Enabling cancer drug target discovery through genome-scale identification of synthetic lethal and tumor suppressive paralog pairs

Phoebe C. R. Parrish[1,2,3*], James D. Thomas[2,4*], Austin Gabel[2,3,4], Shriya Kamlapurkar[1,2], Robert K. Bradley[2,3,4], Alice H. Berger[1,2,3]

1. Human Biology Division, Fred Hutchinson Cancer Research Center; 2. Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center;
3. Department of Genome Sciences, University of Washington; 4. Basic Sciences Division, Fred Hutchinson Cancer Research Center

Background

CRISPR knockout (KO) screens have accelerated the discovery of important cancer genetic dependencies. But traditional CRISPR-Cas9 screens are limited in their ability to assay the function of redundant or duplicated genes. Paralogs in multi-gene families constitute two-thirds of the protein-coding genome, so this blind spot is the rule, not the exception. To overcome the limitations of single gene CRISPR KO screens, we developed paired guide RNAs for Paralog gENetic interaction mapping (pgPEN), a pooled CRISPR/Cas9 approach which targets over 1,000 duplicated human paralogs in single and double KO configurations. We applied pgPEN to two cell lineages and discovered that 12% of human paralogs exhibit synthetic lethality in at least one cellular context. These synthetic lethal paralogs represent new druggable targets for oncology drug discovery.

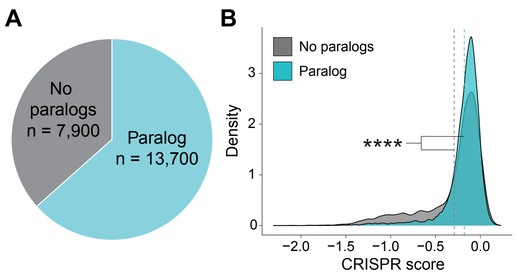

Figure 1. (A) Pie chart of human genes grouped by whether they are part of a paralog gene family with 10-99% amino acid sequence identity. **(B)** Density plot of CRISPR scores (CS) for a single gene CRISPR KO screen in PC9 lung adenocarcinoma cells. Data from Vichas et al., 2021 (Nat Commun). Dashed lines indicate the mean CS of genes in each group. P < 2.2e-16 by Kolmogorov-Smirnov test.

pgPEN library design and methods

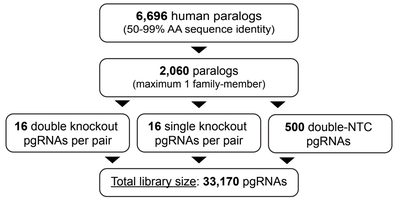

Figure 2. pgPEN library design. Schematic of filtering strategy used to select paralogous genes for inclusion in the pgPEN library. Paralogs were selected from the Ensembl database and filtered for families with a maximum of two members sharing 50-99% amino acid sequence identity. 554 of pgPEN gene products are classified as druggable targets (Finan et al., 2017, Sci Transl Med).

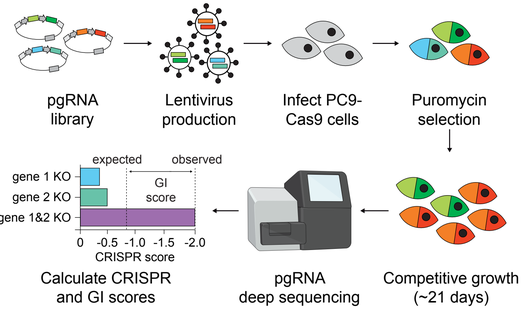

Figure 3. Overview of pgPEN approach for genetic interaction (GI) mapping. The pgPEN library was applied to PC9 cells constituviely expressing Cas9 using a standard pooled CRISPR screening method. GI scores were calculated for each paralog pair by comparing expected vs. observed CS under an additive model. A t-test with Benjamini-Hochberg false discovery rate (FDR) correction was used to compute the statistical significance of the difference in double KO GI scores versus the distribution of control (single KO) GI scores.

pgPEN reveals synthetic lethal human paralog interactions across two cancer cell lines

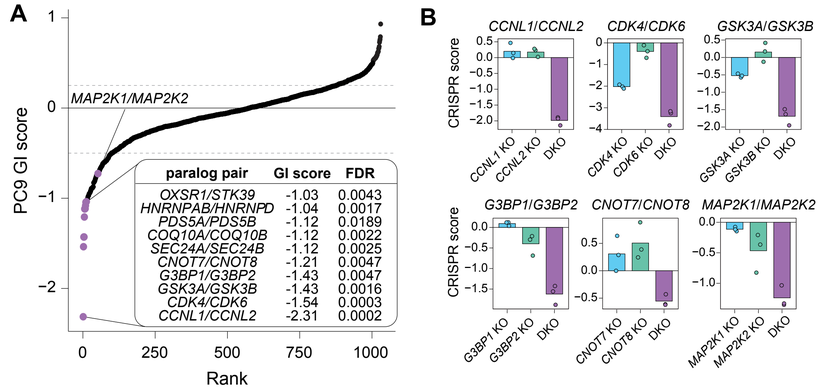

Figure 4. Top PC9 synthetic lethal paralogs. (A) Rank plot of paralog GI scores in PC9 cells. Table insert, top 10 paralogs based on GI score with FDR < 0.1. **(B)** CRISPR scores for representative synthetic lethal paralog pairs. Data shown is the mean CS for each single KO or double KO (DKO) target across 3 biological replicates with replicate data shown in overlaid points. Top synthetic lethal pairs included known synthetic lethal paralogs *MAP2K1/MAP2K2*, drug targets *CDK4/CDK6* and *GSK3A/GSK3B*, and novel synthetic lethal pairs such as splicing regulators *CCNL1/CCNL2*, Ras pathway activators *G3BP1/G3BP2*, and mRNA deadenylase complex members *CNOT7/CNOT8*.

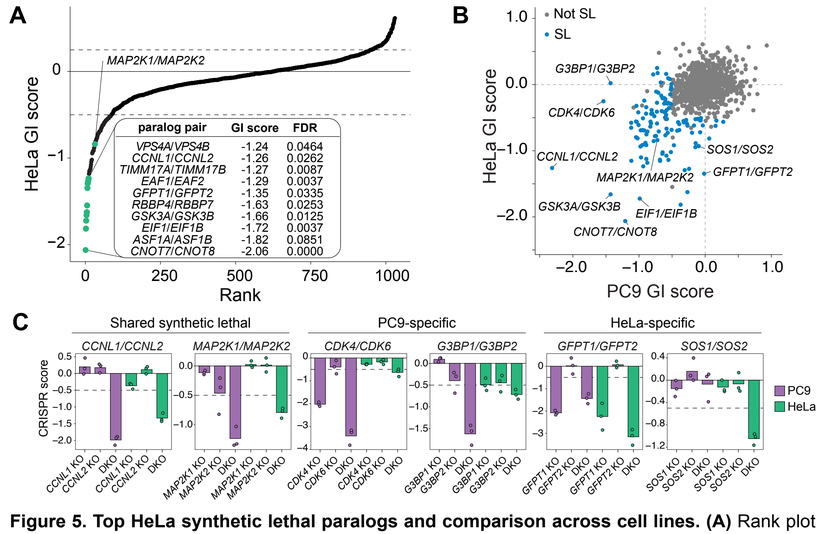

Figure 5. Top HeLa synthetic lethal paralogs and comparison across cell lines. (A) Rank plot of paralog GI scores in HeLa cells. Table insert, top 10 paralogs based on GI score with FDR < 0.1. **(B)** Scatter plot of target-level GI scores for paralog pairs in PC9 versus HeLa cells. Blue, synthetic lethal paralog pairs with GI < -0.5 and FDR < 0.1 in either PC9 or HeLa cells; gray, all paralog pairs with GI ≥ -0.5 or FDR ≥ 0.1. **(C)** CRISPR scores for representative synthetic lethal paralog pairs identified in the PC9 and HeLa cell screens. Data shown is the mean CS for each single KO or DKO target across three biological replicates with replicate data shown in overlaid points. Shared synthetic lethal paralogs have FDR < 0.1 in both cell lines, PC9-specific paralogs have FDR < 0.1 in PC9 only, and HeLa-specific paralogs have FDR < 0.1 in HeLa only. RNAseq data (not shown) revealed that all genes shown were expressed at TPM > 2.0 in both cell lines.

Putative tumor suppressor paralogs

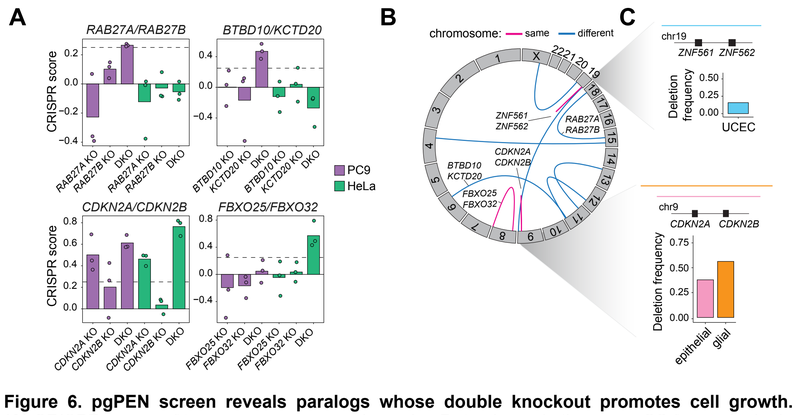

Figure 6. pgPEN screen reveals paralogs whose double knockout promotes cell growth. (A) CRISPR scores of tumor suppressor paralog pairs. Data shown is the mean CS for each single KO or DKO target across three biological replicates with replicate data shown in overlaid points. **(B)** Circos plot showing the genomic locations of tumor suppressive paralog pairs. Blue arcs indicate paralog pairs located on different chromosomes, while pink arcs represent paralog pairs located on the same chromosome. **(C)** Top, diagram of a recurrent deletion seen in uterine corpus endometrial carcinoma (UCEC) data from TCGA that spans the genomic locus containing *ZNF561* and *ZNF562*, and a bar plot indicating the deletion frequency. Bottom, diagram of recurrent deletions in epithelial and glial cancers that span the genomic locus containing *CDKN2A* and *CDKN2B*, and a bar plot showing the deletion frequency in each cancer subtype.

Conclusions

- 12% of duplicate paralogs exhibit synthetic lethality, demonstrating that paralogs are a rich source of genetic interactions.
- Synthetic lethal interactions among paralogs could be harnessed for cancer therapy, since the aneuploid genomes typical of cancer cells commonly harbor deletions and inactivating mutations in one or more paralogs. Moreover, homology among paralogs enables simultaneous targeting with a single small molecule.
- We provide the first systematic identification of tumor suppressive paralog pairs. We identified ten paralog pairs whose combined loss significantly promotes cancer cell line growth.

Acknowledgments

We thank Dr. Athea Vichas (Fred Hutchinson Cancer Research Center) for advice on CRISPR screens in PC9 cells. Sequencing was performed by the Fred Hutch Genomics Shared Resource. Computational studies were supported in part by FHCRC's Scientific Computing Infrastructure.

Funding sources:

NIH NATIONAL CANCER INSTITUTE

GRFP

LUNG CANCER RESEARCH FOUNDATION

For more information and results from the pgPEN screens, please see: https://doi.org/10.1016/j.celrep.2021.109597