2014. Loss of CSMD1 or 2 may contribute to the poor prognosis of colorectal cancer patients. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 35(5), pp.4419–23.

Zhao, S. et al., 2010. Expression of oncogenic K-ras and loss of Smad4 cooperate to induce the expression of EGFR and to promote invasion of immortalized human pancreas ductal cells. *International journal of cancer*, 127(9), pp.2076–87.

Zhu, Q. et al., 2016. miR-10b exerts oncogenic activity in human hepatocellular carcinoma cells by targeting expression of CUB and sushi multiple domains 1 (CSMD1). *BMC cancer*, 16(1), p.806.

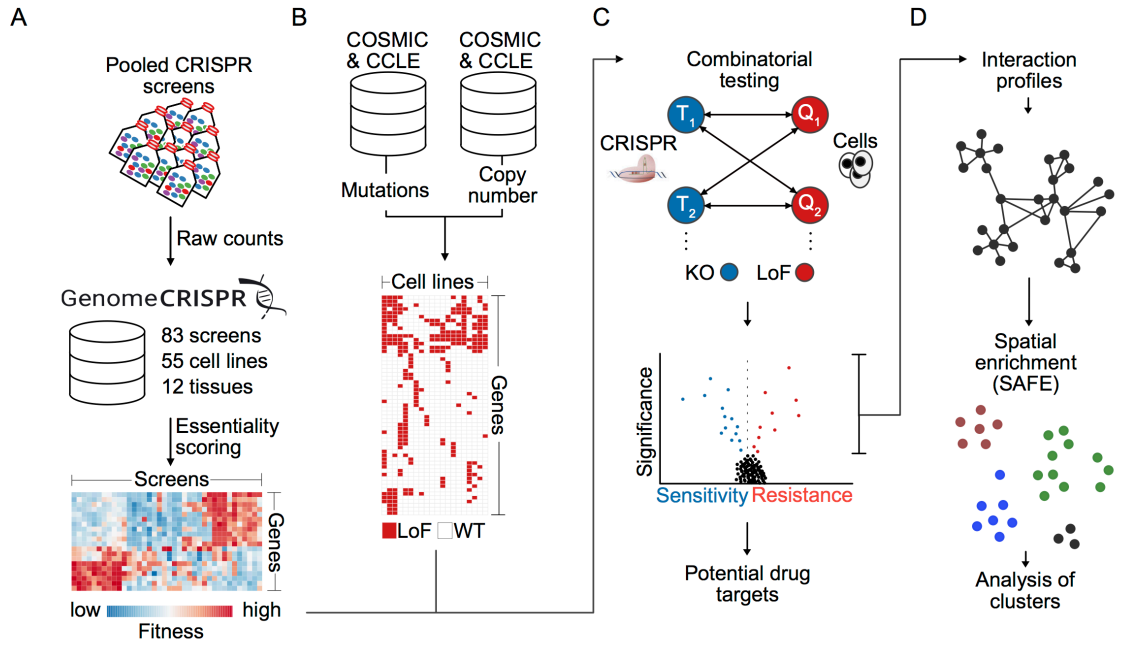

Figure 1 An integrated analysis approach to find genetic interactions in cancer cells. (A) Data from CRISPR-Cas9 screens in 83 cancer cell lines were reanalyzed and integrated. The results were combined into a global perturbation response profile. (B) Mutation and copy number data from the COSMIC and CCLE databases were used to create a map of loss-of-function alterations across over 1,000 cancer cell lines. (C) To identify synthetic lethal dependencies between gene combinations that could serve as potential targets for new drug development, perturbation response of more than 2.2 million gene-gene combinations were examined using a linear mixed-effects model. (D) Interaction profiles were calculated for gene combinations based on the correlation of their interactions as determined by interaction scores (π scores) and statistical significance. Spatial enrichment analysis was performed to identify functional modules in the network.

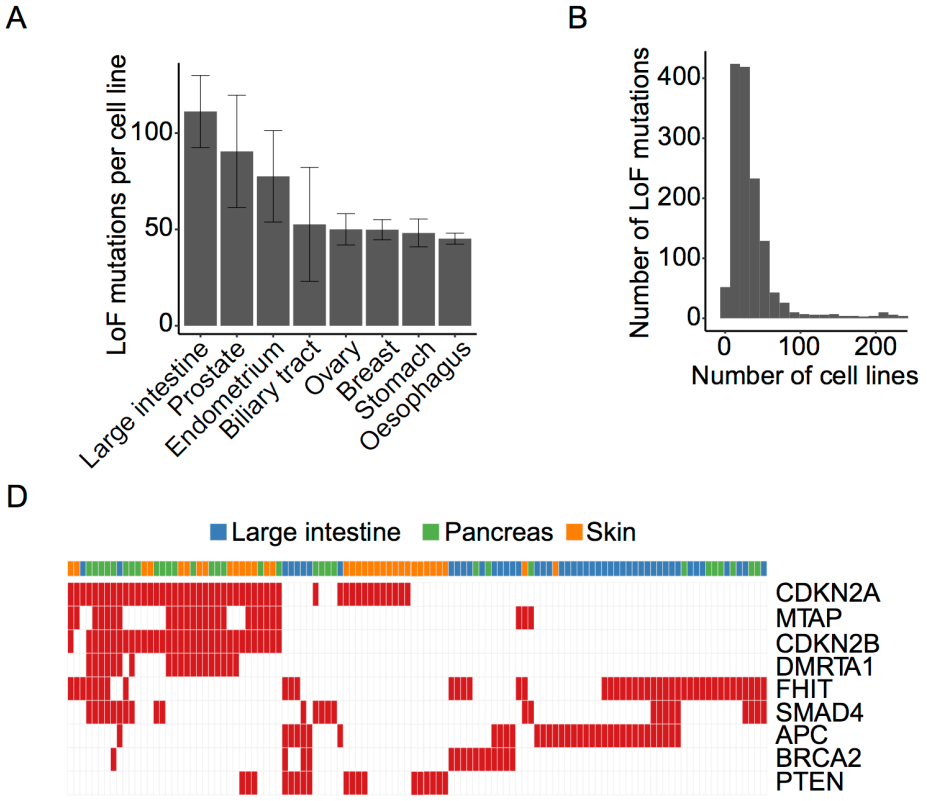

Figure 2 Mapping loss-of-function variations in cancer cells. (A) A summary of 8 tissues in which corresponding cell lines show the highest abundance of loss-of-function events. (B) Distribution of the number of loss-of-function mutations in the cell lines. (C) Clustering of loss-of-function profiles separates cells that originated from large intestine, pancreatic and skin cancers. Red squares indicate a loss-of-function mutation.

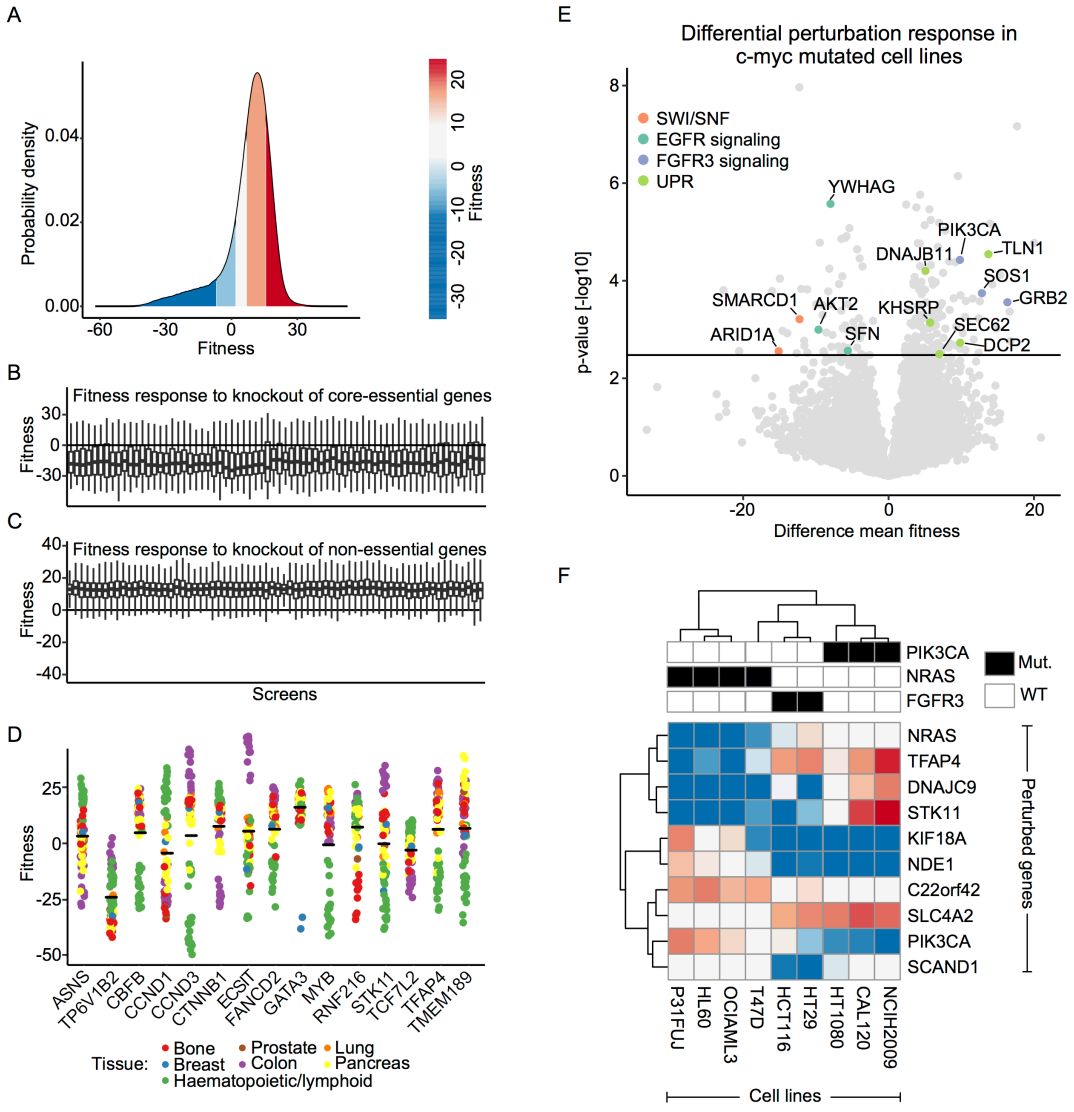

Figure 3 Analysis of cell fitness phenotypes highlights context-dependent vulnerabilities. (A) Distribution of the calculated fitness scores. (B, C) Fitness scores of core-essential and non-essential gold standard gene set across screens. (D) Tissue-specific effects of gene perturbation on the fitness of cancer cell lines. Shown here are 15 genes in which highly significant fitness differences were observed across tissues. Each dot represents one experiment. (E) Volcano plot showing differential fitness response to CRISPR/Cas9 perturbations in cell lines with a *c-Myc* mutation. The horizontal line indicates the 20% false discovery rate (FDR) threshold. Negative values on the x-axis indicate increase vulnerability of *c-Myc* mutated cells, and positive values suggest resistance to the perturbation. Genes that belong to pathways that are enriched among significant genes are highlighted. (F) Clustering of the differential responses to genetic perturbation based on the mutation status of the *PIK3CA*, *NRAS* and *FGFR3* oncogenes in pan-cancer cell lines.

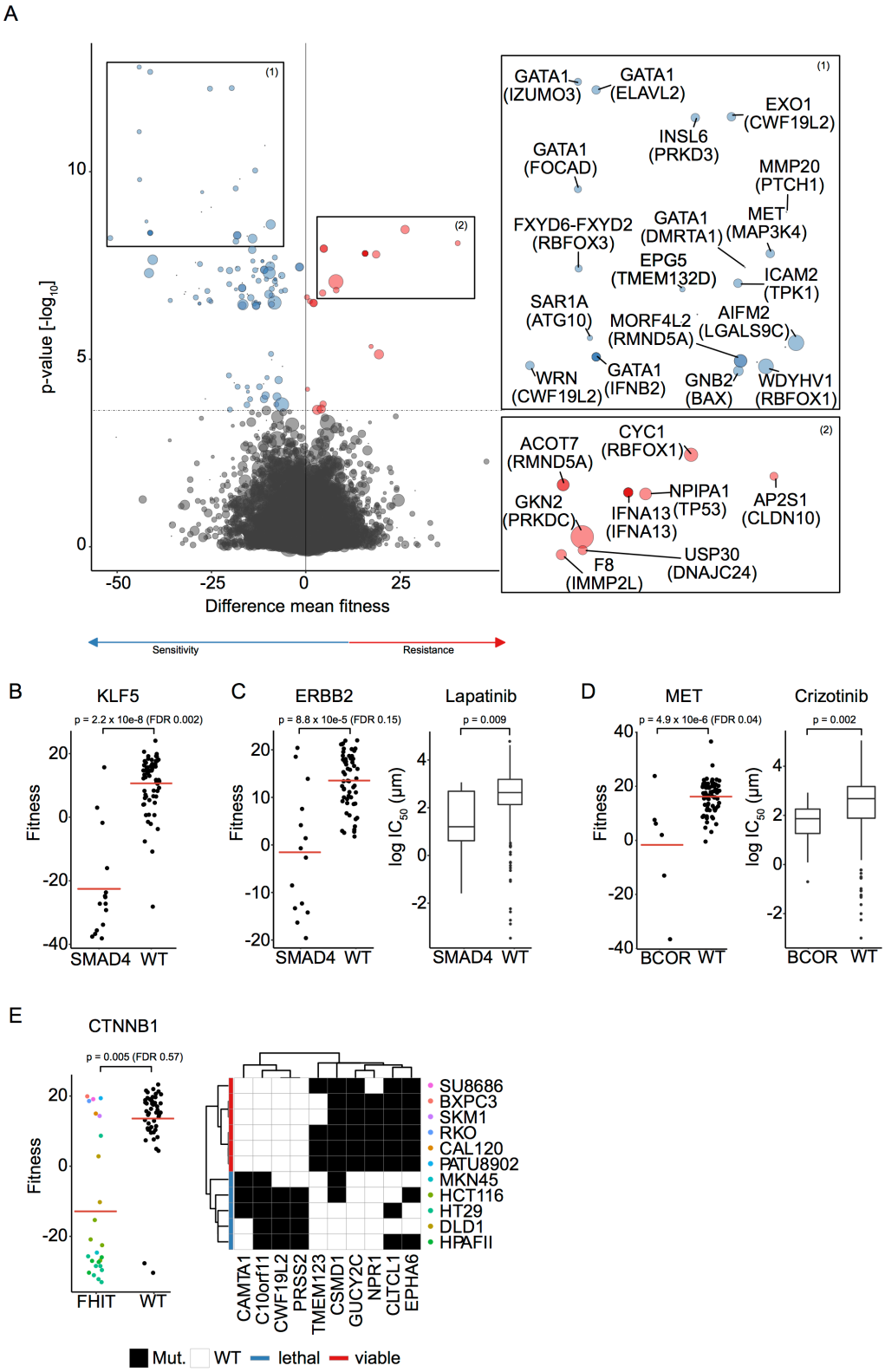
A
p-value [-log10]
Difference mean fitness
Sensitivity
Resistance
GATA1 (IZUMO3)
GATA1 (ELAVL2)
EXO1 (CWF19L2)
INSL6 (PRKD3)
GATA1 (FOCAD)
MMP20 (PTCH1)
FXYD6-FXYD2 (RBFOX3)
GATA1 (DMRTA1)
MET (MAP3K4)
EPG5 (TMEM132D)
ICAM2 (TPK1)
SAR1A (ATG10)
MORF4L2 (RMND5A)
AIFM2 (LGALS9C)
WRN (CWF19L2)
GATA1 (IFNB2)
GNB2 (BAX)
WDYHV1 (RBFOX1)
ACOT7 (RMND5A)
CYC1 (RBFOX1)
GKN2 (PRKDC)
IFNA13 (IFNA13)
NPIPA1 (TP53)
AP2S1 (CLDN10)
F8 (IMMP2L)
USP30 (DNAJC24)
B
KLF5
p = 2.2 x 10e-8 (FDR 0.002)
Fitness
SMAD4 WT
C
ERBB2
p = 8.8 x 10e-5 (FDR 0.15)
Lapatinib
p = 0.009
log IC50 (µm)
D
MET
p = 4.9 x 10e-6 (FDR 0.04)
Crizotinib
p = 0.002
BCOR WT
E
CTNNB1
p = 0.005 (FDR 0.57)
FHIT WT
CAMTA1
C10orf11
CWF19L2
PRSS2
TMEM123
CSMD1
GUCY2C
NPR1
CLTCL1
EPHA6
SU8686
BXPC3
SKM1
RKO
CAL120
PATU8902
MKN45
HCT116
HT29
DLD1
HPAFII
Mut.
WT
lethal
viable

Figure 4 Predicted gene-gene associations. (A) A volcano plot showing effect size (difference between the mean fitness effect of a CRISPR perturbation in cells with and without a specific loss-of-function mutation) and the significance of the associations. The horizontal line indicates a 0.2 false discovery rate threshold. Each circle represents one genetic interaction. The circle size indicates the number of cell lines with the respective loss-of-function variation (minimum 3). Insets 1 and 2 show a magnified view of highly significant positive and negative interactions, respectively. Labels indicate the perturbed gene and the loss-of-function gene (in brackets). (B-D) Loss-of-function-dependent effects of the genetic perturbations for 3 selected candidates, including the response of to the corresponding inhibitors. (E) Details of the response to *β-catenin* perturbation in *FHIT*-deficient cells. Colors link *FHIT*-deficient cell lines to a mutation profile detailing the mutational status of 10 genes, thereby identifying cells that are sensitive and resistant to *β-catenin* knockout, respectively.

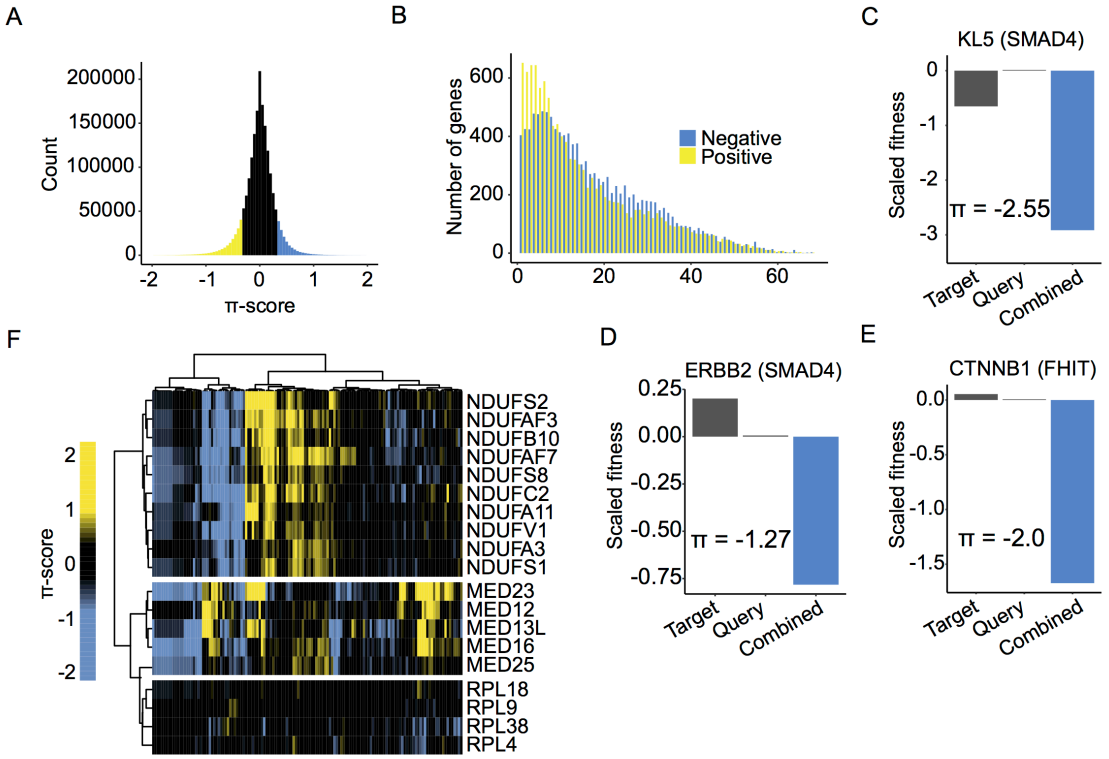

Figure 5 Scoring of synthetic genetic interactions based on π scores**.** (A) Distribution of the interaction scores (π scores). (B) Distribution of the positive and negative interactions of genes. (C-E) Comparison of the combined (measured) effect and the main effects of both target gene and query gene for the selected interactions. (I) Clustering of members of the NADH-ubiquinone oxidoreductase, mediator complex proteins and ribosomal proteins by their genetic interaction profiles.

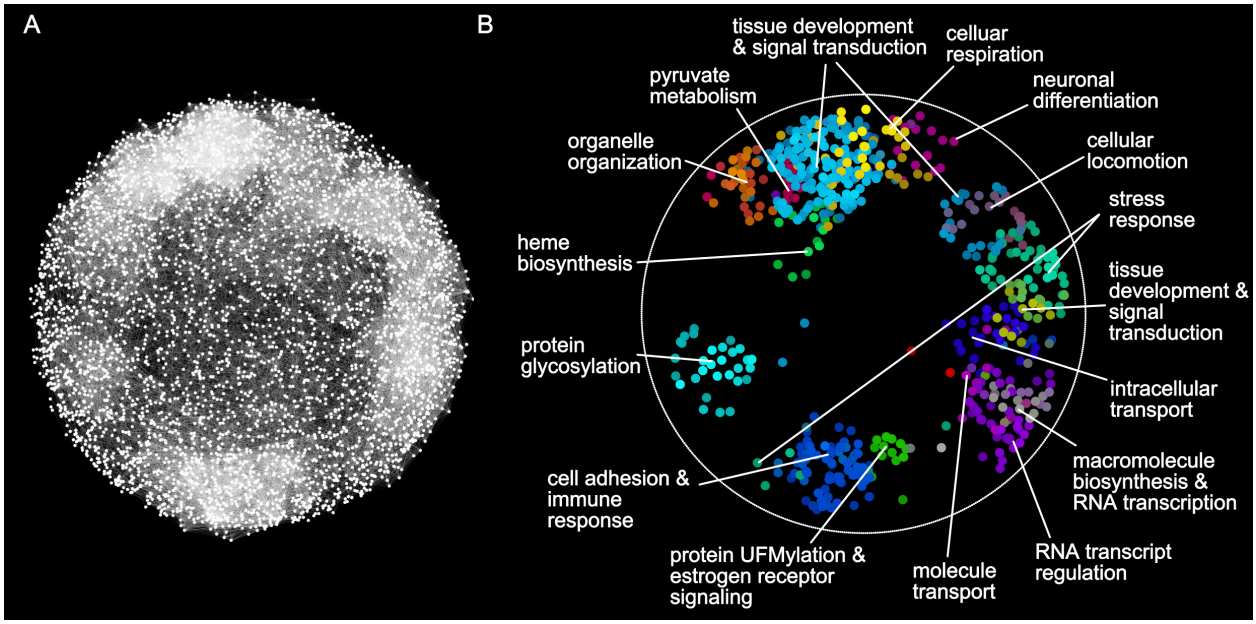

Figure 6 A map of genetic interactions in a human cancer cells. (A) In the correlation-based network, 4,616 nodes (genes) are connected by 39,805 links (Pearson correlation of interaction profiles >0.6). An edge-weighted spring embedded layout was used to position the nodes. Nodes with similar interaction profiles are located proximal to each other. (B) Spatial enrichment analysis with the SAFE algorithm highlights network modules that contain genes with similar functional annotations based on Gene Ontology biological processes. These GO terms are summarized by the module labels in the figure.