Mapping Genetic Interactions of All Paralogs in the Human Genome

Author(s): Michael J Flister[1], Daniel Verduzco[1], Sakina Petiwala[1], Christos Ghekas[2], Xu Shi[2], Charles Lu[1], Zoltan Dezso[2]

Department: [1]AbbVie Inc., 1 North Waukegan Rd., North Chicago, IL 60064; [2]AbbVie Bay Area, 1000 Gateway Boulevard, South San Francisco, CA 94080

Correspondence: Michael J. Flister (michael.flister@pfizer.com) and Zoltan Dezso (zoltan.dezso@abbvie.com)

ABSTRACT

Background: Functionally redundant paralogs in the human genome are the most common source of synthetic lethality (i.e., loss of one paralog conveys dependency to another). However, most paralogs have yet to be experimentally tested and the human paralogome remains largely uncharacterized.

Results: We performed the first pairwise genetic screen of all human paralogs using a multiplexed CRISPR-Cas12 library, which revealed that digenic synthetic lethalities are relatively rare (<0.5% of all paralog pairs) and varied in penetrance across different cancer models. We hypothesized that the variable penetrance of digenic synthetic lethalities was a result of complex polygenic interactions with endogenous factors. A multivariable regression analysis of 1,278 pairs and endogenous cellular features across 30 cancer models revealed that perturbations of related pathways were frequently predictive of paralog synthetic lethality. A machine learning classifier was also built to predict synthetic lethalities using a weighted set of true positives, accounting for the variable penetrance of synthetic lethal interactions. Intuitively, the predictive scores revealed that the penetrance of synthetic lethal interactions was driven by the overlap and essentiality of the protein-protein interactions for each paralog pair.

Conclusions: This study provided a comprehensive analysis of all digenic interactions in the human paralogome, as well as the key features that underlie the heterogeneity in synthetic lethalities that have been reported here and elsewhere.

BACKGROUND

Paralogs are functionally related genes that arise from genomic duplications and serve as a major source of genetic evolution through selective adaptations in most species, including humans [1,2]. The human genome has an estimated 13,320 paralogs belonging to ~3,500 families (ranging from 2 to 48 members) [3]. Many parlogs are thought to share compensatory functions and are therefore more likely to exhibit synthetic lethality (i.e., the loss of one paralog conveys dependency to another) [4]. The loss of a paralog in tumor cells can be driven by a selective advantage or as a genomic passenger that is co-deleted with another tumor suppressor gene, which renders the tumor cells sensitive to selective inhibition of the remaining paralog. Because normal cells retain both paralogs, a selective inhibitor of a single paralog is well tolerated in normal tissues, whereas it is synthetic lethal in the tumor cells [2,5].

Although the concept of synthetic lethal paralogs has existed for nearly a century [6], the majority of paralogs remain uncharacterized and only recently have genomic sequences for multiple organisms enabled the prediction of synthetic lethal interactions [3]. Additionally, the number of potential genetic interactions across all paralog families is immense ($>10^{12}$ interactions) and prohibitive for comprehensive genetic screening [4]. The expansion of genomewide loss-of-functions (LoF) CRISPR screens has uncovered novel synthetic lethal interactions, yet these data are limited by the incomplete representation of LoF events in human patients cancer cell models and therefore are likely to miss patient-relevant synthetic lethal interactions [2,7]. Alternatively, while synthetic lethal interactions of select candidate pairs have been tested through combinatorial LoF CRISPR screening, these pairs are typically limited to a small subset of the known human paralogs. As such, the potential functional interactions of most paralog pairs in the human genome remain unknown.

Here, we present the first genetic screen of the 36,648 known paralog pairs in the human genome using a multiplexed CRISPR-Cas12 library, which identified multiple synthetic lethal and buffering interactions. This unbiased screen also confirmed a wide variability in the penetrance of synthetic lethal interactions across different cell lineages and genetic backgrounds, suggesting that other endogenous factors influence the sensitivity to synthetic lethal mechanisms. To test this possibility, we designed a compact multiplexed CRISPR-Cas12 library to screen a subset of paralog pairs across 20 cancer cell models, which revealed the cellular contexts associated with sensitivity to synthetic lethal interactions. These data were integrated with a recently published paralog screen across 10 additional cancer cell models [8] to form the largest synthetic lethality map to date. Finally, this comprehensive dataset was used to refine a machine learning model capable of predicting synthetic lethal interactions.

RESULTS

Mapping genetic interactions of 36,648 paralog pairs in the human genome

Multiple previous attempts to characterize paralog synthetic lethalities have been limited to testing approximately 10% or less of the human paralogome due to size restrictions of conventional multiplexed CRISPR libraries. To overcome these limitations, we developed a CRISPR-Cas12 screening approach to test all 36,648 paralog pairs in the human genome using a multiplexed library design with six independent constructs per paralog pair (**Table S1**). This CRISPR-Cas12 library design utilized arrays of four guides per construct and two guides per paralog, which performed optimally for combinatorial screens [9] and enabled a minimal library size (n = 113,502 constructs) for screening all paralogs (n = 13,128 singletons and n = 36,648 pairs) (**Figure 1A**). Pairwise screening of all human paralogs was performed using NCI-H1299 and MDA-MB-231 cells that stably expressed enAsCas12a, and synthetic lethal interactions were identified by comparing the observed effects of the double knockout (KO) with the additive effects of each single KO (**Figure 1B** and **Table S2**). Library performance was assessed by comparing the effects of single gene KOs in the CRISPR-Cas12 library with the DepMap essentiality profiles of NCI-H1299 and MDA-MB-231, which demonstrated comparable performance to traditional CRISPR-Cas9 screening (**Figure 1C**, **D**).

Out of the 36,648 paralog pairs that were tested, surprisingly few synthetic lethal interactions were observed in NCI-H1299 (n = 144 pairs) or MDA-MDA-231 (n = 45 pairs), relative to the essential singletons that were detected in NCI-H1299 (n = 406 singletons) or MDA-MB-231 (n = 384 singletons) (**Figure 1E**). Of the 224 essential singletons that were shared by both cell lines, 193 were classified as common essential by the DepMap (86%). In comparison, proportionally fewer common digenic paralog dependencies (~10%) contributed to the overall dependency profile of both cancer models, and NCIH1299 cells unexpectedly had roughly six times the amount of unique synthetic lethal paralog pairs. Of the 93 unique synthetic lethal paralog pairs that were detected in either cell model, 53 have been replicated elsewhere and 40 paralog pairs were not previously reported as (**Figure 1F**, **G** and **Table S3**). It was notable that 20 out of the 26 (77%) synthetic lethalities that were detected in both cell models have been replicated elsewhere (**Table S3**), indicating that these pairs likely represent a small class of common synthetic lethalities. One previously unreported pair of common synthetic lethal paralogs was the TATA-box binding protein associated factors, TAF9 and TAF9b, which have overlapping roles in regulating the core promoters of gene transcription [10] (**Figure H**). Another novel common synthetic lethal paralog pair was the heat shock protein family A (Hsp70) members, HSPA8 and HSPA1A, which can chaperon the same client proteins interchangeably [11–13] (**Figure 1I**). Intriguingly, the loss of HSPA8 and the closely related HSP70 proteins, HSPA1B or HSPA2, also resulted in synthetic lethal interactions, whereas the loss of other expressed family members was well tolerated (**Figure 1I**). Collectively, these

data highlight the power of unbiased paralog mapping to discover synthetic lethal interactions, as well delineate lethalities within large families, such as the HSP70 family.

In addition to synthetic lethalities, combinatorial CRISPR screens also capture positive interactions [14], whereby the loss of two paralogs results in a net positive advantage that is greater than the sum of either paralog alone (i.e., synthetic buffering). To test this possibility across all paralog pairs in the human genome, digenic buffering interactions were identified by comparing the observed effects of the double knockout (KO) with the additive effects of each single KO (**Figure 2A**). Synthetic buffering interactions of 18 paralog pairs were observed in NCI-H1299 (n = 17) or MDA-MDA-231 (n = 6), of which five buffering pairs were shared across both models (**Figure 2B**). Notably, 11 of 18 (61%) of the unique buffering pairs were closely associated by STRING analysis with tumor suppressor pathways, including MOB1B/MOB2 (**Figure 2C**, **D**), ARHGAP12/ARHGAP21 (**Figure 2E**, **F**), and EXOSC7/EXOSC8 (**Figure 2G**, **H**). Although these proteins, such as MOB1B and MOB2, have been individually associated with tumor suppressor pathways [15], this is the first evidence of functional compensation by these tumor suppressor paralogs. For example, MOB1B and MOB2 are members of the MOB family (monopolar spindle-one-binder proteins) with distinct and overlapping roles in regulating the Hippo pathway [15]. Taken together, these data indicate the presence of functionally redundant paralog pairs, such as MOB1B and MOB2, exist within tumor suppressor pathways and likely convey a selective advantage for tumor cells upon loss of both paralogs [16].

Context dependency of paralog synthetic lethalities

Genetic mapping of all known human paralog pairs revealed relatively few synthetic lethalities that were common between cancer models, a phenomenon that has also been observed in more focused screening efforts [3,8,17]. One possible explanation is that most synthetic lethalities are complex polygenic interactions that are modified in different cellular contexts and by other endogenous factors [4]. We hypothesized that disentangling complex polygenic synthetic lethalities could be achieved by correlating pairwise synthetic lethalities with endogenous factors across many cell models. However, a barrier to this approach remains in the practical limitations of screening large multiplexed CRISPR libraries across many cell lines.

Since screening all paralog pairs across many cell lines would be technically challenging, we instead focused on screening the 5,116 paralog pairs with solved structures, as these are typically well-studied proteins with defined functional domains. To further streamline screening, a minimal CRISPR-Cas12 library was designed to interrogate dKO effects with four guides per construct and two guides per paralog (n = 10,886 constructs). The dKO effects of this compact library were then compared with the additive effects of each single KO

in the Cancer Dependency Map, which enabled efficient screening across 20 cancer models. An ROC-AUC analysis revealed that synthetic lethal interactions could be accurately detected (AUC = 0.77) by comparing the normalized observed effects of the double KO (CRISPR-Cas12) with the normalized expected effects of the single KOs in the Cancer Dependency Map (**Figure 3A**). A total of 780 synthetic lethal interactions were detected in at least one cell line out of 5,318 paralog pairs that were tested (14.6%), with wide variability in the number of synthetic lethalities per cell line (n = 17 - 230; median = 115) (**Table S5**). These synthetic lethal interactions could be classified as common (n = 63; >33% incidence), selective (n = 301; 33% to 10% incidence), and rare (n = 446; <10% incidence). The histone deacetylases, HDAC1 and HDAC2, were the most commonly detected synthetic lethal interactions in 18 out of 20 cancer models (**Figure 3B**), agreeing with multiple other reports [8,18,19]. Likewise, 17 of the top 30 synthetic lethalities (represented in >10 cancer models) have been reported previously, including SMARCA2/SMARCA4 [19,20], VPS4A/VPS4B [14,21], and P300/CBP [22]. Finally, this approach could also distinguish synthetic lethal interactions within family members, such as the synthetic lethal interactions of HSPA8 with HSPA1A, HSPA1B or HSPA2, but not other family members, as was detected in the screen of all human paralogs in NCIH1299 and MDAMB231 (**Figure 3C**). Collectively, these data demonstrated that this hybrid approach was robust in detecting synthetic lethal interactions.

A second digenic paralog study recently assessed the synthetic lethal interactions of 5,065 paralog pairs using combinatorial CRISPR/Cas9 screening, which revealed 686 synthetic lethal interactions (13.5% of tested) [8] and provided an opportunity to expand the cohort of cell models tested per paralog pair. Across both studies, a total of 1,278 pairs were tested in 30 unique cancer models, of which 241 paralog pairs were predicted to be synthetic lethal (**Table S6**) and the most common synthetic lethal pairs remained unchanged (**Figure 3D**, **E**). To identify the cellular modifiers of the digenic synthetic lethal paralog pairs, a LASSO regression analysis was performed using gene expression, copy number variants, and somatic mutations as input variables. A pathway enrichment analysis was then performed for the synthetic lethal pairs that were represented in at least four cancer models and with at least five predictive features that were selected by the LASSO analysis (n = 47 pairs) (**Table S7**). Seven of these synthetic lethal paralog interactions could be associated with enriched pathways at a nominal cutoff (p<0.001), including CSNK2A1/CSNK2A2, SMARCA2/SMARCA4, CDK4/CDK6, HDHD2/LHPP, CARM1/PRMT6, PIK3CA/PIK3CG, and PIK3R1/PIK3R2 (**Table S7**). For example, the sensitivity to synthetic lethal interactions between loss of CSNK2A1 and CSNK2A2 was predicted by features enriched in the FGFR signaling pathway, which has been linked previously to CK2 signaling [23]. The most predictive features of CDK4 and CDK6 synthetic lethality were associated with regulation of EIF4F activity, which is required for protein translation to support cell cycle [24]. Finally,

PIK3CA mutations were predictive of the synthetic lethal interactions between PIK3CA/PIK3CG and PIK3R1/PIK3R2, agreeing with prior reports that conditional pathway activation modifies synthetic lethal interactions within the pathway [25]. Combined, these data demonstrated that endogenous cellular context modified the sensitivity to synthetic lethal interactions, which can be mapped across a panel of cancer models.

Predicting paralog synthetic lethality

Although more than 60% of human genes are paralogous, only ~10% have been empirically tested for synthetic lethal interactions prior to this study [3]. As such, computational approaches for predicting synthetic lethality have been limited by the lack of a robust and unbiased set of experimentally validated synthetic lethalities for training and evaluation of predictive models. To our knowledge, this study represents the largest set of paralog screens to date, which we hypothesized could be used to train an ensemble classifier of paralog synthetic lethality. To account for the variability in synthetic lethal interactions across different cellular contexts, weighted sets of true positives (n = 223 pairs per set) based on the overall representation across six distinct CRISPR screening datasets. These types of unbalanced positive-only learning tasks can be addressed by using randomly-sampled negative training sets of equivalent size to the positive set. Accordingly, size-matched negative sets were randomly sampled without replacement from the remaining set of paralog pairs without observed lethalities and a random forest model was built using the averaged predictions over 10,000 models that incorporated 25 sets of features. The feature sets spanned three broad categories (sequence, neighborhood, and expression) that were predefined to capture unique aspects of synthetic lethal biology [3], as well as an additional fourth category that incorporated protein structural similarity from AlphaFold [26]. The resulting averaged model was then used to assign the probability of a synthetic lethal interaction between a paralog pair.

As shown in **Figure 4A-B**, the probability of synthetic lethal interactions were correlated with the prevalence of the observed synthetic lethality, demonstrating that the random forest model is able to differentiate common, selective and rare synthetic lethal interactions. The average decrease in the Gini metric was used to assess the relative importance of the features in the models, which revealed that extend of protein-protein interactions (PPI) of the paralog pairs and the essentiality of the interaction network were strong predictors of synthetic lethality (**Figure 4C**). The average expression of either paralog in normal tissues was also highly predictive of paralog synthetic lethality (**Figure 4C**). The SHapley Additive exPlanations (SHAP) scores were also calculated to determine the relative contribution of the features to individual paralog pairs (**Table S8**), which confirmed the importance of PPI networks and normal tissue expression in predicting synthetic lethality (**Figure 4D-F**). Moreover, comparing the SHAP scores across the classifications of common, selective, and rare synthetic lethal interactions revealed that the size and essentiality of the network of PPI

shared by the paralog pairs made the largest contributions to whether a synthetic lethality was common across cancer models (**Figure 4G**, **H**). To test this hypothesis further, the average essentialities of the 780 synthetic lethal interactions that were classified as common (n = 63; >33% incidence), selective (n = 301; 33% to 10% incidence), and rare (n = 446; <10% incidence), based on experimental observations across the 20 cancer models. This revealed that the shared PPI network of common synthetic lethal paralog pairs had significantly more essential genes on average (n = 19 per pair) than selective synthetic lethalities (n = 6 per pair) or rarely synthetic lethal pairs (n = 3 per pair) (**Figure 4I**). Several known broadly acting synthetic lethal paralog pairs were ranked in the top most essential PPI networks among common synthetic lethalities (**Figure 4J**), as were multiple synthetic lethal paralog pairs in Hsp70 family, including HSPA2/HSPA8 (**Figure 4K**). Taken together, these data demonstrated that the essentiality of a PPI network that is shared by a paralog pair is a major determinant of the commonality of the synthetic lethality.

DISCUSSION

This study addressed the significant challenge of empirically testing the genetic interactions of all known paralog pairs in the human genome, which has been limited by the experimental capacity of existing technologies for combinatorial genetic perturbation screens [4]. To overcome these limitations, we developed a CRISPR-Cas12 screening approach, which performed well in detecting synthetic lethal and buffering interactions in two distinct cancer models. This study revealed that synthetic lethal interactions are relatively rare on a per tumor basis (~0.4% of all paralogs). In addition to rarity, a meta-analysis of a subset of 1,278 paralog pairs in 30 cancer models confirmed that the penetrance of synthetic lethal interactions is variable, and in some cases, can be predicted by polygenic interactions with endogenous factors. Acknowledging that the ability to predict digenic interactions is likely limited by the variability across cancer models, we built a machine learning classifier that utilized a true positive set that was weighted by the penetrance of synthetic lethal interactions across cancer models. The weighted machine learning classifier revealed that the penetrance of a synthetic lethal pair is modified by the breadth and essentiality of the mutual PPI network that interacts with a paralog pair, which is both a novel and intuitive finding. Thus, taken together, this study provides a comprehensive analysis of all digenic interactions in the human genome, as well as the key features that underlie the heterogeneity in synthetic lethalities that have been reported here and elsewhere.

Pairwise interactions of paralogs are heterogenous and relatively rare

Multiple studies have demonstrated a wide variability in the penetrance of digenic synthetic lethal interactions using focused CRISPR screens [8,14,17,19,27,28], yet interpreting these data have been difficult due to the limited overlap in the paralog sets and technologies used in each study [29]. Here, we screened all 36,648 paralog pairs in the human genome using a single gene-editing technology (enCas12a), which confirmed unambiguously that a minority of synthetic lethalities (~10%) are shared by two cancer models. Fitting with recent observations that few synthetic lethal interactions are common in all contexts [29], the expanded meta-analysis of 30 cancer models revealed no synthetic lethalities that were completely penetrant, with HDAC1/2 reaching the highest penetrance of 77%. We hypothesized that the penetrance of synthetic lethal interactions was modified by cellular context, which was supported by a multivariable regression analysis that identified the predictive features of some digenic synthetic lethalities. Notably, the most predictive features typically fell within the same pathway or related cellular function, suggesting that these synthetic lethalities are likely complex polygenic interactions that are context dependent. These data agree with several reports of digenic synthetic lethality that were dependent on driver mutations within the pathway [8,25], as well as the recently reported IN4MER approach that identified several trigenic synthetic lethalities [29]. Taken together, the data presented in our study confirmed and extended the examples of polygenic synthetic lethalities through complex interactions with endogenous factors in different cancer models.

Another key finding of this study was that most digenic paralog synthetic lethalities are relatively rare when compared to monogenic dependencies. Approximately 0.4% of all 36,648 paralog pairs in the human genome were synthetic lethal in MDAMB231 or NCIH1299 cells, whereas ~3% of the tested singletons were essential in either model. Even fewer paralog pairs were synthetic lethal in both cancer models (0.06%), while the majority of essential singletons shared by both models, and were defined as “common essential” by the DEPMAP. To our knowledge, the meta-analysis of the 30 cancer models represents the largest combined screen of a subset of paralog pairs to date, providing an opportunity to more broadly assess the penetrance of synthetic lethality across different cellular contexts. Notably, this meta-analysis confirmed that the vast majority of digenic synthetic lethalities were rarely detected in multiple cancer models, which could be explained by complex polygenic interactions that vary based on cellular context and endogenous features [4]. Alternatively, these observations might also suggest tumors evolve paralog synthetic lethality profiles based on lineage and genetic drivers, as has been proposed recently [30].

In addition to synthetic lethal interactions, the 36,648 paralog screen also captured positive interactions (i.e., synthetic buffering), whereby the loss of two paralogs drives a net positive advantage in cancer cells. Similar to a previous report [14], we found that the synthetic buffering paralog pairs were extremely rare when compared to synthetic lethal interactions and overlapped strongly with tumor suppressor genes (TSGs). Combined with the observation that paralogs are more likely to be homozygously deleted with TSGs [30], these data suggest that synthetic buffering by paralogs is likely a mechanism that enables tumorigenicity. Moreover, based on the frequency of homozygous deletions in TCGA patients [31], it is likely that the rarity of synthetic buffering is driven by incidence homozygous deletions of TSGs.

Predicting digenic synthetic lethality of paralogs

Multiple attempts have been made to predict paralog synthetic lethality [3], yet these efforts have been limited by an incomplete set of validated synthetic lethal interactions for the training and evaluation of machine learning models. De Kegel et al [3] have recently reported what was arguably the most comprehensive study of paralog synthetic lethality to date, prompting us to adopt a similar approach with several key changes. Firstly, rather than focusing the training sets (both positive and negative) on observations made from genetic interactions within the DepMap [3], the positive training set used in our approach was defined by empirical observations and the negative set was represented as a random sampling of all other paralog pairs. Thus, the machine learning classifier used in the present study is distinct in that the training sets were not reliant on a limited representation of synthetic lethalities in the DepMap [7]. Secondly, our machine learning classifier was unique in that it used a weighted positive-only training set, which accounts for the widely variable penetrance of paralog synthetic lethalities that was reported here and elsewhere [8,14,17,19,27,28]. To our knowledge, this new classifier revealed for the first time that the overlap and essentiality of

the PPI network that is shared by a paralog pair are the best predictors of the most penetrant synthetic lethalities.

Moving beyond digenic synthetic lethal interactions

The incomplete penetrance of synthetic lethal interactions suggests that most synthetic lethal interactions are not digenic and will require further exploration to understand the molecular underpinnings of synthetic lethality [4]. One approach used in this study was to test the association of a subset of 1,278 paralog pairs across 30 cancer models, which to our knowledge was the largest cohort of cancer models to be used for a meta-analysis of synthetic lethal interactions to date. Notably, this approach revealed multiple digenic synthetic lethalities that were associated with endogenous factors, suggesting the possibility that larger screens of digenic interactions will unveil additional endogenous modifiers of polygenic synthetic lethalities in the future. An empirical approach to testing and validating polygenic synthetic lethalities has recently been demonstrated using multiplexed CRISPR-Cas12, in which families of up to four paralogs could be empirically tested [29]. However, although this exciting new approach offers to expand the repertoire of polygenic synthetic lethal interactions, there remains practical limitations in the feasibility of testing all polygenic interactions [4]. Thus, to expand beyond digenic synthetic lethalities, it is likely that a combination of approaches will be required, including the observation of polygenic interactions from digenic screens and empirical validation using multiplexed screening approaches to test two or more paralogs simultaneously.

Conclusions

The concept of synthetic lethal paralogs has existed for nearly a century [6], yet most human paralogs had yet to be empirically tested. Here, we presented the first screening approach to test synthetic lethalities across all 36,648 paralog pairs in the human genome. Combined with focused paralog synthetic lethality studies across 30 cancer models, this work identified >800 synthetic lethalities with varied penetrance across different cancer models, which in some cases could be related to endogenous perturbations of related pathways. Based on the observation that the penetrance of digenic synthetic lethal paralogs varies widely and are likely polygenic in most cases (i.e., three or more interactions), we built a new machine learning classifier that weighted the most penetrant synthetic lethal paralog pairs. Thus, this new classifier is better suited to modeling the predictive features of truly digenic paralogs, which subsequently revealed that highly penetrant digenic synthetic lethal interactions were predicted by the essentiality of the mutual PPI network of the paralog pair. This observation indicates that machine learning approaches could be further refined to predict polygenic synthetic lethalities. To this end, future work is needed to expand the catalog of polygenic synthetic lethalities beyond the few examples that have been reported in the literature [4].

MATERIALS AND METHODS

Multiplexed CRISPR/Cas12 screening of all human paralogs

Guides were designed for the enAsCas12a enzyme using the CRISPick (https://portals.broadinstitute.org/gppx/crispick/public) and synthesized into 4-guide arrays with direct repeats (DR)-0, -1, -2, and -3 preceding each guide, followed by cloning into a guide-only lentiviral vector (pRDA_052), as described previously [9,32]. Two double knockout constructs consisting of 2 guides x 2 genes (n = 4 guides total per construct) were designed for each pair of paralogs. Two single knockout constructs were also designed for each paralog pair: 2 guides x 1 gene + 2 non-targeting guides (n= 4 guides total per construct). A total of 500 constructs with 4 non-targeting guides were also included in the library as negative controls. The final library to test all paralog pairs in the human genome consisted of 113,502 constructs (n = 13,128 singletons and n = 36,648 pairs). Screens were performed in triplicate using NCI-H1299 and MDA-MB-231 cells that stably expressed enAsCas12a, as described previously [9]. Cells were infected at a multiplicity of infection (MOI) = 0.3 and cultured for 14 days while continuously maintaining 500X coverage, followed by DNA extraction and PCR-barcoding using the p5 Agon and p7 Kermit primers [9]. The PCR-barcoded libraries were single-end sequenced using an Illumina HiSeq4000 (300X cycle), followed by demultiplexing of sequencing reads (bcl2fastq, Illumina) and quantification of guide array abundance across all samples was done with a custom Perl script. Sequences between the flanking sequences or by location were extracted and compared to a database of sgRNA for each library. Only perfectly matched sgRNA sequences were kept and used in the generation of count matrix. Normalization between all samples was done using the “TMM” method [33] implemented in the edgeR R Bioconductor package. Log2 fold-changes (L2FC) of guide array abundance were calculated by comparing day 14 libraries with the plasmid library using limma-voom [34]. Genetic interactions (GI) were calculated by comparing the expected and observed L2FC of double and single knockout constructs following z-score transformation (zL2FC) [17]. Paralog pairs were filtered for expression (>2.5 TPM) and an observed activity of the median double knockout effect of <-1 zL2FC and >1 zL2FC for synthetic lethal and synthetic buffering pairs, respectively. A paralog pair with a delta zL2FC (dzL2FC) of >1.64 (90% confidence interval) between the observed and predicted effects was considered significant.

Hybrid screening approach for testing a subset human paralogs

To test a subset of 5,116 paralog pairs by dKO, a CRISPR-Cas12 library design utilized arrays of four guides per construct and two guides per paralog (n = 10,886 constructs). Two double knockout constructs consisting of 2 guides x 2 genes (n = 4 guides total per construct) were designed for each pair of paralogs. Library synthesis, cloning, and screening was performed as described above using the following enAsCas12a transgenic cancer models:

NCIH1299, A549, MDAMB231, LS180, RKO, PANC1, MKN74, T84, SNU1, MIAPACA2, DETROIT562, HSC2, HT29, SW837, SW1463, LS513, NCIH2087, NCIH520, COLO678, and LK2. Cells were infected at a multiplicity of infection (MOI) = 0.3 and cultured for 14 days while continuously maintaining 500X coverage, followed by DNA extraction, PCR-barcoding, and sequencing, as described above. Sequences between the flanking sequences or by location were extracted and compared to a database of sgRNA for each library. As described above, perfectly matched sgRNA sequences were used in the generation of the count matrix. Normalization between all samples was done using the "TMM" method [33] implemented in the edgeR R Bioconductor package and the L2FC of guide array abundance were calculated by comparing day 14 libraries with the plasmid library using limma-voom [34]. To estimate single gene knockout effects, sgRNA count matrices for each cancer model were directly downloaded from the DepMap (AvanaRawReadcounts.csv) and L2FC changes of single gene effects were calculated as described above. The GI were calculated by comparing the expected and observed L2FC of double and single knockout constructs following z-score transformations. A paralog pair with a dzL2FC > 2 (95% confidence interval) between the observed and predicted effects was considered significant. A receiver operating characteristic (ROC) curve was used to assess the efficacy of the hybrid screening approach using CRISPR-Cas12 to measure double KO effects and the DEPMAP to measure single gene KO effects. The R pROC package (version 1.18.5) was used to plot the ROC curve and compute the AUC value.

Meta-analysis of 1,278 paralog pairs in 30 cancer models

Ito et al [8] recently tested 5,065 paralog pairs using combinatorial CRISPR-Cas9 screening across 11 cancer models, which revealed 686 synthetic lethal interactions (13.5% of tested). When combined with the present study, a total of 1,278 pairs in 30 unique cancer models remained for a multivariable regression analysis of endogenous modifiers of digenic synthetic lethality. The L2FC data were directly downloaded from Ito et al [8] and z-transformed to yield zL2FC values. The GI scores were calculated, as described above, by comparing the expected and observed L2FC of double and single knockout constructs following z-score transformations. A paralog pair with a dzL2FC > 2 (95% confidence interval) between the observed and predicted effects was considered significant. The gene expression, mutations and copy number variants data were downloaded for each cancer model (n = 30) from the DEPMAP (release 21Q1). The gene expression data were filtered for a variance of greater than 0.5 across the cancer models (n = 11,412 genes). The mutation data were filtered for a frequency of greater than 2% across the cancer models (n = 9,157 mutations). The copy number variant data were filtered to identify deletions with a relative ploidy of less than 1.2 and a frequency of greater than 5% (n = 6,757 deletions). The 'glmnet' package (version

4.1.3) was used to build a LASSO model between synthetic lethal interactions and combined features from gene expression, mutations and deletions. The models were trained using a 10-fold cross-validation using cv.glmnet function. The lambda value was determined by “lambda.min” with alpha = 1. The paralog pairs with summed coefficients less than 0.01 were filtered from the analysis.

Machine Learning Predictions of Synthetic Lethal Paralogs

The random forest models and predictions were performed using the “randomForest (v4.6-3)” and “caret (v6.0-93)” libraries in R statistical software. The number of trees in each model was fixed at 1000, as any increase in trees above this value did not yield any significant changes in the prediction probability. The tuneRF function in the random forest package was used to determine the optimal number of variables (“mtry”) sampled at each split. The stepfactor was set at 0.01 and the “improve” parameter at 0.01 value. A weighted positive set was randomly sampled without replacement (n = 223 per set) to match the penetrance of the 825 synthetic lethal paralogs observed in this study. The weight was defined as the number of cell lines where the synthetic lethality was observed divided by the total number of cell lines tested. To balance the training set for the model, size-matched negative training sets were generated by random sampling without replacement of genes and excluding the positive training set. A total of 10,000 random forest models using each of the random negative sets were trained and predictions were made based on each model, followed by averaging the predictions over 10,000 models to assign the synthetic lethality score to each paralog pair. There were 25 features which included 23 from the DeKegel et al. study and three additional features. Two features were the correlation values of the paralog pairs based on the DepMap CRISPR screen values and the co-expression from the TCGA. An additional feature was included by calculating the structure similarity of the pairs based on the AlphaFold 3D structure prediction values. The structure similarity was determined based on the root-mean-square deviation of atomic positions (RMSD) of the superimposed protein structures. The RMSD was calculated using the bio3D (2.4-4) R package. For feature importance both the Gini metric and the SHAP values were calculated. The Gini values were calculated using the R randomForest packaged and averaged over the 10,000 model to determine the final contribution. The SHAP scores were generated using the shapper R package. The contribution of features to each individual paralog pair was averaged over 1,000 models.

DECLARATIONS

Competing Interest Statement

The design, study conduct, and financial support for all other research were provided by AbbVie. AbbVie participated in the interpretation of data, review, and approval of the publication. All authors were employees of AbbVie at the time of the study. M.J.F. is currently a full time employee of Pfizer.

Author Contributions

M.J.F., Z.D., C.G., and D.V. conceptualized the study. M.J.F, D.V., and S.P. performed the experiments. M.J.F., Z.D., X.S., C.G., D.V., S.P., and C.L. analyzed and interpreted data. M.J.F. and Z.D. wrote the original manuscript draft. All authors reviewed and edited the manuscript draft.

Acknowledgements

We thank AbbVie employees Aparna Vasanthakumar, Archana Iver, Kari Barlan, and Kevin Zhao for the helpful discussions and critical feedback. We also thank Erin Murphy, Tifani Anton, and the other members of the Genome Technology team in the AbbVie Genome Research Center for their expertise and technical support.

REFERENCES

1. Koonin EV. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet [Internet]. 2005 [cited 2024 Mar 16];39. Available from: https://pubmed.ncbi.nlm.nih.gov/16285863/

2. Ryan CJ, Mehta I, Kebabci N, Adams DJ. Targeting synthetic lethal paralogs in cancer. Trends Cancer Res [Internet]. 2023 [cited 2024 Mar 16];9. Available from: https://pubmed.ncbi.nlm.nih.gov/36890003/

3. De Kegel B, Quinn N, Thompson NA, Adams DJ, Ryan CJ. Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. cels. 2021;12:1144–59.e6.

4. Ryan CJ, Devakumar LPS, Pettitt SJ, Lord CJ. Complex synthetic lethality in cancer. Nat Genet. 2023;55:2039–48.

5. Huang A, Garraway LA, Ashworth A, Weber B. Synthetic lethality as an engine for cancer drug target discovery. Nat Rev Drug Discov. 2020;19:23–38.

6. Bridges CB. The Origin of Variations in Sexual and Sex-Limited Characters. Am Nat [Internet]. 1922 [cited 2024 Mar 16]; Available from: https://www.journals.uchicago.edu/doi/10.1086/279847

7. Boehm JS, Garnett MJ, Adams DJ, Francies HE, Golub TR, Hahn WC, et al. Cancer research needs a better map. Nature. 2021;589:514–6.

8. Ito T, Young MJ, Li R, Jain S, Wernitznig A, Krill-Burger JM, et al. Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. Nat Genet. 2021;53:1664–72.

9. DeWeirdt PC, Sanson KR, Sangree AK, Hegde M, Hanna RE, Feeley MN, et al. Optimization of AsCas12a for combinatorial genetic screens in human cells. Nat Biotechnol [Internet]. 2021 [cited 2024 Feb 25];39. Available from: https://pubmed.ncbi.nlm.nih.gov/32661438/

10. Frontini M, Soutoglou E, Argentini M, Bole-Feysot C, Jost B, Scheer E, et al. TAF9b (Formerly TAF9L) Is a Bona Fide TAF That Has Unique and Overlapping Roles with TAF9. Mol Cell Biol. 2005;25:4638.

11. Kabani M, Martineau CN. Multiple hsp70 isoforms in the eukaryotic cytosol: mere redundancy or functional specificity? Curr Genomics [Internet]. 2008 [cited 2024 Mar 9];9. Available from: https://pubmed.ncbi.nlm.nih.gov/19471609/

12. Brocchieri L, Conway de Macario E, Macario AJL. hsp70 genes in the human genome: Conservation and differentiation patterns predict a wide array of overlapping and specialized functions. BMC Evol Biol. 2008;8:19.

13. Kampinga HH, Craig EA. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol. 2010;11:579–92.

14. Parrish PCR, Thomas JD, Gabel AM, Kamlapurkar S, Bradley RK, Berger AH. Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. Cell Rep. 2021;36:109597.

15. Gundogdu R, Hergovich A. MOB (Mps one Binder) Proteins in the Hippo Pathway and Cancer. Cells [Internet]. 2019 [cited 2024 May 21];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6627106/

16. Anvar NE, Lin C, Ma X, Wilson LL, Steger R, Sangree AK, et al. Efficient gene knockout and genetic interaction screening using the in4mer CRISPR/Cas12a multiplex knockout platform. Nat Commun [Internet]. 2024 [cited 2024 Jun 29];15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11055879/

17. Shi X, Gekas C, Verduzco D, Petiwala S, Jeffries C, Lu C, et al. TCGADEPMAP – Mapping Translational Dependencies and Synthetic Lethalities within The Cancer Genome Atlas [Internet]. bioRxiv. 2022 [cited 2024 Feb 24]. p. 2022.03.24.485544. Available from: https://www.biorxiv.org/content/10.1101/2022.03.24.485544v2.abstract

18. Zhang Y, Remillard D, Onubogu U, Karakyriakou B, Asiaban JN, Ramos AR, et al. Collateral lethality between HDAC1 and HDAC2 exploits cancer-specific NuRD complex vulnerabilities. Nat Struct Mol Biol. 2023;30:1160–71.

19. Dede M, McLaughlin M, Kim E, Hart T. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. Genome Biol. 2020;21:1–19.

20. Hoffman GR, Rahal R, Buxton F, Xiang K, McAllister G, Frias E, et al. Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc Natl Acad Sci U S A. 2014;111:3128–33.

21. Neggers JE, Paolella BR, Asfaw A, Rothberg MV, Skipper TA, Yang A, et al. Synthetic Lethal Interaction between the ESCRT Paralog Enzymes VPS4A and VPS4B in Cancers Harboring Loss of Chromosome 18q or 16q. Cell Rep [Internet]. 2020 [cited 2024 Mar 11];33. Available from: http://www.cell.com/article/S2211124720314820/abstract

22. Ogiwara H, Sasaki M, Mitachi T, Oike T, Higuchi S, Tominaga Y, et al. Targeting p300 Addiction in CBP-Deficient Cancers Causes Synthetic Lethality by Apoptotic Cell Death due to Abrogation of MYC Expression. Cancer Discov [Internet]. 2016 [cited 2024 Mar 11];6. Available from: https://pubmed.ncbi.nlm.nih.gov/26603525/

23. Intracellular partners of fibroblast growth factors 1 and 2 - implications for functions. Cytokine Growth Factor Rev. 2021;57:93–111.

24. Kong T, Xue Y, Cencic R, Zhu X, Monast A, Fu Z, et al. eIF4A Inhibitors Suppress Cell-Cycle Feedback Response and Acquired Resistance to CDK4/6 Inhibition in Cancer. Mol Cancer Ther. 2019;18:2158–70.

25. Chang L, Jung NY, Atari A, Rodriguez DJ, Kesar D, Song T-Y, et al. Systematic profiling of conditional pathway activation identifies context-dependent synthetic lethalities. Nat Genet. 2023;55:1709–20.

26. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

27. Thompson NA, Ranzani M, van der Weyden L, Iyer V, Offord V, Droop A, et al. Combinatorial CRISPR screen identifies fitness effects of gene paralogues. Nat Commun. 2021;12:1302.

28. Gonatopoulos-Pournatzis T, Aregger M, Brown KR, Farhangmehr S, Braunschweig U, Ward HN, et al. Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9-Cas12a platform. Nat Biotechnol. 2020;38:638–48.

29. Anvar NE, Lin C, Ma X, Wilson LL, Steger R, Sangree AK, et al. Efficient gene knockout and genetic interactions: the IN4MER CRISPR/Cas12a multiplex knockout platform [Internet]. bioRxiv. 2023 [cited 2024 Mar 24]. p. 2023.01.03.522655. Available from:

https://www.biorxiv.org/content/10.1101/2023.01.03.522655v2.abstract

30. De Kegel B, Ryan CJ. Paralog dispensability shapes homozygous deletion patterns in tumor genomes. Mol Syst Biol [Internet]. 2023 [cited 2024 Mar 27];19. Available from: https://pubmed.ncbi.nlm.nih.gov/37963083/

31. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173:321–37.e10.

32. Petiwala S, Modi A, Anton T, Murphy E, Kadri S, Hu H, et al. Optimization of Genomewide CRISPR Screens Using AsCas12a and Multi-Guide Arrays. The CRISPR journal [Internet]. 2023 [cited 2024 Feb 29];6. Available from: https://pubmed.ncbi.nlm.nih.gov/36787117/

33. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:1–9.

34. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15:1–17.

FIGURE LEGENDS

Figure 1. Multiplexed CRISPR/Cas12 screening of synthetic lethal interactions across all 36,648 paralog pairs in the human genome. (**A**) Schematic of the CRISPR/Cas12 library multiplexed guide arrays targeting one or two genes per array. (**B**) Schematic of the synthetic lethality screening approach using the CRISPR/Cas12 library. (**C**, **D**) Library performance was assessed by comparing the effects of single gene KOs in the CRISPR-Cas12 library with the DepMap essentiality profiles of NCI-H1299 (**C**) and MDA-MB-231 (**D**). (**E**) A venn diagram of the essential singletons and synthetic lethal paralog pairs in NCIH1299 and MDAMB231 cells. (**F**, **G**) Manhattan plots of the dzL2FC values of 36,648 paralog pairs in NCI-H1299 (**F**) and MDA-MB-231 (**G**). (**H**) An example of a previously unreported pair of common synthetic lethal paralogs, TAF9 and TAF9b. $^{**}P < 0.01$ and $^{*}P < 0.05$, as determined by unpaired Student's t-test. (**I**) An example of a previously unreported family (Hsp70) with a subset of synthetic lethal paralogs. $^{***}P < 0.001$ and $^{*}P < 0.05$, as determined by two-way ANOVA with Dunnett's test for multiple comparisons.

Figure 2. Multiplexed CRISPR/Cas12 screening of synthetic buffering interactions across all 36,648 paralog pairs in the human genome. (**A**) Schematic of the synthetic lethality screening approach using the CRISPR/Cas12 library. (**B**) A heatmap of synthetic buffering interactions in NCIH1299 and MDAMB231 cells. (**C**-**H**) STRING network and pathway enrichment analysis of synthetic buffering pairs, MOB1/MOB1B (MOB4) (**C**, **D**), ARHGAP12/ARHGAP21 (**E**, **F**), and EXOSC7/EXOSC8 (**G**, **H**). Note the connections with established TSGs and related pathways.

Figure 3. Meta-analysis of paralog synthetic lethalities across cancer models. (**A**) An ROC-AUC analysis revealed that synthetic lethal interactions could be accurately measured using a hybrid approach that compared the normalized observed doubled KO effects using multiplexed CRISPR-Cas12 with the normalized expected effects of the single KOs in the DEPMAP. (**B**) The hybrid screening approach recapitulated the high penetrance of HDAC1/HDAC2 synthetic lethality across 20 cancer models tested. (**C**) The hybrid screening approach also accurately detected synthetic lethalities within paralog families, such as the HSP70 family, across the 20 cancer models. (**D**) A meta-analysis of 1,278 pairs tested across 30 unique cancer models detected prevalent synthetic lethal interactions, many of which have been reported elsewhere. (**E**) A heatmap of the most prevalent synthetic lethal interactions across the 30 unique cancer models.

Figure 4. A weighted random forest classifier for predicting digenic synthetic lethal paralogs. (**A**) The distribution of synthetic lethality probability scores across the 36,648 paralog pairs in the human genome. (**B**) A violin plot showing the distribution of probability scores across synthetic lethal paralog pairs that were classified as common (>33% prevalence), selective (33%-10% prevalence), rare (<10% prevalence), or never detected. (**C**) The mean decrease in the Gini metric was used to assess the relative importance of the features in the model.

(**D**-**F**) The distribution of SHAP scores across paralog synthetic lethalities that were classified as common, selective, or rare. (**G**, **H**) The comparison of SHAP scores across classes revealed that the size (**G**) and essentiality (**H**) of the PPI network shared by a paralog pair are the major contributors to whether a synthetic lethality was common across cancer models. ***$P < 0.001$, as determined by one-way ANOVA with Tukey's multiple comparison test. (**I**) Average number of essential genes in the shared PPI network of synthetic lethal interactions classified as common (n = 63; >33% incidence), selective (n = 301; 33% to 10% incidence), and rare (n = 446; <10% incidence). Data are presented as mean +/- SEM. ***$P < 0.001$, as determined by one-way ANOVA with Dunnett's multiple comparison test. (**J**) The distribution of essential and non-essential genes in the shared PPI network of the top common synthetic lethal paralog pairs. (**K**) Plot of the PPI network shared by the synthetic lethal paralog pairs, HSPA2 and HSPA8. Essential genes (DepMap Chronos Score < -1) are indicated by blue shading.

Figure 1

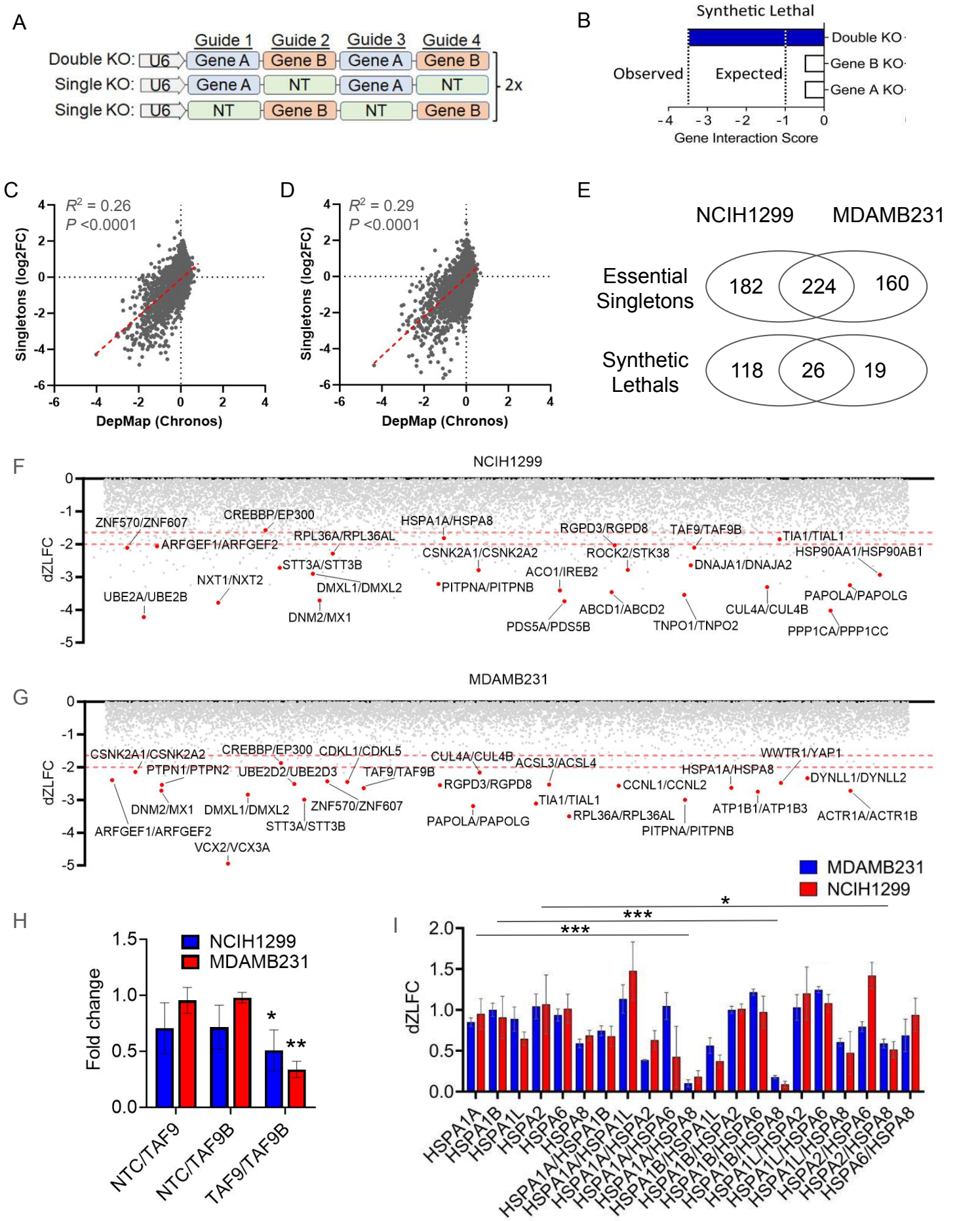

A
Guide 1
Guide 2
Guide 3
Guide 4
Double KO: U6 Gene A Gene B Gene A Gene B
Single KO: U6 Gene A NT Gene A NT
Single KO: U6 NT Gene B NT Gene B
2x
B
Synthetic Lethal
Double KO
Gene B KO
Gene A KO
Observed
Expected
Gene Interaction Score
C
R^2 = 0.26
P <0.0001
Singletons (log2FC)
DepMap (Chronos)
D
R^2 = 0.29
P <0.0001
Singletons (log2FC)
DepMap (Chronos)
E
NCIH1299
MDAMB231
Essential Singletons
182
224
160
Synthetic Lethals
118
26
19
F
NCIH1299
dZLFC
ZNF570/ZNF607
CREBBP/EP300
HSPA1A/HSPA8
RGPD3/RGPD8
TAF9/TAF9B
TIA1/TIAL1
ARFGEF1/ARFGEF2
RPL36A/RPL36AL
CSNK2A1/CSNK2A2
ROCK2/STK38
HSP90AA1/HSP90AB1
STT3A/STT3B
DNAJA1/DNAJA2
NXT1/NXT2
ACO1/IREB2
DMXL1/DMXL2
PITPNA/PITPNB
UBE2A/UBE2B
PAPOLA/PAPOLG
ABCD1/ABCD2
CUL4A/CUL4B
DNM2/MX1
PDS5A/PDS5B
TNPO1/TNPO2
PPP1CA/PPP1CC
G
MDAMB231
dZLFC
CSNK2A1/CSNK2A2
CREBBP/EP300
CDKL1/CDKL5
CUL4A/CUL4B
ACSL3/ACSL4
WWTR1/YAP1
PTPN1/PTPN2
UBE2D2/UBE2D3
TAF9/TAF9B
HSPA1A/HSPA8
DYNLL1/DYNLL2
RGPD3/RGPD8
CCNL1/CCNL2
TIA1/TIAL1
DNM2/MX1
DMXL1/DMXL2
ZNF570/ZNF607
ATP1B1/ATP1B3
PAPOLA/PAPOLG
RPL36A/RPL36AL
ACTR1A/ACTR1B
ARFGEF1/ARFGEF2
STT3A/STT3B
PITPNA/PITPNB
VCX2/VCX3A
MDAMB231
NCIH1299
H
NCIH1299
MDAMB231
Fold change
NTC/TAF9
NTC/TAF9B
TAF9/TAF9B
I
dZLFC
HSPA1A
HSPA1B
HSPA1L
HSPA2
HSPA6
HSPA8
HSPA1A/HSPA1B
HSPA1A/HSPA1L
HSPA1A/HSPA2
HSPA1A/HSPA6
HSPA1A/HSPA8
HSPA1B/HSPA1L
HSPA1B/HSPA2
HSPA1B/HSPA6
HSPA1B/HSPA8
HSPA1L/HSPA2
HSPA1L/HSPA6
HSPA1L/HSPA8
HSPA2/HSPA6
HSPA2/HSPA8
HSPA6/HSPA8

Figure 2

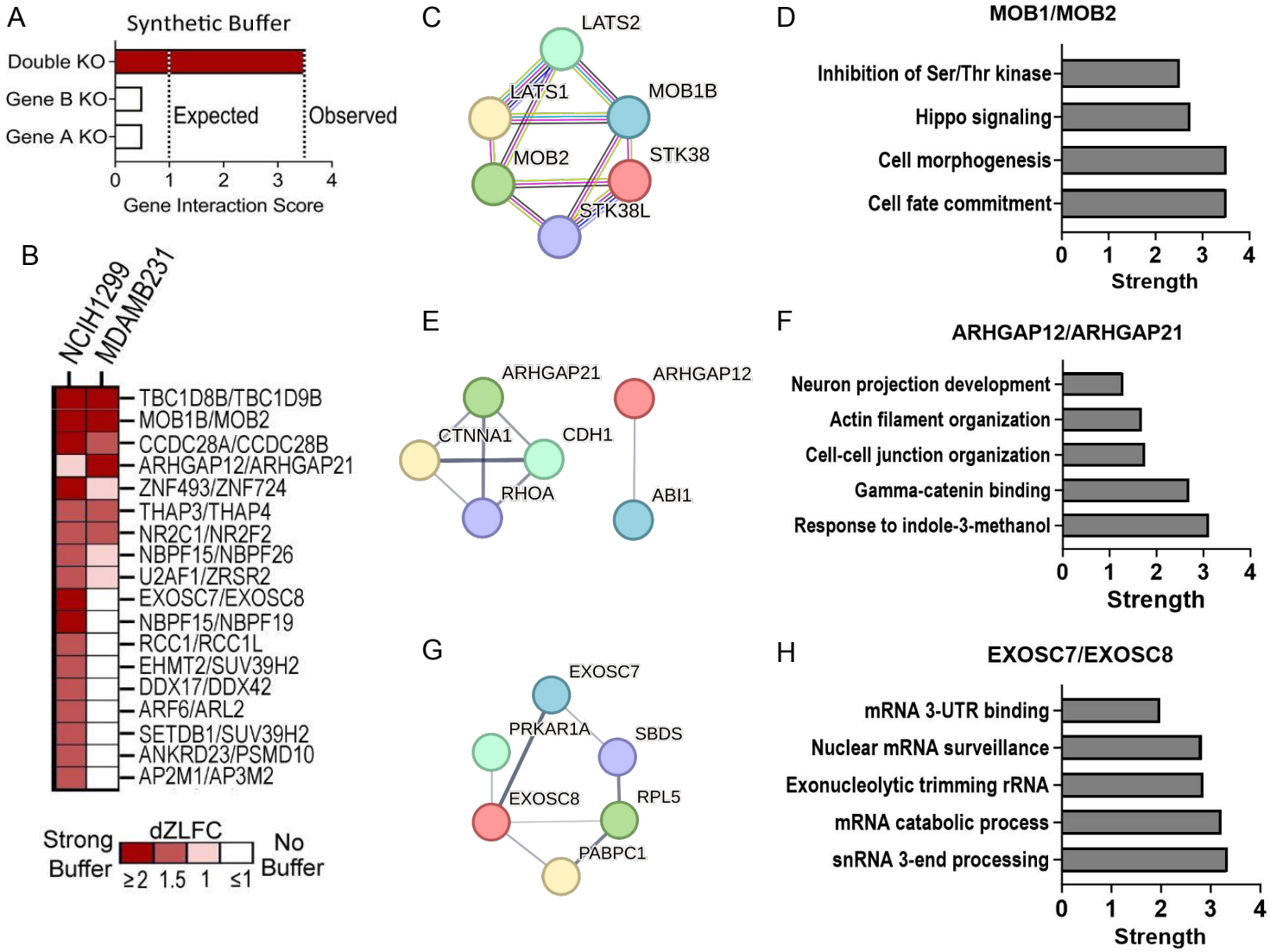

A
Synthetic Buffer
Double KO
Gene B KO
Gene A KO
Expected
Observed
0 1 2 3 4
Gene Interaction Score
B
NCIH1299
MDAMB231
TBC1D8B/TBC1D9B
MOB1B/MOB2
CCDC28A/CCDC28B
ARHGAP12/ARHGAP21
ZNF493/ZNF724
THAP3/THAP4
NR2C1/NR2F2
NBPF15/NBPF26
U2AF1/ZRSR2
EXOSC7/EXOSC8
NBPF15/NBPF19
RCC1/RCC1L
EHMT2/SUV39H2
DDX17/DDX42
ARF6/ARL2
SETDB1/SUV39H2
ANKRD23/PSMD10
AP2M1/AP3M2
Strong Buffer
dZLFC
≥2 1.5 1 ≤1
No Buffer
C
LATS2
LATS1
MOB1B
MOB2
STK38
STK38L
D
MOB1/MOB2
Inhibition of Ser/Thr kinase
Hippo signaling
Cell morphogenesis
Cell fate commitment
0 1 2 3 4
Strength
E
ARHGAP21
ARHGAP12
CTNNA1
CDH1
RHOA
ABI1
F
ARHGAP12/ARHGAP21
Neuron projection development
Actin filament organization
Cell-cell junction organization
Gamma-catenin binding
Response to indole-3-methanol
0 1 2 3 4
Strength
G
EXOSC7
PRKAR1A
SBDS
EXOSC8
RPL5
PABPC1
H
EXOSC7/EXOSC8
mRNA 3-UTR binding
Nuclear mRNA surveillance
Exonucleolytic trimming rRNA
mRNA catabolic process
snRNA 3-end processing
0 1 2 3 4
Strength

Figure 3

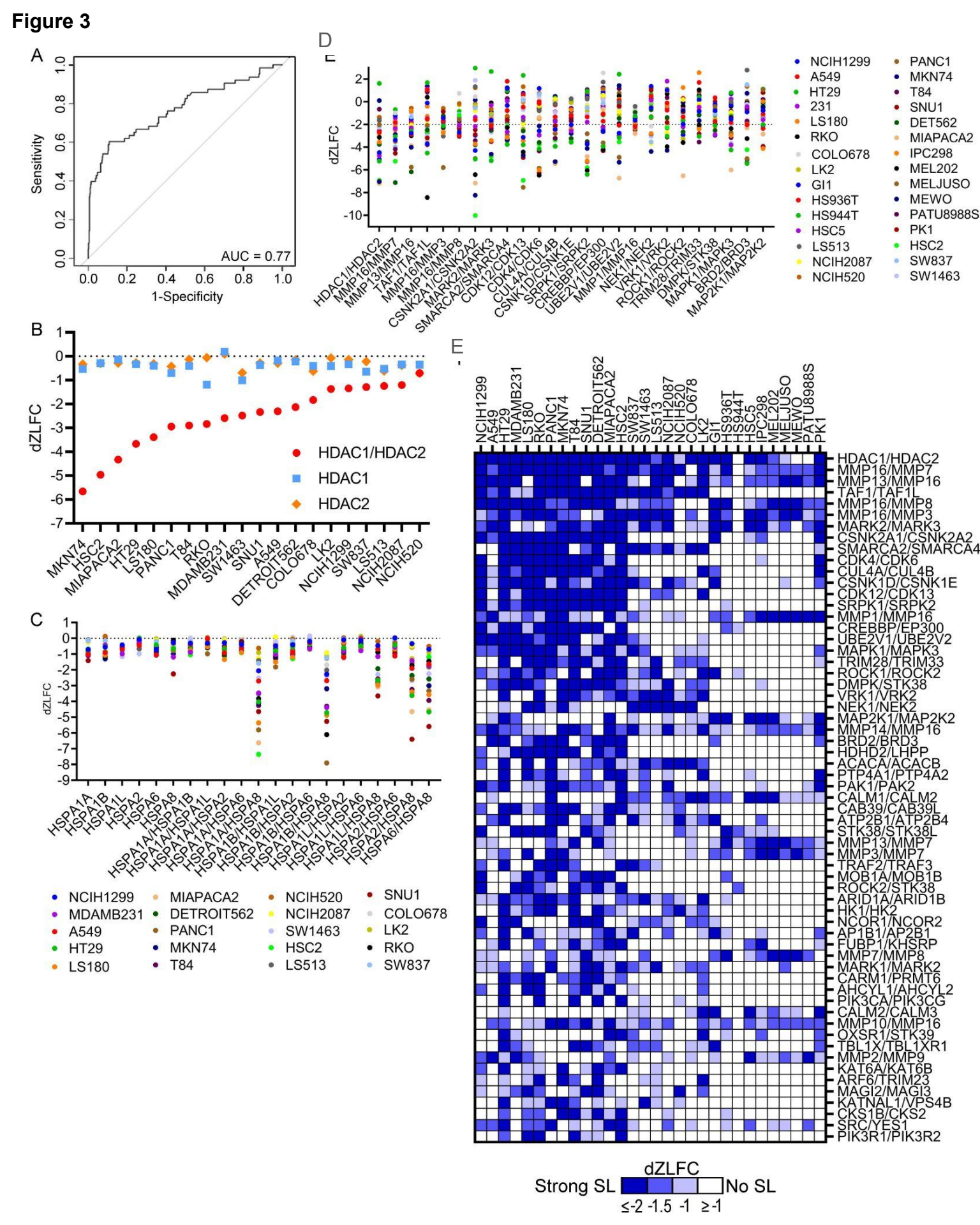
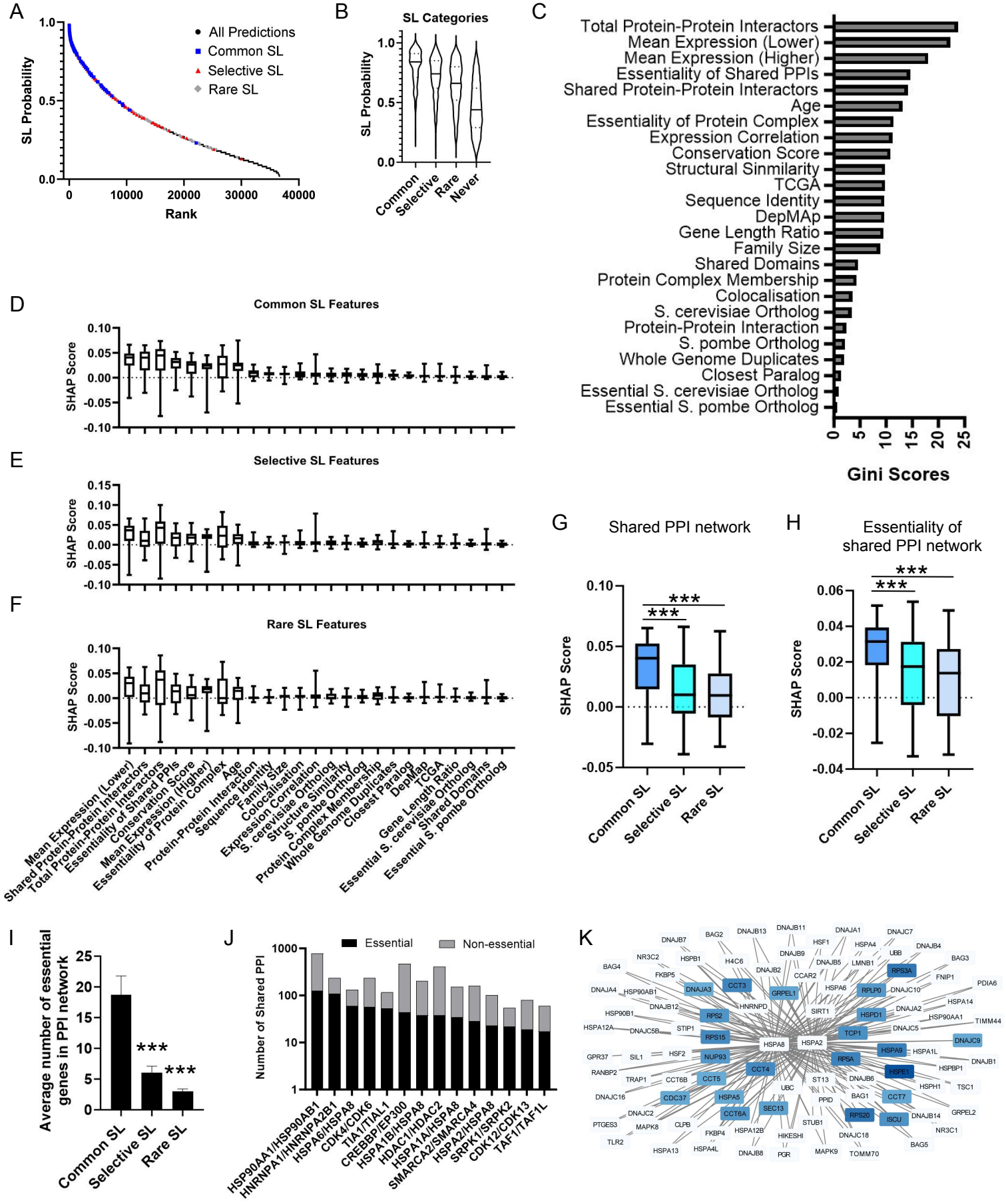

A
Sensitivity
1-Specificity
AUC = 0.77
D
dZLFC
HDAC1/HDAC2
MMP16/MMP7
MMP13/MMP16
TAF1/TAF1L
MMP16/MMP8
MMP16/MMP3
CSNK2A1/CSNK2A2
MARK2/MARK3
SMARCA2/SMARCA4
CDK12/CDK13
CDK4/CDK6
CUL4A/CUL4B
CSNK1D/CSNK1E
SRPK1/SRPK2
CREBBP/EP300
UBE2V1/UBE2V2
MMP1/MMP16
NEK1/NEK2
VRK1/VRK2
ROCK1/ROCK2
TRIM28/TRIM33
DMPK/STK38
MAPK1/MAPK3
BRD2/BRD3
MAP2K1/MAP2K2
NCIH1299
A549
HT29
231
LS180
RKO
COLO678
LK2
GI1
HS936T
HS944T
HSC5
LS513
NCIH2087
NCIH520
PANC1
MKN74
T84
SNU1
DET562
MIAPACA2
IPC298
MEL202
MELJUSO
MEWO
PATU8988S
PK1
HSC2
SW837
SW1463
B
dZLFC
HDAC1/HDAC2
HDAC1
HDAC2
MKN74
HSC2
MIAPACA2
HT29
LS180
PANC1
T84
RKO
MDAMB231
SW1463
SNU1
A549
DETROIT562
COLO678
LK2
NCIH1299
SW837
LS513
NCIH2087
NCIH520
C
dZLFC
HSPA1A
HSPA1B
HSPA1L
HSPA2
HSPA6
HSPA8
HSPA1A/HSPA1B
HSPA1A/HSPA1L
HSPA1A/HSPA2
HSPA1A/HSPA6
HSPA1A/HSPA8
HSPA1B/HSPA1L
HSPA1B/HSPA2
HSPA1B/HSPA6
HSPA1B/HSPA8
HSPA1L/HSPA2
HSPA1L/HSPA6
HSPA1L/HSPA8
HSPA2/HSPA6
HSPA2/HSPA8
HSPA6/HSPA8
NCIH1299
MDAMB231
A549
HT29
LS180
MIAPACA2
DETROIT562
PANC1
MKN74
T84
NCIH520
NCIH2087
SW1463
HSC2
LS513
SNU1
COLO678
LK2
RKO
SW837
E
NCIH1299
A549
HT29
MDAMB231
LS180
RKO
PANC1
MKN74
T84
SNU1
DETROIT562
MIAPACA2
HSC2
SW837
SW1463
LS513
NCIH2087
NCIH520
COLO678
LK2
GI1
HS936T
HS944T
HSC5
IPC298
MEL202
MELJUSO
MEWO
PATU8988S
PK1
HDAC1/HDAC2
MMP16/MMP7
MMP13/MMP16
TAF1/TAF1L
MMP16/MMP8
MMP16/MMP3
MARK2/MARK3
CSNK2A1/CSNK2A2
SMARCA2/SMARCA4
CDK4/CDK6
CUL4A/CUL4B
CSNK1D/CSNK1E
CDK12/CDK13
SRPK1/SRPK2
MMP1/MMP16
CREBBP/EP300
UBE2V1/UBE2V2
MAPK1/MAPK3
TRIM28/TRIM33
ROCK1/ROCK2
DMPK/STK38
VRK1/VRK2
NEK1/NEK2
MAP2K1/MAP2K2
MMP14/MMP16
BRD2/BRD3
HDHD2/LHPP
ACACA/ACACB
PTP4A1/PTP4A2
PAK1/PAK2
CALM1/CALM2
CAB39/CAB39L
ATP2B1/ATP2B4
STK38/STK38L
MMP13/MMP7
MMP3/MMP7
TRAF2/TRAF3
MOB1A/MOB1B
ROCK2/STK38
ARID1A/ARID1B
HK1/HK2
NCOR1/NCOR2
AP1B1/AP2B1
FUBP1/KHSRP
MMP7/MMP8
MARK1/MARK2
CARM1/PRMT6
AHCYL1/AHCYL2
PIK3CA/PIK3CG
CALM2/CALM3
MMP10/MMP16
OXSR1/STK39
TBL1X/TBL1XR1
MMP2/MMP9
KAT6A/KAT6B
ARF6/TRIM23
MAGI2/MAGI3
KATNAL1/VPS4B
CKS1B/CKS2
SRC/YES1
PIK3R1/PIK3R2
dZLFC
Strong SL
No SL
≤-2 -1.5 -1 ≥-1

Figure 4


A
SL Probability
Rank
All Predictions
Common SL
Selective SL
Rare SL
B
SL Categories
SL Probability
Common
Selective
Rare
Never
C
Total Protein-Protein Interactors
Mean Expression (Lower)
Mean Expression (Higher)
Essentiality of Shared PPIs
Shared Protein-Protein Interactors
Age
Essentiality of Protein Complex
Expression Correlation
Conservation Score
Structural Sinmilarity
TCGA
Sequence Identity
DepMAp
Gene Length Ratio
Family Size
Shared Domains
Protein Complex Membership
Colocalisation
S. cerevisiae Ortholog
Protein-Protein Interaction
S. pombe Ortholog
Whole Genome Duplicates
Closest Paralog
Essential S. cerevisiae Ortholog
Essential S. pombe Ortholog
Gini Scores
D
Common SL Features
SHAP Score
E
Selective SL Features
SHAP Score
F
Rare SL Features
SHAP Score
Mean Expression (Lower)
Shared Protein-Protein Interactors
Total Protein-Protein Interactors
Essentiality of Shared PPIs
Conservation Score
Mean Expression (Higher)
Essentiality of Protein Complex
Protein-Protein Interaction
Age
Sequence Identity
Family Size
Colocalisation
Expression Correlation
S. cerevisiae Ortholog
Structure Similarity
S. pombe Ortholog
Protein Complex Membership
Whole Genome Duplicates
Closest Paralog
DepMap
TCGA
Gene Length Ratio
Essential S. cerevisiae Ortholog
Shared Domains
Essential S. pombe Ortholog
G
Shared PPI network
SHAP Score
Common SL
Selective SL
Rare SL
H
Essentiality of shared PPI network
SHAP Score
Common SL
Selective SL
Rare SL
I
Average number of essential genes in PPI network
Common SL
Selective SL
Rare SL
J
Essential
Non-essential
Number of Shared PPI
HSP90AA1/HSP90AB1
HNRNPA1/HNRNPA2B1
HSPA6/HSPA8
CDK4/CDK6
TIA1/TIAL1
CREBBP/EP300
HSPA1B/HSPA8
HDAC1/HDAC2
HSPA1A/HSPA8
SMARCA2/SMARCA4
HSPA2/HSPA8
SRPK1/SRPK2
CDK12/CDK13
TAF1/TAF1L
K