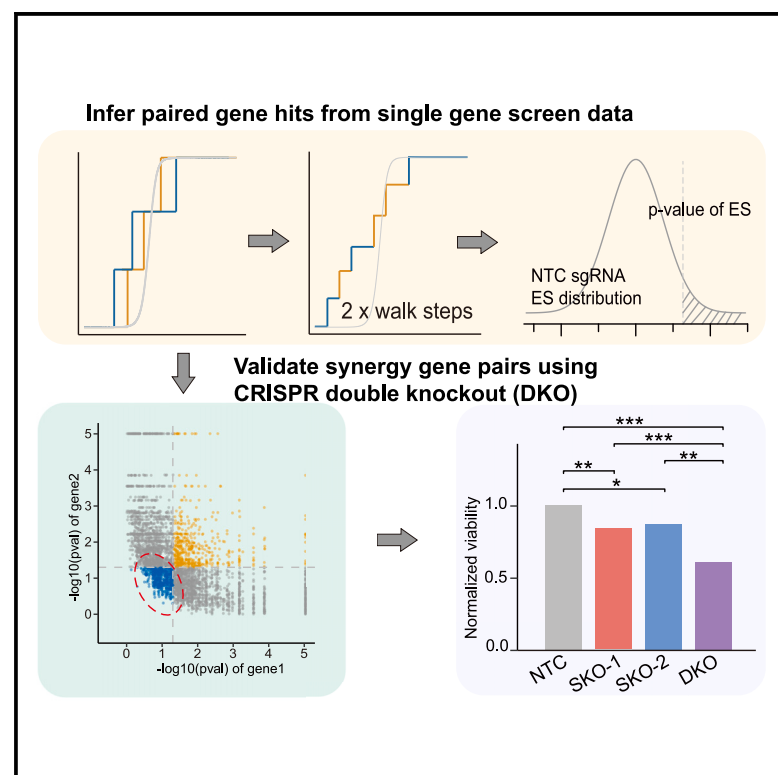


# Patterns

## Sensitive detection of synthetic response to cancer immunotherapy driven by gene paralog pairs

### Graphical abstract



### Authors

Chuanpeng Dong, Feifei Zhang, Emily He, ..., James Cai, Hongyu Zhao, Sidi Chen

### Correspondence

sidi.chen@yale.edu

### In brief

Combinatorial targeting has shown promise in cancer therapies; however, the experimental detection of paired functional targets remains challenging due to the vast number of possible combinations. In this study, the authors introduce a method to repurpose single-gene screen data to identify functional paralog pairs that enhance immunotherapy, extending the search through machine learning approaches. The authors further demonstrate experimentally that synergistic paralog pairs can improve immunotherapy efficacy, even when neither individual target is effective, a scenario often underestimated.

### Highlights

- Computational approaches for detecting paired targets using CRISPR screen data
- Validation of paralog gene pairs' effect on T cell killing using CRISPR double knockouts
- A machine learning model for predicting paralog gene pairs that modulate immunotherapy

Article

# Sensitive detection of synthetic response to cancer immunotherapy driven by gene paralog pairs

Chuanpeng Dong,<sup>1,2,3,4,5,10</sup> Feifei Zhang,<sup>1,2,3,10</sup> Emily He,<sup>1,2,3,7</sup> Ping Ren,<sup>1,2,3</sup> Nipun Verma,<sup>1,2,3</sup> Xinxin Zhu,<sup>4,5</sup> Di Feng,<sup>5,9</sup> James Cai,<sup>5,9</sup> Hongyu Zhao,<sup>1,4,5,8</sup> and Sidi Chen<sup>1,2,3,4,5,6,11,\*</sup>

<sup>1</sup>Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

<sup>2</sup>System Biology Institute, Yale University, West Haven, CT, USA

<sup>3</sup>Center for Cancer Systems Biology, Yale University, West Haven, CT, USA

<sup>4</sup>Center for Biomedical Data Science, Yale University School of Medicine, New Haven, CT, USA

<sup>5</sup>Yale-Boehringer Ingelheim Biomedical Data Science Fellowship Program, Yale University School of Medicine, New Haven, CT, USA

<sup>6</sup>Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, CT, USA

<sup>7</sup>Yale College, Yale University, New Haven, CT, USA

<sup>8</sup>Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA

<sup>9</sup>Global Computational Biology and Digital Sciences, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA

<sup>10</sup>These authors contributed equally

<sup>11</sup>Lead contact

\*Correspondence: [sidi.chen@yale.edu](mailto:sidi.chen@yale.edu)

<https://doi.org/10.1016/j.patter.2025.101184>

**THE BIGGER PICTURE** Immunotherapies leverage the patient's own immune system to fight cancer, and emerging therapies in this domain have revolutionized cancer treatment; however, resistance and relapse remain challenges. Recent efforts to identify combinatorial targets, or pairs of genes that could be used together for cancer treatment, are hindered by limitations in current high-throughput screening approaches. This type of screening often misses promising target pairs. In this study, we leverage public genomic datasets to identify and validate gene pairs that synergistically enhance effector T cell function, ultimately promoting immune-mediated tumor eradication. Our findings underscore the overlooked potential of paralog gene pairs in driving functional synergy and offer a framework for designing more effective combinatorial screen libraries. While this study focuses on cancer immunotherapy, the methods we developed for identifying combinatorial targets may be broadly applicable to other viability functional screening asks, where large-scale datasets are more readily available.

## SUMMARY

Immunotherapies, including checkpoint blockade and chimeric antigen receptor T cell (CAR-T) therapy, have revolutionized cancer treatment; however, many patients remain unresponsive to these treatments or relapse following treatment. CRISPR screenings have been used to identify novel single gene targets that can enhance immunotherapy effectiveness, but the identification of combinatorial targets remains a challenge. Here, we introduce a computational approach that uses sgRNA set enrichment analysis to identify cancer-intrinsic paralog pairs for enhancing immunotherapy using genome-wide screens. We have further developed an ensemble learning model that uses an XGBoost classifier and incorporates features to predict paralog gene pairs that influence immunotherapy efficacy. We experimentally validated the functional significance of these predicted paralog pairs using CRISPR double knockout (DKO). These data and analyses collectively provide a sensitive approach to identifying previously undetected paralog gene pairs that can significantly affect cancer immunotherapy response, even when individual genes within the pair have limited effect.

## INTRODUCTION

Emerging immunotherapies, especially immune checkpoint blockade (ICB) and adoptive cell therapy, have revolutionized cancer treatment for multiple cancer types; however, a significant fraction of patients fail to respond to immunotherapies or relapse following treatment.<sup>1</sup> Recent efforts have focused on using CRISPR knockout (KO) or activation screening to identify targets that enhance T cell effector function and augment immune killing capability.<sup>2</sup> Nevertheless, overcoming resistance by manipulating a single gene remains challenging due to the compensatory effects of other genes. Combination therapies have been proposed as a promising strategy to overcome tumor resistance to monotherapy when single agents are ineffective.<sup>3</sup>

Paralogs, genes that originate from the same ancestors and share similar functions, often work in concert to promote cell survival. Numerous focused studies have utilized genomic data and pooled screening techniques to explore how targeting paralogous genes can impact cancer cell viability.<sup>4–6</sup> Recent research has shown that targeting these genes simultaneously in cancer cells can enhance the effectiveness of immunotherapy by preventing the cells from evading T cell-mediated destruction. For instance, Park et al. demonstrated that treating mice with both anti-PD-L1 and anti-PD-L2 antibodies significantly reduced tumor growth in specific mouse models.<sup>7</sup> Additionally, signals from both *PARP-1* and *PARP-2* are essential for initiating an effective T cell response against breast cancer in mice.<sup>8</sup> While pooled CRISPR screening is popular for identifying single genes that modulate specific responses, including those relevant to immunotherapy, double-knockout screens aiming to identify paired targets remain rare. Recent efforts in double-knockout screens, covering dozens to hundreds of gene pairs, have been conducted to identify novel gene combinations that lead to synthetic lethality and boost T cell function.<sup>5,9,10</sup> However, a comprehensive, high-throughput strategy to screen all potential paralog pairs related to immunotherapy is challenging because a very large screening library that includes hundreds of thousands of gene combinations would be both costly and unfeasible. To solve this problem, we have developed a computational approach to identify the most promising paralog gene combinations, which can then be used to build a focused and manageable library for experimental validation.

Here, we used a single guide RNA (sgRNA) set enrichment method to identify cancer-intrinsic paralog pairs that enhance T cell killing from prior pooled screening data. We show that our method identifies novel synergistic paralog pairs that are successfully validated experimentally. We have also constructed an ensemble learning XGBoost classifier to predict true-positive paralog pairs enhancing immunotherapy by incorporating features such as gene characteristics, sequence and structural similarity, protein-protein interaction (PPI) networks, and gene co-evolution data. To enhance the reliability of the XGBoost classifier predictions, we conducted further experimental validations on the highest-ranked paralog pairs identified. This study is expected to uncover novel therapeutic targets and inspire new combination immunotherapies to aid cancer treatment.

## RESULTS

### Identifying cancer-intrinsic paired paralogs that enhance T cell-mediated killing using CRISPR screen data

A large proportion of genes have paralogs that perform redundant functions. Inactivation of functionally important paralog pairs can lead to synthetic lethality, as well as more efficient modulation of biological pathways.<sup>9,11</sup> Unfortunately the identification of these paralog pairs is limited in single-gene perturbation screens due to compensation between paralogous genes,<sup>12</sup> and dual gene screening is often unfeasible due to the large size of CRISPR libraries required to encompass all potential combinations of paralogous gene pairs. Current double-knockout screens cover only a limited number of genes and require a substantial number of gRNA pairs (Table S1). Even the largest paired gene screens to date cover less than 0.1% of all possible combinations.<sup>5,9,12–18</sup> To overcome this limitation, we proposed an unbiased computational approach to filter and predict functional gene pairs from prior published genome-wide CRISPR screen data.

We initially attempted to search for immunotherapy-aiding paralog pairs using conventional MAGeCK results in Lawson's CRISPR screen dataset,<sup>19</sup> which was originally aimed at identifying genes involved in evading T cell-mediated cytotoxicity across six murine tumor cell lines. Based on a strict dual-hit gene criterion, we observed that few or no suitable pairs identified (Figure S1A); however, relaxing the criteria to include single hit genes leads to a significant increase in uncertain pairs. To overcome this, we developed an enrichment-based methodology named paralog sgRNA set enrichment analysis (pSSEA), which integrates the sgRNAs of gene pairs to enhance the identification of high-potential paralog pairs. We provide a schematic of the pSSEA algorithm in Figure 1B. Briefly, the process begins with the results of differential expression analysis (using DESeq2, MAGeCK, etc.) as input. The pSSEA evaluates the enrichment score by employing a Kolmogorov-Smirnov (KS)-like random walk statistic, a method similar to those used in GSVA<sup>20</sup> and ASSESS.<sup>21</sup> The detailed steps include

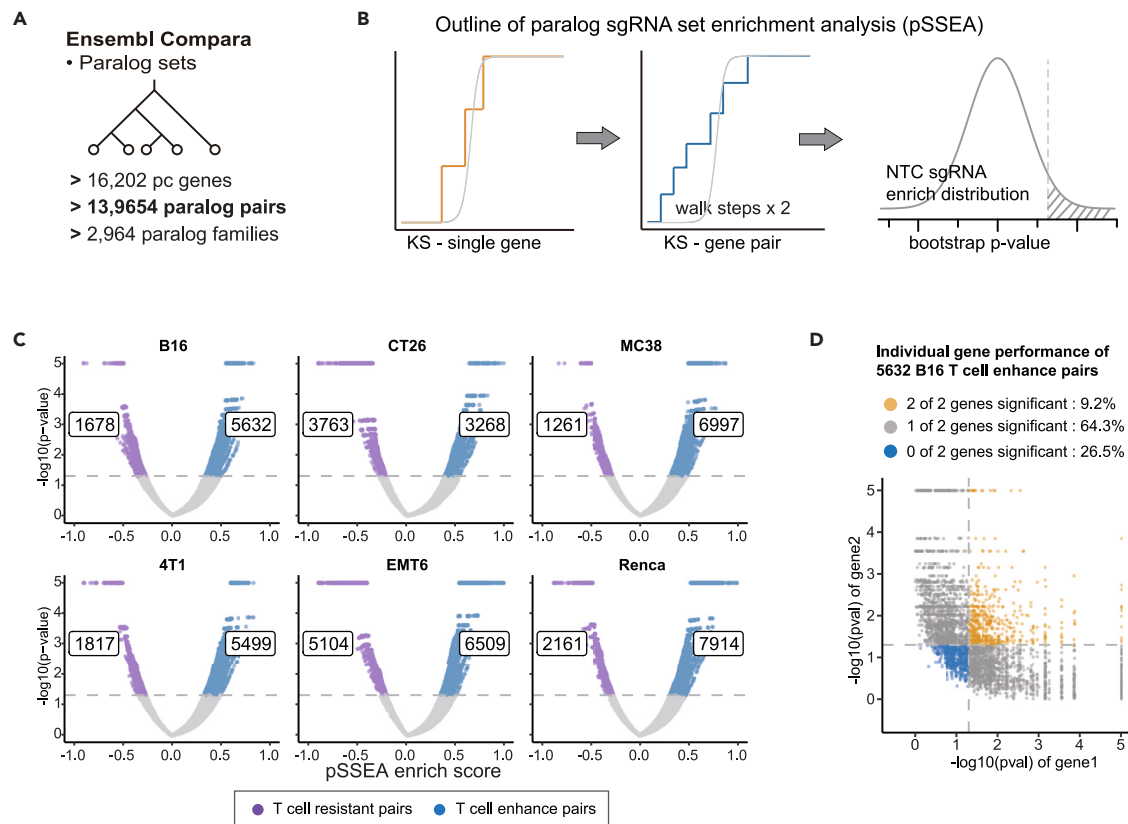
$$v(\ell) = \frac{\sum_{i=1}^{\ell} |r_i| I(g_i \in \gamma)}{\sum_{i=1}^p |r_i| I(g_i \in \gamma)} - \frac{\sum_{i=1}^{\ell} |r_i| I(g_i \notin \gamma)}{p - |\gamma|} \quad (\text{Equation 1})$$

where  $r_i$  is the rank order of sgRNAs, sorted by the log fold changes;  $\gamma$  is the sgRNA set that belongs to the paralog gene pair;  $I(g_i \in \gamma)$  is the indicator function on whether the  $i$ -th sgRNA is in paralog sgRNA set  $\gamma$ ,  $|\gamma|$  is the number of sgRNAs belonging to the gene pairs, and  $p$  is the number of total sgRNAs in the library.

$$ES = v \left[ \arg \max_{\ell=1, \dots, p} |v(\ell)| \right] \quad (\text{Equation 2})$$

The enrichment scores for paralog pairs were calculated by identifying the maximum deviation from zero during a random walk analysis.

To assess the significance of enrichment, we created an empirical null distribution by randomly selecting an equivalent



**Figure 1. Identification of paralog pairs that enhance T cell killing using pSSEA**

(A) Protein-coding paralogous genes retrieved from Ensembl Compara database using bioMart.

(B) Workflow for identifying paralog pair hits using the paralog sgRNA set enrichment analysis (pSSEA).

(C) Identification of paralog pairs that enhance T cell killing in six cancer models using pSSEA. CRISPR screen data were taken from GEO: GSE149933. The volcano plot of pSSEA outputs, with the x axis representing the enrichment score and the y axis showing the negative logarithm (base 10) of the  $p$  value from pSSEA, is shown.

(D) Scatterplot illustrating the characteristics of identified T cell-enhancing paralog pairs using a single-gene enrichment model. The x and y axes represent the negative logarithm (base 10) of the  $p$  values from the enrichment results of gene1 and gene2, respectively.

number of non-targeting control (NTC) sgRNAs and repeating this process 10,000 times to ensure robustness. The significance of enrichment for each gene pair was then determined by comparing their sgRNA distributions against both the positive and negative tails of the NTC null distribution.

We re-analyzed Lawson's CRISPR screen dataset<sup>19</sup> to identify cancer-intrinsic paralog pairs that enhance T cell-mediated killing function using the pSSEA method. From the Ensembl Compara database (Ensembl, v.102), a total of 139,654 protein-coding gene paralog pairs were retrieved, covering 16,202 protein-coding genes, as shown in Figure 1A. First, DESeq2 was used to quantify the selective depletion of sgRNAs within tumor populations under immune selection from co-cultured OT-1 T cells. Subsequently, using the pSSEA framework, we identified paralogous gene pairs that potentially influence T cell-mediated cytotoxicity based on the differential sgRNA results obtained from DESeq2. Our analysis identified 7,310 significant paralog pairs (5.2% of 139,654) with  $p$  values below 0.05 in B16F10 melanoma cell data. Among these, 5,632 pairs were positively enriched, suggesting vulnerability to T cell killing, and 1,678 pairs were negatively enriched, suggesting resistance to T cell killing

(Figure 1D). Using these same methods, we identified paralog pairs associated with vulnerability and resistance to T cell killing for CT26, MC38, 4T1, EMT6, and Renca cell lines, as shown in Figure 1C.

To investigate the properties of paralog pairs identified as enriched through pSSEA, we assessed individual gene enrichment using pSSEA in single-gene mode. Among the 5,632 pairs associated with vulnerability to T cell killing in B16F10 melanoma cells, 3,620 pairs (64.3%) were single hit, with only one gene showing significance; 520 pairs (9.2%) were dual hits, with both genes significant; and the remaining 1,492 pairs (26.5%) showed no hits of either of the genes within the identified paralog pair (Figure 1D; Table S2). Notably, we observed that over a quarter of the paralog pairs exhibited a synergistic effect despite no hits at the single-gene level. This observation was confirmed using data from Parrish et al., which includes both single knockout (SKO) and double knockout (DKO) screens for tumor viability. By analyzing DKO alongside SKO screens using MaGeck, we found that out of 354 synthetic lethality pairs identified in the DKO screen, 95 were not significant as individual genes in the SKO screen (26.8%; Figure S1B). We also observed that

pSSEA was more effective in identifying synthetic gene pairs compared to selecting pairs based on the significance of individual genes or both genes in SKO screens (Figure S1C). This aligns with our previous findings in Figure 1D that many potential synergistic gene pairs may not be significant as individual genes, highlighting a broad spectrum of potential targets that might be overlooked by obtaining gene pairs according to individual gene results.

### Validation of synergistic cancer-intrinsic paralog pairs that, together, boost T cell killing

To determine if the gene pairs identified by pSSEA truly function synergistically, we focused on the subset of paralog pairs that exhibited a synergistic effect despite neither gene being individually significant. This subset constituted over 25% of the functional paralog pairs identified (Figure 1D). For our experimental validation using the B16F10 murine melanoma cell model, we first screened for paralog pairs in the B16F10 model, where neither gene showed a significant effect individually but demonstrated a significant effect when both genes were perturbed together. This ensured that we could efficiently validate potential gene pairs in this tumor model. Then, to enhance the robustness of our selection, we applied pSSEA to five additional tumor models, including CT26, MC38, 4T1, EMT6, and Renca. We aimed to identify paralog pairs that consistently enhanced T cell killing across multiple models, thereby reducing the likelihood that observed effects were model-specific anomalies. We observed that no paralog pair was significant across all six models. However, some pairs appeared frequently, being significant in up to four models (Figure 2A; Table S3). Ultimately, four paralog pairs, *Rbm45/Rbms2*, *Ppp2r2a+Ppp2r2d*, *Elf2+Etv6*, and *Adam10+Adam15*, were selected for further experimental assessment of their potential to enhance T cell killing efficiency (Figures S1D–S1G).

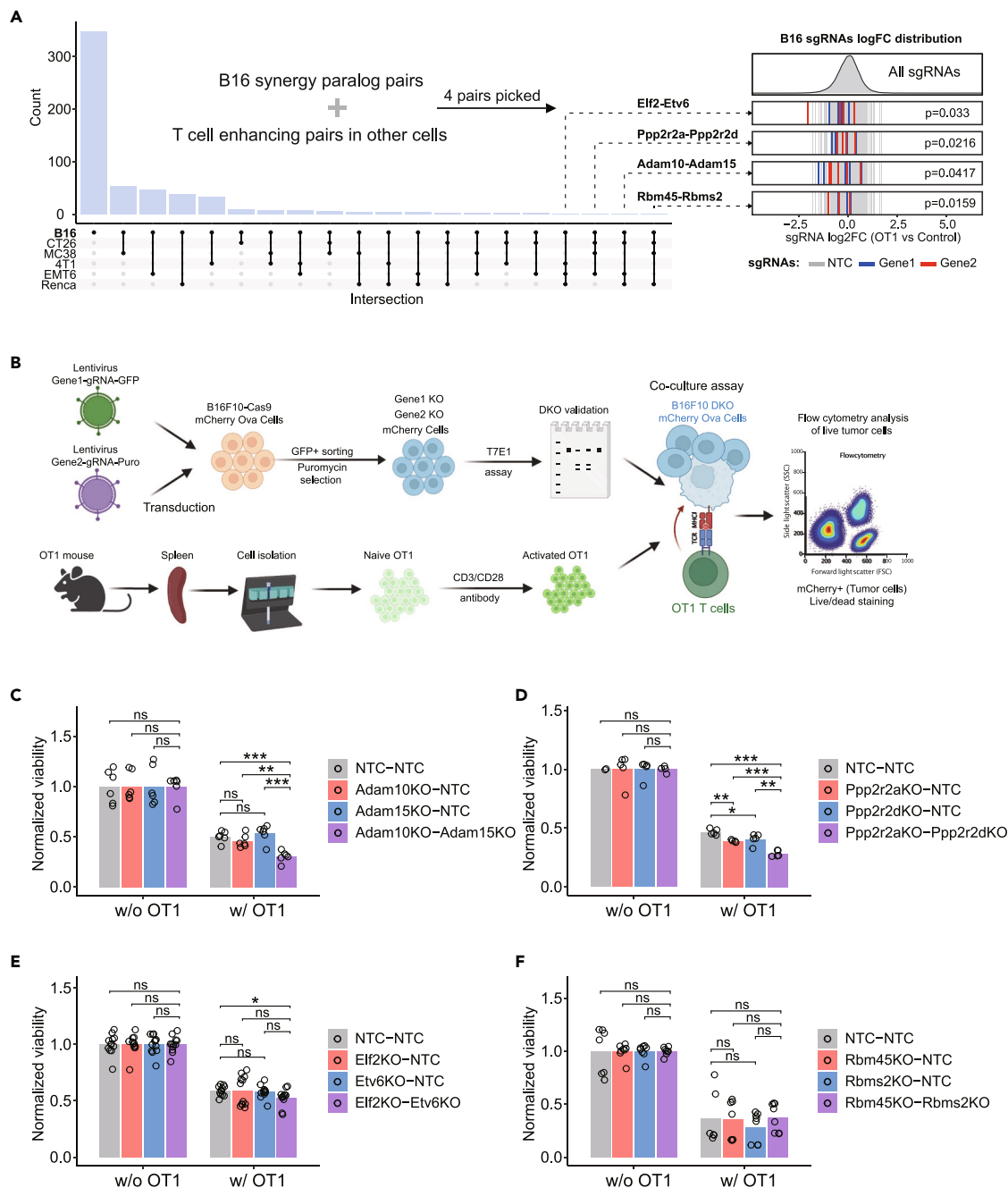
To perform the paired gene KO, we utilized the same gRNAs from the original gRNA-library pool used in the CRISPR screen. For the first gene within the paralog pair, the targeting gRNA was cloned into a GFP-expressing construct, while for the second gene within the paralog pair, the targeting gRNA was inserted into a puromycin-expressing construct (Figure 2B). To evaluate the synergistic function of each gene pair, we established four experimental groups with targeted KOs: NTC paired with NTC (NTC-NTC), gene1 KO paired with NTC (gene1KO-NTC), gene2 KO paired with NTC (gene2KO-NTC), and a double knockout (DKO) of both genes (gene1KO-gene2KO). To generate cells with these specific gene targets, we performed dual transduction of B16F10-Cas9-mCherry-OVA tumor cells, which express the OVA antigen, using lentiviruses carrying the gRNAs with GFP or puromycin markers. GFP-positive cells were sorted and then selected for puromycin-resistant cells to obtain either single- or double-knockout cells. Subsequently, gene editing was verified, and KO efficiency was evaluated using the T7E1 assay. To investigate the role of the paired genes in T cell-mediated cytotoxicity, we co-cultured B16F10-Cas9-mCherry-OVA tumor cells from the four experimental groups described above with OT-I T cells, whose T cell receptor (TCR) recognizes the SIINFEKL peptide presented by Ova antigen-expressing tumor cells, and assessed tumor cell survival.

In these co-culture experiments, we found that simultaneous inactivation of *Adam10* and *Adam15* significantly increased T cell-mediated tumor cell death, while single knockouts of *Adam15* or *Adam10* did not affect T cell activity (Figures 2C and S1H). Similarly, *Ppp2r2a+Ppp2r2d* double knockouts displayed significantly greater vulnerability to T cell cytotoxicity, whereas individual knockouts of *Ppp2r2a* or *Ppp2r2d* showed only modestly enhanced T cell killing (Figures 2D and S1I). The *Elf2* and *Etv6* combination showed significantly enhanced T cell killing compared to the NTC control but not compared to the single-gene *Elf2* and *Etv6* KO cells (Figures 2E and S1J). The *Rbm45/Rbms2* pair showed no discernible impact on T cell efficiency in either single- or double-knockout cells compared to the NTC group (Figures 2F and S1K). Furthermore, we evaluated the synergistic effects of combination targets in response to T-cells mediated killing, synergy analysis showed that *Adam10+Adam15*, *Ppp2r2a+Ppp2r2d*, and *Elf2+Etv6* demonstrated positive synergy scores across the HSA, Loewe, and delta log fold change (dLFC) models (see the methods section and Table S4). In total, three out of four predicted paralog pairs were successfully confirmed to exhibit increased T cell vulnerability, highlighting previously overlooked synergistic gene pairs. These findings also provide fundamental experimental evidence for the accuracy of using the pSSEA method to effectively identify paralog pairs that enhance T cell killing efficacy.

### Building an ensemble XGBoost classifier to predict paralog pairs that enhance T cell-mediated cytotoxic killing

Using a CRISPR screen dataset, we have demonstrated that paralog pairs can work together to enhance T cell-mediated cytotoxicity. However, this CRISPR screen dataset was derived from *in vitro* co-culture of cancer cell lines with cytotoxic T cells, an assay with significant limitations and potential experimental biases. Therefore, we developed a system for deeper characterization of paralog pairs and prediction of their potential for enhancing cancer immunotherapy. Utilizing T cell-enhancing paralog pairs identified in B16F10 melanoma data with pSSEA as true positives and those identified as insignificant as false positives, we compiled 32 features representing various aspects of paralogous gene interactions. These features include sequence characterization, expression abundance and correlation, shared PPIs, complex membership, co-evolution, tumor microenvironment (TME) associations, perturbation similarities, and combined survival associations in patients with cancer (Figure 3A; Table S5). Then, an ensemble classifier was trained, utilizing the identified 32 features, to predict potential paralog pairs. In order to ensure the quality of model training and testing, the paralog pairs were filtered using several criteria: paralog pairs identified as T cell-enhancing pairs or insignificant pairs in B16 data; a sequence BLAST identity larger than 30% with an *e*-value < 1e–5; measurements in structural similarity, synthetic lethality, and survival associations; and paralog pairs with family sizes of 20 or fewer to prevent overrepresentation of specific ortholog families. A total of 3,204 paralog pairs were selected from the initial set of 139,654 paralog pairs. Furthermore, 80% of these 3,204 pairs were randomly assigned to the training dataset (*n* = 2,547), which includes negative pairs (*n* = 2,274) and





**Figure 2. Synergistic paralog pairs that, together, boost T cell killing function in B16F10 tumor model**

(A) The T cell synergistic paralog pairs identified in the B16F10 melanoma model and the intersection across the other 5 tumor models. T cell-enhanced paralog pairs were identified using pSSEA, and synergistic pairs were denoted as pairs where neither of the two genes could reach significance using a single-gene enrichment approach. Four paralog pairs were identified in 4 out of 6 tumor models. Left: UpSet plot of overlap synergistic paralog pairs identified from the B16F10 melanoma model that were also significant in the 5 other cancer cell models in the Lawson study.

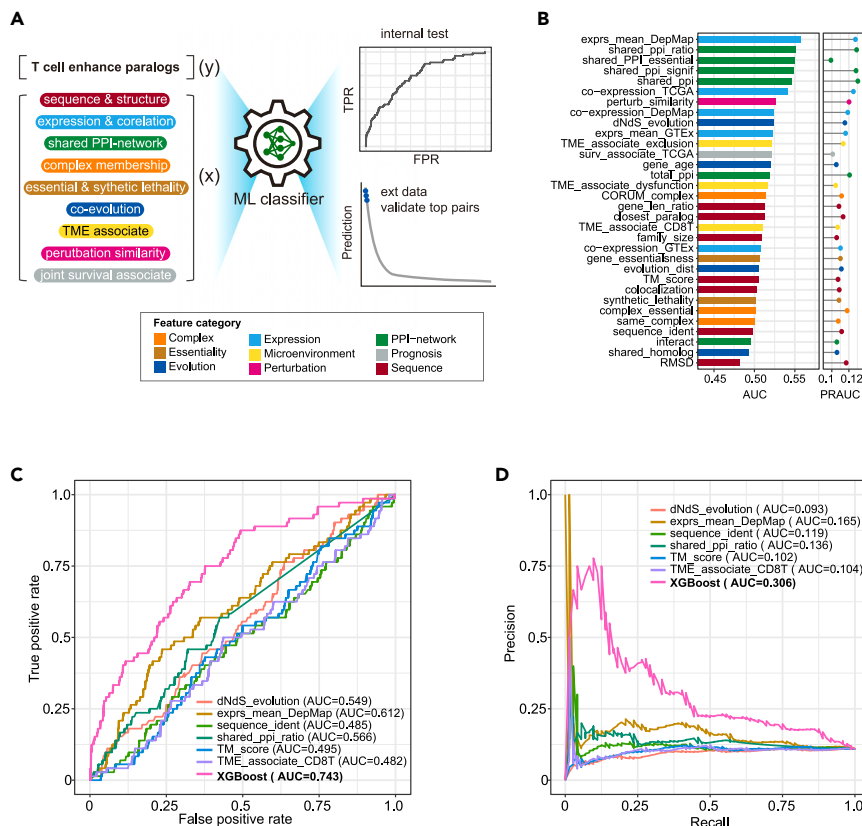
(B) Experimental design for the validation of selected paralog pairs predicted to enhance T cell killing. Naive CD8a<sup>+</sup> T cells were isolated from the spleen of OT-I mice and stimulated with anti-mouse CD3/CD28 antibody for 2 days. Then, B16F10-Cas9-mCherry-Ova gene knockout cells were co-cultured with OT-I T cells at an E:T = 1:1 ratio. Tumor cell killing were tested at 24 h by flow cytometry.

(C) Bar chart of mean viability of B16F10 cells with *Adam10* and *Adam15* single- or double-gene knockout(s) following culture with or without OT-1 T cells. Individual data points are shown as black dots. The *p* values come from two-sided heteroscedastic t tests (ns, *p* > 0.05, \**p* ≤ 0.05, \*\**p* ≤ 0.01, and \*\*\**p* ≤ 0.001).

(D) Bar chart of mean viability of *Ppp2r2a* and *Ppp2r2d* knockout B16F10 cells following culture with or without OT-1 T cells.

(E) Bar chart of mean viability of *Elf2* and *Etv6* knockout B16F10 cells following culture with or without OT-1 T cells.

(F) Bar chart of mean viability of *Rbm45* and *Rbms2* knockout B16F10 cells following culture with or without OT-1 T cells.



**Figure 3. Development and evaluation of XGBoost classifier for predicting T cell enhancing synergistic paralog pairs**

(A) Schematic for constructing a machine learning classifier to predict T cell-enhancing paralog pairs. (B) ROC AUC and PRC AUC measures of 32 paralog features for predicting paralog pairs that boost T cell killing. (C) ROC curves of XGBoost and top individual features in the internal test dataset. The XGBoost classifier outperformed the traditional individual features. (D) Precision recall curves of the XGBoost classifier and individual features.

positive pairs ( $n = 273$ ). The remaining 20% were saved for internal testing later (Figures S2A and S2B). Prior to building the ensemble model, our analysis revealed that certain individual features exhibited a stronger predictive effect in the training data. The top influential feature was gene expression abundance with the highest ROC AUC (area under receiver operating characteristic curve) values, indicating that gene expression itself could be a critical indicator of a paralog pair (Figure 3B). Notably, the features ranked second to fifth were four PPI-network features (Figure 3A), highlighting the crucial role of gene networks in determining whether two genes could interact in an immunotherapy context. Synthetic lethality was identified as a mid-level feature, suggesting that the collaborative roles of paralogs of cancer cells in response to immune context might differ from their roles in cell viability.

Next, the XGBoost classifier, an ensemble-boosted tree learner, was trained, utilizing 5-fold cross-validation for hyperparameter tuning. We then applied the optimal parameters to re-train the classifier using the entire training dataset (details in methods). Variable importance analysis revealed that complex, co-localization, and PPI features were the top-ranking features (Figure S2C). We then evaluated the XGBoost classifier's performance using testing data. Notably, the XGBoost classifier significantly outperformed all single features used as a baseline for comparison (Figure 3C). Specifically, the classifier achieved an AUC of 0.743, surpassing the best-performing individual-feature average paralog gene expression in cancer cell lines, which had an AUC of 0.612. Considering the imbalanced nature of the training dataset, with more negative pairs than positive ones,

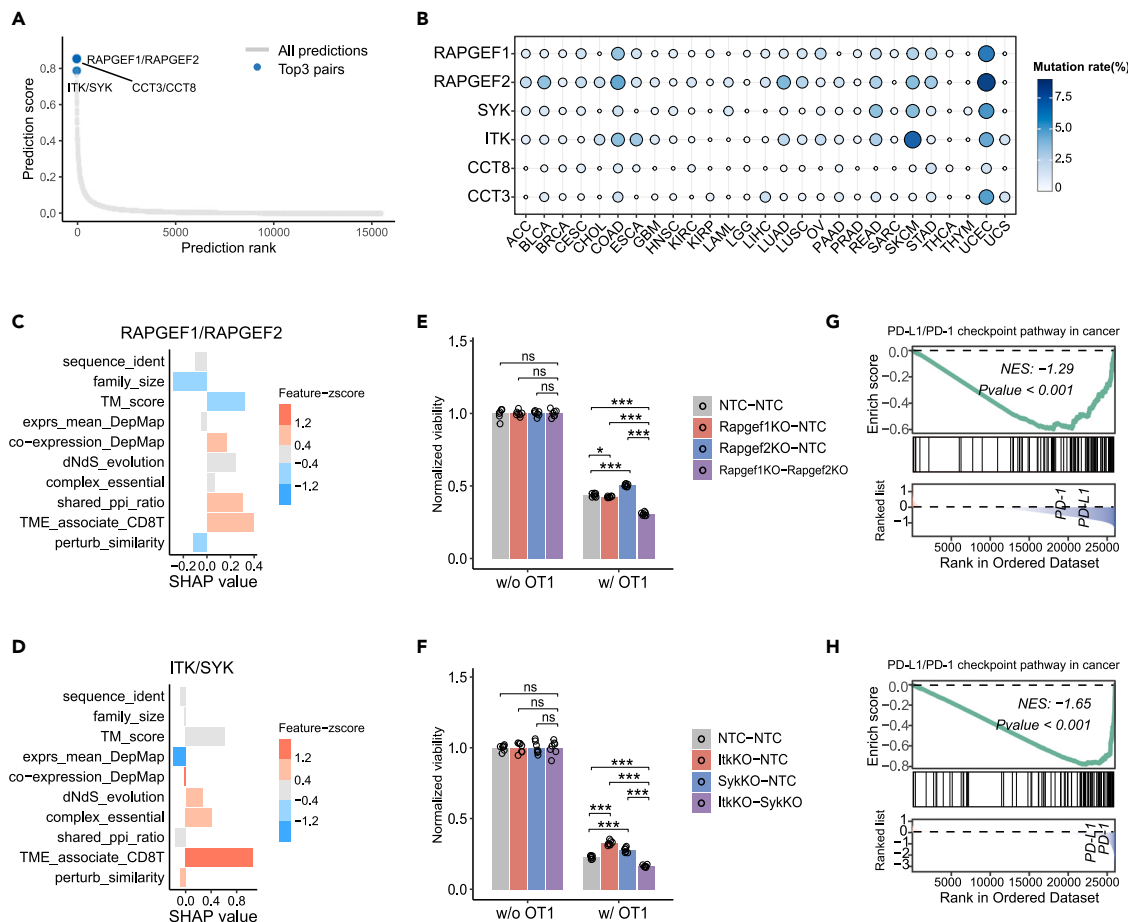
we also assessed the classifier's performance using the precision-recall curve (PRC; Figure 3D). Similarly, the XGBoost classifier showed better performance compared to all individual features. Overall, the results underscore the enhanced predictive capability of our XGBoost classifiers in effectively capturing interactions between features and non-linear relationships, thereby enhancing T cell killing potential.

Furthermore, we evaluated the classifier's performance on paralog pairs in cells other than B16F10 cells. ROC analysis revealed that the classifier trained on B16F10 cell data could not predict paralog pairs enhancing T cell killing in other cell types, except for CT26 colon cells (Figures S3A–S3E), indicating inherent heterogeneity among cancer cell models.

### Interpreting the prediction of immunotherapy-favorable paralog pairs from the machine learning classifier

Using the XGBoost classifier, we comprehensively analyzed the probability of enhancing T cell cytotoxicity for additional paralog pairs outside the training and internal testing sets. By filtering with criteria including measurements in structural similarity, synthetic lethality, and survival associations and excluding pairs from the training and testing sets, we obtained 15,414 paralog pairs out of the total 139,654 pairs for prediction. Each paralog pair was calculated for its potential impact, assigned a probability score, and ranked accordingly. The Shapley additive explanations (SHAP) tree explainer method assessed the contribution of each feature to the final predictions.

The paralog pairs were ranked by their prediction scores, and subsequent pathway enrichment analysis revealed that genes in the top 5% of paralog pairs were significantly enriched in functions related to protein phosphorylation and ubiquitination modification (Figure S4A). The top three ranked predicted paralog pairs were *Rapgef1+Papgef2*, *Cct8+Cct3*, and *Syk+Itk* (Figure 4A; Table S5). Notably the expression of *Syk* in cancer cells has shown to be associated with both tumor promotion and suppression.<sup>22</sup> *Itk* has been reported to enhance ICB response in solid tumors,<sup>23</sup> and *Rapgef1* was identified as essential for melanocyte growth through a CRISPR proliferation screen.<sup>24</sup> Analysis



**Figure 4. Visualization and validation of top 3 predicted paralogs from XGBoost classifier**

(A) Paralog pairs with the top 3 prediction scores are displayed in a scatterplot.

(B) Mutation profile of top 3 paralog pairs *RAPGEF1/RAPGEF2*, *SYK/ITK*, and *CCT8/CCT3* in The Cancer Genome Atlas (TCGA) cohorts.

(C and D) SHAP profile for *RAPGEF1/RAPGEF2* (C) and *SYK/ITK* (D).

(E and F) Bar chart of mean viability of *Rapgef1/Rapgef2* (E) and *Itk/Syk* (F) knockout treated B16F10 melanoma cells following culture with or without OT-1 T cells. The *p* values come from two-sided heteroscedastic *t* tests. (ns, *p* > 0.05, \**p* ≤ 0.05, \*\**p* ≤ 0.01, and \*\*\**p* ≤ 0.001).

(G and H) Gene sets enrichment analysis (GSEA) revealed that patients with low expression of *RAPGEF1/RAPGEF2* (G) and *SYK/ITK* (H) were associated with downregulation of PD-L1/PD-1 signaling pathway. The GSEA was performed using the differential expression output between patients with low (<25%) expression of both paralog genes versus all remaining patients.

of The Cancer Genome Atlas (TCGA) cohort revealed that four genes (*Rapgef1*, *Rapgef2*, *Itk*, and *Syk*) had mutation rates exceeding 1% in the pan-cancer cohort. At the individual cancer-type level, four genes (*Cct3*, *Rapgef1*, *Rapgef2*, and *Syk*) had mutation rates exceeding 5% in highly mutated endometrial carcinoma, while in melanoma, three genes (*Rapgef1*, *Rapgef2*, and *Syk*) had mutation rates exceeding 3%; *Itk* had a rate of 7.3% (Figure 4B). The SHAP profile revealed differential feature attributions among the top pairs. For example, sequence identity and paralog family size showed negative contributions to the prediction in *Rapgef1+Rapgef2* (Figure 4C) and *Itk+Syk* (Figure 4D) but positive contributions in *Cct3+Cct8* (Figure S4B).

To evaluate the synergistic effect of the paralog genes identified by the XGBoost classifier, we conducted DKO-B16F10-OT1 T cell co-culture assays. Knocking out *Rapgef1* alone did not affect T cell-mediated tumor cell killing, while knocking out *Rapgef2* alone increased tumor cell resistance to T cell killing.

However, a combined KO of *Rapgef1* and *Rapgef2* significantly improved T cell efficiency in killing tumor cells (Figures 4E and S4C). A similar pattern was observed with the *Syk+Itk* gene pair: knocking out individual genes increased resistance to T cell-mediated killing, but a double knockout enhanced T cell killing activity (Figures 4F and S4D). In contrast, double knockout of the *Cct8+Cct3* gene pair impaired T cell killing ability against tumor cells (Figures S4E and S4F). Synergy analysis revealed that *Rapgef1+Rapgef2* and *Syk+Itk* exhibited positive synergy effects (see the methods section and Table S4).

To further investigate why *Cct3/Cct8* might fail while the other two pairs work, we performed gene set enrichment analysis (GSEA) on TCGA-SKCM data, comparing patients with low expression in both paired genes (≤25%) to those with high expression (≥75%). We observed significant downregulation of the PD-L1/PD-1 signaling pathway in patients with melanoma with low *Rapgef1+Rapgef2* (Figure 4G) and *Itk+Syk*



(Figure 4H) expression. Downregulation of the PD-L1/PD-1 signaling pathway augments T cell killing by interrupting inhibitory signals that suppress T cell activity. Conversely, patients with low *Cct3+Cct8* expression (Figure S4G) showed insignificant upregulation of this pathway compared to others. This finding suggests that the failure of *Cct3+Cct8* to enhance T cell cytotoxicity may be due to the activation of other immunosuppressive pathways and the elevated expression of inhibitory molecules.

## DISCUSSION

Although numerous CRISPR screen studies aim to identify novel cancer cell targets for enhancing cancer immunotherapies, few have focused on functional gene pairs, thus limiting the development of combination strategies. To address this, we developed a computational enrichment-based approach, pSSEA, to identify potential paralog gene pairs using genome-wide CRISPR screen datasets by combining sgRNAs from two paralogs. We demonstrated that pSSEA enables the identification of cancer-intrinsic paralog pairs that can synergistically enhance T cell killing. Subsequently, we constructed an ensemble-learning XGBoost classifier to predict additional cancer-intrinsic, immunotherapy-aiding paralog pairs and experimentally tested the top predictions. Notably, all analyses in this study were based on CRISPR screen data. We observed that only a small subset of the screen-derived T cell-enhancing paralog pairs demonstrated a combined effect in predicting ICB responses when using patient-derived transcriptomic data (data not shown). This suggests that these two data types may capture different aspects of immunotherapy-related signals.

We developed an enrichment-based approach pSSEA to identify key paralog gene pairs, building on two main considerations: (1) single-gene screens well correlated with combinatorial screens effects from previously studies<sup>9,14,16</sup> and (2) although combinatorial targeting remains a compelling area of interest, designing a library that comprehensively covers all possible gene pairs is technically unfeasible.

## Limitations of the study

Limitations of this study should also be acknowledged. First, the paralog pairs selected and used for modeling were not derived from actual combinatorial screens targeting immunotherapy. The absence of ground-truth data may influence the accurate identification of true paired hits, potentially impacting the model's performance. Secondly, we did not incorporate a synergistic effect module, as used in other combinatorial screening studies,<sup>13,25,26</sup> which defines synergy as the difference between a gene pair's observed effects and the expected sum of their individual effects. While concerns remain about the comparability of enrichment scores generated from sgRNA sets of varying sizes, we plan to further investigate this issue and evaluate the feasibility of integrating similar functionality. Nevertheless, this study has the potential to offer valuable insights into the rapidly advancing field of immunotherapy and harness the power of well-established single-gene screens to uncover meaningful gene pair interactions.

In summary, we envision our work providing novel methods to leverage genome-wide screen data for selecting combination

targets in cancer immunotherapy. We believe that the computational framework in our study could be adapted to prioritize non-paralogous protein-coding gene pairs, broadening its applicability beyond paralogs.

## METHODS

### Paralog pair information

Information on protein-coding paralog pairs was obtained from the Ensembl Compara database via biomaRt (*hsapiens\_gene\_ensembl*, v.102) using custom R scripts. Duplicated paralog pairs were filtered out. A total of 139,654 protein-coding gene paralog pairs, covering 16,202 protein-coding genes, were used for the analyses in this study.

### Collect features for paralog pairs

#### Sequence similarity

Amino acid sequence of paralog genes were obtained from Ensembl (release 102).<sup>27</sup> Paralog sequence similarities were measured using BLASTP command line tools.<sup>28</sup> The longest translated peptide was used for genes with multiple transcripts. Gene pair sequence identity, paralog family size, and whether they are the closest pair in the same gene family were included as sequence-related features for paralog pairs.

We also used high-resolution structure PDB files from <https://alphafold.ebi.ac.uk/>.<sup>29</sup> The structural similarity of paralogous proteins was measured using template modeling score (TM-score) software.<sup>30</sup> The TM-score and root-mean-square deviation (RMSD) were included as additional sequence-related features.

### Shared gene network, complex, and co-localization

We collected shared PPI features to characterize the role of paralogs in the gene interaction network, including total PPIs, shared PPIs, and two statistical measures: the Jaccard index and Fisher's exact test  $-\log_{10}(p \text{ value})$  for shared PPIs. Gene subcellular location data were obtained from the Human Protein Atlas (v.22.0, file: *subcellular\_location.tsv*). The co-localization of paralogous genes was also measured using the Jaccard index. The complex membership of paralog pairs and gene-protein complex membership data were obtained from CORUM<sup>31</sup> via <https://maayanlab.cloud/Harmonizome/>.

### Gene co-evolution

Gene ages were obtained from ProteinHistorian (<https://proteinhistorian.docpollard.org/>),<sup>32</sup> and we used the average age of the two paralogs as the age for the pair (Wagner age reconstruction algorithm). The conservation level of an individual gene was calculated as the number of species with orthologs (BLASTP e-value < 1-e-5, identity > 30%) out of 269 species from Ensembl genomes (v.102). The shared homologs of a gene pair were calculated as the number of species having both gene homologs. The ratio of the number of non-synonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site (Ka/Ks) for gene pairs was calculated using the codeml program in the PAML package (v.4.10.6).<sup>33,34</sup> We adopted the phylogenetic distance method developed by Tabach et al. to measure the co-evolution of a pair of genes.<sup>35</sup> A non-negative matrix factorization (NMF)-derived distance considering phylogenetic tree

information was calculated using the phylogenetic profiles of the paralogous genes.<sup>36,37</sup>

### Essentiality and synthetic lethality

Paralogs with available evidence of essentiality from *S. pombe* and *S. cerevisiae* ortholog data from OGEE v.3 were marked as essential pairs.<sup>38</sup> We adopted Kegel's method to identify synthetic lethality pairs<sup>6</sup> using DepMap CRISPR data (<https://depmap.org/portal>, 23Q2 release). Paralog pairs were considered synthetic if the A1 gene was an essential gene (Chronos score<sup>39</sup> < −0.6 in 1% or 10% of cell lines), and its dependency was significantly associated with A2 loss status with a *p* value less than 0.05.

### Co-expression pattern

Gene pair co-expression was measured using Spearman correlation coefficient  $\rho$  at multiple levels: cancer cell lines, normal tissues, and cancer scenarios. Cancer Cell Line Encyclopedia (CCLE) gene expression data were downloaded from DepMap.<sup>40</sup> Tissue-specific gene expression profiles from healthy donors were downloaded from Genotype-Tissue Expression (GTEx; <https://gtexportal.org/>) public releases.<sup>41</sup> Cancer-specific gene expression profiles for 33 cancer types were downloaded from TCGA GDC Portal (<https://portal.gdc.cancer.gov/>).<sup>42</sup>

### Perturbation similarity

Due to the limited number of gene perturbation experiments in Connectivity Map (L1000, <https://clue.io/data/>),<sup>43</sup> we conducted pseudo perturbation using TCGA mRNA sequence data to mimic the transcriptomic gene expression changes associated with single-gene manipulation. Practically, the differential expression profile of specific gene inhibition was calculated as the log fold change in gene expression between patients with expression levels in the lowest 25th percentile versus those in the highest 75th percentile of the query gene. The perturbation similarity of paralogous genes was measured using recovery AUC with the paralog gene differential expression profiles.

### Microenvironment associations

The TME plays an important role in cancer progression and treatment outcome. We attempted to uncover features quantifying the association of the paralogs with multiple microenvironment indicators. CD8<sup>+</sup> T cell infiltration was measured by the single-stranded GSEA (ssGSEA) method.<sup>20</sup> We calculated T cell exclusion and dysfunction scores using another software, the TIDE algorithm developed by Jiang et al.<sup>44</sup> The association with gene pairs and microenvironmental characteristics was calculated by Kendall's rank correlation test.<sup>45</sup>

### Survival associations

To determine gene pairs with a synergistic effect on survival outcomes, we calculated the survival associations to assess whether a paralog pair acts in concert to influence cancer prognosis. First, patients were categorized into three groups based on the expression levels of the paired paralog genes: groups were designated as 1 for both genes showing low expression, 3 for both showing high expression, and 2 for any other expression combination. The survival association for paralog pairs was then calculated using the formula

$$\text{survival association} = -\log_{10}(p) * \text{sign}(\beta_{\text{gene1}} * \beta_{\text{gene2}}).$$

In this formula, *p* refers to the *p* value from the Cox regression analysis of the paralog group, adjusted for sex and age, while  $\beta_{\text{gene1}}$  and  $\beta_{\text{gene2}}$  represent the exponentiated coefficients of the individual genes from the Cox proportional hazards model. Clinical data and genomics data for TCGA-SKCM were downloaded from the GDC Portal (<https://portal.gdc.cancer.gov/>).<sup>46</sup>

### XGBoost ensemble classifier

The popular supervised-learning algorithm XGBoost—a gradient boosting ensemble learner with regularization parameters—was used in this study (implementation in the XGBoost package).<sup>47</sup> We performed a grid search based on 5-fold cross-validation of our paralog pair dataset to explore the space of potential hyperparameters for the XGBoost classifier. The parameters used in the final model are as follows: objective = reg: logistic, n\_estimators = 400, max\_depth = 10, min\_child\_weight = 1, min\_split\_loss = 0.1, subsample = 1, colsample\_bytree = 1, learning\_rate = 0.01, and random\_state = 8 (for reproducibility). All parameters not listed used default settings.

### Performance evaluation

The ROC AUC and PRC AUC for the feature values as well as the classifiers were computed to evaluate the performance of individual features and trained models.

### Lentivirus purification

For lentivirus production, 20  $\mu\text{g}$  of lenti-U6-sgRNA-EFS-GFP-WPRE or 20  $\mu\text{g}$  of lenti-U6-sgRNA-EFS-Puro-WPRE, 10  $\mu\text{g}$  of pMD2.G, and 15  $\mu\text{g}$  of psPAX2 were co-transfected into LentiX-293 cells plated in a 150 mm dish at 80%–90% confluency using 130  $\mu\text{g}$  polyethylenimine (PEI). 6 h later, the media were replaced with fresh DMEM+10% FBS. Virus supernatant was collected 48 h post-transfection and centrifuged at 1,500g for 15 min to remove the cell debris. The virus supernatant was concentrated by Lenti-X concentrator at 4°C for 30 min (1 vol of Lenti-X concentrator with 3 vol of supernatant), followed by centrifugation at 1,500g for 45 min at 4°C. Finally, the virus was resuspended in DMEM, aliquoted, and stored at −80°C.

### Generation of single-gene-knockout or double-gene-knockout cells

B16F10 cells stably expressing Cas9 were generated by transducing B16F10 cells with lentiviral EF1a-NLS-Cas9-2A-Blast-WPRE, followed by 5 days of selection under 20  $\mu\text{g}/\text{mL}$  blasticidin. B16F10-Cas9 cells were further transfected with lentiviral EF1a-mCherry-2A-OVA-WPRE and sorted for mCherry-positive cells to generate B16f10-Cas9-mCherry-OVA cells. The two gRNAs targeting each gene within the paralog pair were transduced into the B16F10-Cas9-mCherry-Ova cells simultaneously. gRNA virus-infected cells were cultured at 37°C for more than 24 h, and then GFP-positive cells were sorted and selected under 5  $\mu\text{g}/\text{mL}$  puromycin for 3 days to generate the double-knockout cells.

### T7E1 assay

T7E1 was used to estimate gRNA cutting efficiency. In brief, gDNA was extracted by using Genomic DNA prep with QuickExtract (QE) buffer. PCR amplification of the genomic regions flanking the CRISPR RNA (crRNA) was performed using the primers listed in Table S3. Using Phusion Flash High Fidelity Master Mix (Thermo Fisher Scientific), the thermocycling parameters for PCR were 98°C for 1 min; 35 cycles of 98°C for 1 s, 60°C for 5 s, and 72°C for 15 s; and 72°C for 2 min. The PCR amplicons were then used for T7E1 assays according to the manufacturer's protocol.

### Naive OT-I CD8a+ T cell isolation and co-culture assay

The naive OT-I CD8a+ T cells were isolated using the Naive CD8a+ T Cell Isolation Kit (Miltenyi Biotec) following the manufacturer's instructions. Naive CD8a+ T cells were isolated from the spleen of OT-I mice and stimulated with anti-mouse CD3/CD28 antibody for 2 days. For the tumor cell and OT-I T cell co-culture assay, B16F10-Cas9-mCherry-Ova gene KO cells were co-cultured with OT1 cells at effector-to-target cell ratios (E:T) = 1:1 ratios. Tumor cell killing was tested at 24 h by flow cytometry. LIVE/DEAD near-infrared (IR) was diluted at 1:1,000 to identify the dead cells. Tumor cells were identified as mCherry-positive cells.

### Synergy analysis

Statistical analysis for evaluating synergistic interactions between paralog pairs was calculated by the SynergyFinder R package<sup>48</sup> using four models, including the highest single agent (HSA), Bliss, Loewe, and zero interaction potency (ZIP) models. The *p* values for the average synergy score were computed using bootstrapping of the dose-response matrix. We also adopted the dLFC method to measure the additive effect of gene pairs.<sup>13</sup> The dLFC was calculated as the observed mean relative killing rate of the DKO minus the expected sum of the relative killing rates of the individual SKO. The relative killing rate was determined by subtracting the values from the NTC group.

### RESOURCE AVAILABILITY

#### Lead contact

Requests for information and resources used in this article should be addressed to the lead contact, Sidi Chen ([sidi.chen@yale.edu](mailto:sidi.chen@yale.edu)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

The source code for processed data and modeling is available under GitHub (<https://github.com/cpdong/imParalog>) and Zenodo (<https://doi.org/10.5281/zenodo.14759518>)<sup>49</sup> and included in this article and its supplemental information files. An interactive web application of this study is deployed at <https://sidichenlab.shinyapps.io/imparalog> for free visiting and querying. Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

### ACKNOWLEDGMENTS

We thank all members of the Chen laboratory, as well as various colleagues at Yale Genetics, SBI, CSBC, MCGD, Immunobiology, BBS, YCC, YSCC, and CBDS, for assistance and/or discussions. We thank various Yale Core Facil-

ities for technical support. S.C. is supported by NIH (R33CA281702), DoD (W81XWH-17-1-0235, W81XWH-20-1-0072, W81XWH-21-1-0514, and HT94252310472), the Cancer Research Institute Lloyd J. Old STAR Award (CRI4964), the Alliance for Cancer Gene Therapy, the Sontag Foundation (DSA), the Pershing Square Sohn Cancer Research Alliance, Dexter Lu, the Ludwig Family Foundation, the Blavatnik Family Foundation, and the Chenevert Family Foundation. C.D. is supported by the Boehringer Ingelheim Biomedical Data Science Fellowship. N.V. is supported by the American Board of Radiology's B. Leonard Holman Research Pathway Fellowship and ASTRO Seed Grant.

### AUTHOR CONTRIBUTIONS

S.C., D.F., and H.Z. contributed to the conception of the study and offered method guidance. C.D. and F.Z. designed the algorithm, performed formal analysis and visualization, and drafted the manuscript. E.H., P.R., N.V., X.Z., and J.C. provided valuable suggestions for the experiments and manuscript. All authors contributed to the review and editing of the manuscript.

### DECLARATION OF INTERESTS

D.F. and J.C. are employers of Boehringer Ingelheim Pharmaceuticals. S.C. is a (co)founder of EvolveImmune Tx, Cellinfinity Bio, MagicTime Med, and Chen Consulting, all of which are unrelated to this study.

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2025.101184>.

Received: June 29, 2024

Revised: November 4, 2024

Accepted: January 29, 2025

Published: February 25, 2025

### REFERENCES

- Sharma, P., Hu-Lieskovan, S., Wargo, J.A., and Ribas, A. (2017). Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell* 168, 707–723.
- Wang, G., Chow, R.D., Zhu, L., Bai, Z., Ye, L., Zhang, F., Renauer, P.A., Dong, M.B., Dai, X., Zhang, X., et al. (2020). CRISPR-GEMM pooled mutagenic screening identifies KMT2D as a major modulator of immune checkpoint blockade. *Cancer Discov.* 10, 1912–1933.
- Madani Tonekaboni, S.A., Soltan Ghoraie, L., Manem, V.S.K., and Haibe-Kains, B. (2018). Predictive approaches for drug combination discovery in cancer. *Briefings Bioinf.* 19, 263–276.
- Magen, A., Sahu, A.D., Lee, J.S., Sharmin, M., Lugo, A., Gutkind, J.S., Schäffer, A.A., Rupp, E., and Hannonhall, S. (2019). Beyond synthetic lethality: charting the landscape of pairwise gene expression states associated with survival in cancer. *Cell Rep.* 28, 938–948.e936.
- Parrish, P.C.R., Thomas, J.D., Gabel, A.M., Kamapurkar, S., Bradley, R.K., and Berger, A.H. (2021). Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Rep.* 36, 109597.
- De Kegel, B., Quinn, N., Thompson, N.A., Adams, D.J., and Ryan, C.J. (2021). Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. *Cell Syst.* 12, 1144–1159.e6.
- Park, J.S., Gazzaniga, F.S., Wu, M., Luthens, A.K., Gillis, J., Zheng, W., LaFleur, M.W., Johnson, S.B., Morad, G., Park, E.M., et al. (2023). Targeting PD-L2-RGMB overcomes microbiome-related immunotherapy resistance. *Nature* 617, 377–385.
- Moreno-Lama, L., Galindo-Campos, M.A., Martínez, C., Comerma, L., Vazquez, I., Vernet-Tomas, M., Ampurdanés, C., Lutfi, N., Martín-Caballero, J., Dantzer, F., et al. (2020). Coordinated signals from PARP-1 and PARP-2 are required to establish a proper T cell immune response to breast tumors in mice. *Oncogene* 39, 2835–2843.

9. Park, J.J., Codina, A., Ye, L., Lam, S., Guo, J., Clark, P., Zhou, X., Peng, L., and Chen, S. (2022). Double knockout CRISPR screen for cancer resistance to T cell cytotoxicity. *J. Hematol. Oncol.* **15**, 172–175.
10. Zhao, D., Badur, M.G., Luebeck, J., Magaña, J.H., Birmingham, A., Sasik, R., Ahn, C.S., Ideker, T., Metallo, C.M., and Mali, P. (2018). Combinatorial CRISPR-Cas9 metabolic screens reveal critical redox control points dependent on the KEAP1-NRF2 regulatory axis. *Mol. Cell.* **69**, 699–708.e7.
11. Tang, S., Wu, X., Liu, J., Zhang, Q., Wang, X., Shao, S., Gokbag, B., Fan, K., Liu, X., Li, F., et al. (2022). Generation of dual-gRNA library for combinatorial CRISPR screening of synthetic lethal gene pairs. *STAR Protoc.* **3**, 101556.
12. Ito, T., Young, M.J., Li, R., Jain, S., Wernitznig, A., Krill-Burger, J.M., Lemke, C.T., Monducci, D., Rodriguez, D.J., Chang, L., et al. (2021). Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. *Nat. Genet.* **53**, 1664–1672.
13. Dede, M., McLaughlin, M., Kim, E., and Hart, T. (2020). Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol.* **21**, 262.
14. Chow, R.D., Wang, G., Ye, L., Codina, A., Kim, H.R., Shen, L., Dong, M.B., Errami, Y., and Chen, S. (2019). In vivo profiling of metastatic double knockouts through CRISPR–Cpf1 screens. *Nat. Methods* **16**, 405–408.
15. Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178.
16. Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N., et al. (2018). Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189.
17. Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463–474.
18. Shao, S., Li, S., Tang, S., Fan, K., and Li, L. (2024). CRISPR Double Knockout Library Reveals Synthetic Lethal Gene Pairs in Triple-Negative Breast Cancer. Preprint at bioRxiv. <https://doi.org/10.1101/2024.2005.2014.594157>.
19. Lawson, K.A., Sousa, C.M., Zhang, X., Kim, E., Akthar, R., Caumanns, J.J., Yao, Y., Mikolajewicz, N., Ross, C., Brown, K.R., et al. (2020). Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature* **586**, 120–126.
20. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* **14**, 1–15.
21. Edelman, E., Porrello, A., Guinney, J., Balakumaran, B., Bild, A., Febbo, P.G., and Mukherjee, S. (2006). Analysis of sample set enrichment scores: assaying the enrichment of sets of genes for individual samples in genome-wide expression profiles. *Bioinformatics* **22**, e108–e116.
22. Aguirre-Ducier, A., Gianino, N., Villarroel-Espindola, F., Desai, S., Tang, D., Zhao, H., Syrigos, K., Trepicchio, W.L., Kannan, K., Gregory, R.C., and Schalper, K.A. (2022). Tumor cell SYK expression modulates the tumor immune microenvironment composition in human cancer via TNF- $\alpha$  dependent signaling. *J. Immunother. Cancer* **10**, e005113.
23. Zhao, M., Li, L., Kiernan, C.H., Castro Eiro, M.D., Dammeyer, F., van Meurs, M., Brouwers-Haspels, I., Wilmsen, M.E.P., Grashof, D.G.B., van de Werken, H.J.G., et al. (2023). Overcoming immune checkpoint blockade resistance in solid tumors with intermittent ITK inhibition. *Sci. Rep.* **13**, 15678.
24. Thakur, R., Xu, M., Thornock, A., Sowards, H., Long, E., Rheling, T., Funderburk, K., Yin, J., Hennessey, R., Chari, R., et al. (2023). Integrative analysis of 3D chromatin organization at GWAS loci identifies RAGEF1 as a melanoma susceptibility gene. *Cancer Res.* **83**, 5239.
25. Ward, H.N., Aregger, M., Gonatopoulos-Pournatzis, T., Billmann, M., Ohsumi, T.K., Brown, K.R., Blencowe, B.J., Moffat, J., and Myers, C.L. (2021). Analysis of combinatorial CRISPR screens with the Orthrus scoring pipeline. *Nat. Protoc.* **16**, 4766–4798.
26. Zamanighomi, M., Jain, S.S., Ito, T., Pal, D., Daley, T.P., and Sellers, W.R. (2019). GEMINI: a variational Bayesian approach to identify genetic interactions from combinatorial CRISPR screens. *Genome Biol.* **20**, 137.
27. Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* **19**, 327–335.
28. Boratyn, G.M., Camacho, C., Cooper, P.S., Coulouris, G., Fong, A., Ma, N., Madden, T.L., Matten, W.T., McGinnis, S.D., Merezuk, Y., et al. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **41**, W29–W33.
29. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444.
30. Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710.
31. Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2023). CORUM: the comprehensive resource of mammalian protein complexes–2022. *Nucleic Acids Res.* **51**, D539–D545.
32. Capra, J.A., Williams, A.G., and Pollard, K.S. (2012). ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput. Biol.* **8**, e1002567.
33. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
34. Harikrishnan, S.L., Pucholt, P., and Berlin, S. (2015). Sequence and gene expression evolution of paralogous genes in willows. *Sci. Rep.* **5**, 18662.
35. Tabach, Y., Golan, T., Hernández-Hernández, A., Messer, A.R., Fukuda, T., Kouznetsova, A., Liu, J.G., Lilienthal, I., Levy, C., and Ruvkun, G. (2013). Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. *Mol. Syst. Biol.* **9**, 692.
36. Tabach, Y., Billi, A.C., Hayes, G.D., Newman, M.A., Zuk, O., Gabel, H., Kamath, R., Yacoby, K., Chapman, B., Garcia, S.M., et al. (2013). Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* **493**, 694–698.
37. Sherill-Rofe, D., Rahat, D., Findlay, S., Mellul, A., Guberman, I., Braun, M., Bloch, I., Lalezari, A., Samiei, A., Sadreyev, R., et al. (2019). Mapping global and local coevolution across 600 species to identify novel homologous recombination repair genes. *Genome Res.* **29**, 439–448.
38. Gurumayum, S., Jiang, P., Hao, X., Campos, T.L., Young, N.D., Korhonen, P.K., Gasser, R.B., Bork, P., Zhao, X.-M., He, L.-j., and Chen, W.H. (2021). OGEE v3: Online GENE Essentiality database with increased coverage of organisms and human cell lines. *Nucleic Acids Res.* **49**, D998–D1003.
39. Dempster, J.M., Boyle, I., Vazquez, F., Root, D.E., Boehm, J.S., Hahn, W.C., Tsherniak, A., and McFarland, J.M. (2021). Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol.* **22**, 343.
40. Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508.
41. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585.
42. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120.
43. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A next



- generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17.
44. Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., Li, Z., Traugh, N., Bu, X., Li, B., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558.
  45. Sen, P.K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **63**, 1379–1389.
  46. Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X., Noushmehr, H., Colaprico, A., and Papaleo, E. (2019). New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput. Biol.* **15**, e1006701.
  47. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.
  48. Zheng, S., Wang, W., Aldahdooh, J., Malyutina, A., Shadbahr, T., Tanoli, Z., Pessia, A., and Tang, J. (2022). SynergyFinder plus: toward better interpretation and annotation of drug combination screening datasets. *Dev. Reprod. Biol.* **20**, 587–596.
  49. Dong, C., Zhang, F., He, E., Ren, P., Verma, N., Zhu, X., Feng, D., Cai, J., Zhao, H., and Chen, S. (2025). Sensitive detection of synthetic response to cancer immunotherapy driven by gene paralog pairs. Zenodo. <https://doi.org/10.5281/zenodo.14759518>.