

# Discovery of synthetic lethal interactions from large-scale pan-cancer perturbation screens

Received: 22 October 2019

Accepted: 29 November 2022

Published online: 14 December 2022

 Check for updates

Sumana Srivatsa <sup>1,2,7</sup>, Hesam Montazeri <sup>3,7</sup>, Gaia Bianco <sup>4,7</sup>,  
Mairene Coto-Llerena <sup>4,5</sup>, Mattia Marinucci <sup>4</sup>, Charlotte K. Y. Ng <sup>2,6</sup>,  
Salvatore Piscuoglio <sup>4,5</sup> ✉ & Niko Beerenwinkel <sup>1,2</sup> ✉

The development of cancer therapies is limited by the availability of suitable drug targets. Potential candidate drug targets can be identified based on the concept of synthetic lethality (SL), which refers to pairs of genes for which an aberration in either gene alone is non-lethal, but co-occurrence of the aberrations is lethal to the cell. Here, we present SLIdR (Synthetic Lethal Identification in R), a statistical framework for identifying SL pairs from large-scale perturbation screens. SLIdR successfully predicts SL pairs even with small sample sizes while minimizing the number of false positive targets. We apply SLIdR to Project DRIVE data and find both established and potential pan-cancer and cancer type-specific SL pairs consistent with findings from literature and drug response screening data. We experimentally validate two predicted SL interactions (*ARIDIA-TEAD1* and *AXINI-UR11*) in hepatocellular carcinoma, thus corroborating the ability of SLIdR to identify potential drug targets.

Synthetic lethality (SL) refers to gene pairs for which an aberration in either gene alone does not affect cell viability, but aberrations in both genes are fatal to the cell. Key to exploiting SL in cancer therapy is the identification of a targetable dependent gene (SL partner) for a given genetically altered gene, such that the two genes form an SL pair (Fig. 1a). A classical example of SL in cancer therapy is the use of *PARP* inhibitors in *BRCA*-mutated cancers. The *BRCA1/2* genes involved in DNA double-strand break repair are often mutated in breast and ovarian cancers<sup>1–3</sup>, and hence such cancer cells rely on alternate DNA repair processes. *PARP1* plays a central role in these alternate DNA repair mechanisms<sup>4,5</sup>, and therefore inhibiting *PARP* results in catastrophic double-strand breaks during replication, ultimately leading to cancer cell death<sup>6,7</sup>.

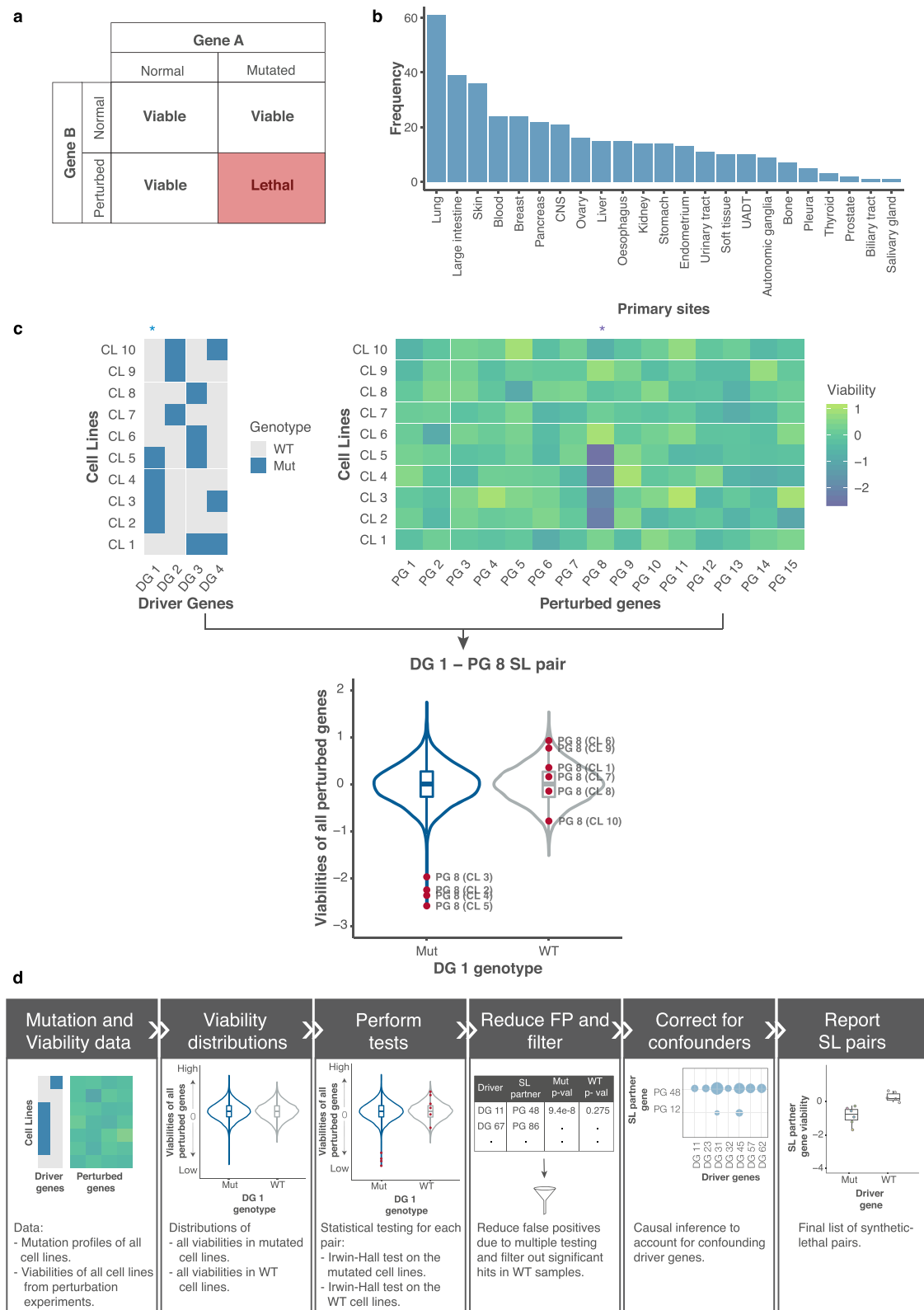
In recent years, high-throughput experiments have enabled the generation of multi-omics observational data and large-scale

interventional data in various cancer types. Several computational methods have since been developed to identify and prioritize SL interactions from such diverse molecular data. Early methods involved the identification of mutually exclusive genetic alterations in functionally related genes. These approaches were either de novo and detected patterns in the mutational data<sup>8–11</sup>, or knowledge-driven and based on pathways or interaction networks<sup>12</sup>. Other computational methods for identifying SL interactions relied on human orthologues of yeast genetic interactions<sup>13</sup>, inter-species network models<sup>14</sup>, signaling networks<sup>15</sup>, and protein-protein interaction networks<sup>16</sup>.

Experimentally, large-scale perturbation screens based on siRNA, shRNA, CRISPR, or small molecules in cell lines have been conventionally used to identify SL interactions<sup>17–27</sup>. McDonald et al.<sup>27</sup> conducted a large-scale deep RNAi screen, entitled Project DRIVE, that targeted 7,837 genes in 398 Cancer Cell Line Encyclopedia<sup>28</sup> (CCLE)

<sup>1</sup>Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. <sup>2</sup>SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland.

<sup>3</sup>Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. <sup>4</sup>Visceral Surgery and Precision Medicine Research Laboratory, Department of Biomedicine, University of Basel, 4031 Basel, Switzerland. <sup>5</sup>Institute of Medical Genetics and Pathology, University Hospital Basel, 4031 Basel, Switzerland. <sup>6</sup>Department for BioMedical Research, University of Bern, 3008 Bern, Switzerland. <sup>7</sup>These authors contributed equally: Sumana Srivatsa, Hesam Montazeri, Gaia Bianco. ✉ e-mail: [s.piscuoglio@unibas.ch](mailto:s.piscuoglio@unibas.ch); [niko.beerenwinkel@bsse.ethz.ch](mailto:niko.beerenwinkel@bsse.ethz.ch)



**Fig. 1 | Overview and SLiDR workflow.** **a** Definition of a synthetic lethal pair: Aberration of gene A (driver gene) or knockdown of gene B (SL partner gene) alone does not affect the viability of the cell. However, the combination of mutated gene A and knockdown of gene B is lethal to the cell. **b** Distribution of the number of cell lines with copy number data from CCLE across 23 different cancer types used in this study. Since most cancers have small sample sizes, the figure indicates the necessity of developing a computational tool with sufficient statistical power on small sample size data. **c** Illustration of the SLiDR algorithm with a toy example. The data consists

of driver genes DG 1-DG 4 and perturbed genes PG 1-PG 15 across cell lines CL 1-CL 10. Cell lines CL 2-5 are mutated in the driver gene DG 1 (Mut), while the remaining cell lines are DG 1 wild-type (WT). Comparison of viability distributions across all perturbed genes PG 1-PG 15 in the DG 1 mutated (Mut) and WT cell lines shows that perturbation of gene 8 (PG 8) results in reduced viability only in CL 2-5 and not the WT cell lines. Thus, PG 8 is a SL partner of DG 1. **d** The computational pipeline illustrating the different steps performed to obtain the candidate SL pairs from mutation profiles and perturbation screen data.

models and provided a rich and robust dataset for the identification of SL pairs. However, the authors primarily analyzed gene interactions in a pan-cancer manner.

Significant efforts have been made in developing integrative computational methods that harness both multi-omics observational data and genetic perturbation screening data. Jerby-Arnon et al.<sup>29</sup> proposed DAISY, a method that predicts a global network of potential SL interactions in human cells from tumor copy number, expression data, and shRNA screening data from cell lines. Similarly, Sinha et al.<sup>30</sup> presented MiSL, an algorithm based on Boolean implications to prioritize mutation-specific SL partners for different cancer types from large pan-cancer multi-omics and perturbation screening data. ISLE is another statistical approach that uses lab-screened SL interactions as inputs and analyzes tumor molecular profiles, patient clinical data, and gene phylogeny relations to identify clinically relevant SL interactions<sup>31</sup>.

While all these methods mark a significant step forward, general prediction of SL interactions in cancer still remains a challenge. Most methods primarily rely on large sample sizes and multi-omics data, and thus are not suitable for rare cancer types. Furthermore, they use genetic perturbation screening data merely to refine the candidate list of SL pairs derived from analyzing multi-omics data. We hypothesize that such rich large-scale perturbation screens can be exploited further to obtain SL pairs.

Here, we present a statistical framework called SLIdR (Synthetic Lethal Identification in R), a rank-based statistical method for predicting SL pairs from perturbation screens in both pan-cancer and cancer type-specific settings. The predicted SL pairs are validated by large-scale drug-response profiles and literature evidence. Subsequently, by benchmarking SLIdR on simulated data and experimentally identified SL interactions, we demonstrate the improved predictive power and advantage of SLIdR in reducing the number of false positives even when the sample size is small. Finally, we validate two SLIdR predictions in hepatocellular carcinoma, namely *ARID1A-TEAD1* and *AXINI-URII*, through comprehensive experiments in patient-derived cell lines.

## Results

### SLIdR workflow

We developed SLIdR, a statistical framework for identifying SL interactions between a genetically altered gene and a perturbed gene from large-scale perturbation screens of cancer cell lines. We focused on significantly mutated genes reported by MutSig 2CV v3.1<sup>32,33</sup> for each cancer type and considered these genes to be genetically altered in cell lines if they were subject to non-synonymous mutations or deep deletions (see Methods). We collectively refer to these altered genes as driver genes and their alterations as mutations. SLIdR aims to find SL partners for such drivers from perturbation data. We applied SLIdR to the Project DRIVE dataset<sup>27</sup>, focusing on cell lines from CCL<sup>28</sup> across various cancer types with available copy number data (Fig. 1b).

In contrast to previous methods which perform statistical tests on the raw viability readouts<sup>27,29,31</sup>, SLIdR uses the normalized ranks of the viabilities across all perturbed genes, for each cell line, in order to increase statistical power for small sample sizes. For each driver gene, SLIdR first stratifies the cell lines into mutated and wild-type based on the mutation status of the driver gene. Subsequently, it tests, for each perturbed gene, whether the perturbation results in lower ranked viabilities in the mutated cell lines but not in the wild-type cell lines (Fig. 1c). SLIdR uses two Irwin-Hall tests to mine for such driver-perturbed SL gene pairs (see Methods). Cell lines with several co-occurring driver mutations can yield multiple SL pairs with the same perturbed gene. To identify the most likely SL pairs, we perform causal inference using matching-based potential outcome models. For a given candidate pair, we match the wild-type to mutated cell lines based on the other co-occurring mutations, thus achieving a covariate

balance. Finally, SLIdR compares the viabilities of the matched groups, and the significant SL pairs are reported (Fig. 1d; see Methods).

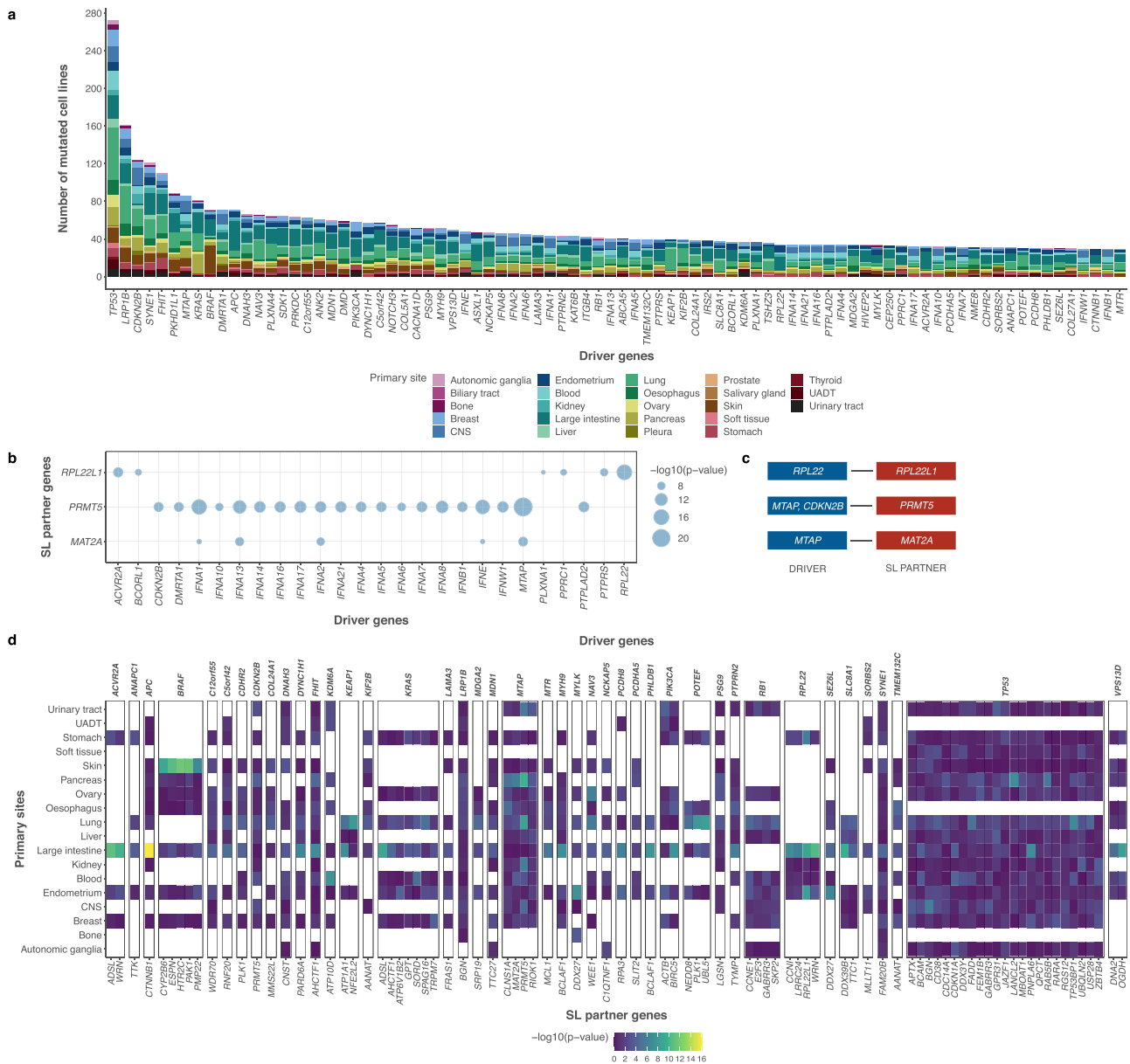
### Enrichment of pan-cancer SL interactions by SLIdR

To identify pan-cancer SL interactions, we first applied SLIdR to the DRIVE data in a pan-cancer setting. We identified 151 SL pairs (Supplementary Data 1) involving 84 driver genes (Fig. 2a). Out of the 151 SL pairs, five pairs involving bona fide driver genes *TP53*, *KRAS*, *BRAF*, *CTNNB1*, and *PIK3CA* exhibited oncogene-addiction, i.e., they paired with themselves as the SL partner gene. This proved to be an efficient quality check for our method as these are well-established drivers and their subsequent knockdown resulted in cellular mortality. We also found that some cell lines with several co-occurring mutations resulted in multiple driver genes pairing with the same SL partner (Fig. 2b; Supplementary Fig. 1a). For example, co-deletion of genes near p16 including *MTAP* and several interferons is common in several cancers, and subsequently all these drivers paired with *MAT2A* as the SL partner. Using causal inference (see Methods), we accounted for these co-occurring mutations and predicted the relevant driver genes for each SL partner (Fig. 2c; Supplementary Fig. 1b), resulting in 90 SL pairs across 42 driver genes (Fig. 2d).

Top predictions of SLIdR included *PRMT5*, *MAT2A*, and *RIOK1* as SL partners of *MTAP* which are all well-established vulnerable targets for *MTAP*-altered cells<sup>27,34</sup>. SLIdR also predicted *E2F3* and *SKP2* as SL partners of *RBI*<sup>27</sup>. Furthermore, *RPL22* showed lethality with its paralog *RPL22L1* confirming the findings of McDonald et al.<sup>27</sup>. *PIK3CA-BIRC5* was another reassuring pair as depletion of survivin (*BIRC5*) has been shown to have a pro-apoptotic effect in breast cancer cells with *PIK3CA* mutations<sup>35,36</sup>. In addition to established pairs, SLIdR also predicted several potential SL pairs, such as *KRAS-TRPM7* and *TP53*-specific SL partners, including *TP53BP1*, *USP28*, *DDX3*, and *PNPLA6*, which were further supported by evidence in the literature (Supplementary Data 2).

Next, we systematically validated the predicted pan-cancer SL pairs on the primary PRISM repurposing dataset, a large-scale drug-response profile from the DepMap consortium<sup>37</sup>. Out of the predicted SL interactions, 25 SL pairs had at least one matching drug compound in the screen. Stratifying the cell lines based on the mutation status of the driver gene for each pair, we tested if the drug targeting the corresponding SL partner gene was more effective in the mutated cell lines using a *t*-test (Supplementary Data 3). Reassuringly, we found 13 SL pairs, including *BRAF-CYP2B6*, *BRAF-HTR2C*, *KEAP1-NFE2L2*, *PIK3CA-BIRC5*, and *TP53-GABRR3*, with significant difference in drug response (significance level  $\alpha=0.1$ ). Of these 13 pairs, *BRAF-HTR2C*, *KEAP1-ATPIA1*, *NOTCH3-WEE1*, and *POTEF-PLK1* were significant after multiple testing correction ( $q$ -values  $\leq 0.2$ ) across 9 different candidate drug compounds (Supplementary Fig. 2, Supplementary Data 3). The lower coverage after FDR correction can be attributed to the heterogeneity and noise of these screens. As a control experiment, we performed permutation tests across 1000 sets of random gene pairs and found these validated hits to be significant (empirical  $p$ -value  $\leq 0.005$ ), thereby confirming the predictions from SLIdR (see Methods).

Pan-cancer analyses offer large sample sizes and the ability to identify shared targets across different cancer types. While the latter property is preferred especially from a therapeutic perspective, it is often difficult to identify such pairs due to the inherent genetic diversity in tumors based on their primary sites. To assess the differential sensitivities of the predicted pan-cancer hits based on primary sites, we computed the SLIdR  $p$ -values for these predicted pan-cancer SL pairs in subsets of cell lines grouped by primary sites. We found that a sizable fraction of the pan-cancer signals is indeed cancer type-specific (Fig. 2d). For example, SLIdR identified *NFE2L2* as the SL partner of the mutated *KEAP1*, both of which play an important role in cancer through Nrf2 pathway activation<sup>38</sup>. However, this SL interaction was largely driven by lung cancer samples (Fig. 2d), in accordance with



**Fig. 2 | Pan-cancer SLiDR predictions.** **a** Stacked barplot indicating the frequencies of 84 mutated driver genes across different cancer types. The number of mutated cell lines of a given cancer type may have an impact on the statistical power of the SLiDR framework. **b** Bubble-plot summarizing the significance ( $-\log_{10}(p\text{-value})$ ) of different driver genes (x-axis) pairing with the same SL partner gene (y-axis) as predicted by SLiDR in the pan-cancer analysis after filtering out false positives from multiple testing. The  $p$ -values are computed using one-sided IH-test. **c** Corresponding list of significant SL pairs after accounting for confounding

mutations and performing causal inference using matching-based potential outcome models. **d** Differential sensitivities of pan-cancer SL pairs in subsets of cell lines grouped by primary sites (y-axis). Each panel corresponds to a specific driver gene (x-axis top) and encapsulates the significance profiles of all its SL-partners (x-axis bottom) across various primary sites. Each column in a given panel depicts the significance profile of the SL pair in subsets of cell lines grouped by primary sites. The  $p$ -values are computed using one-sided IH-test.

Leiserson et al.<sup>39</sup> who also reported the pair to be mutually exclusive in their pan-cancer TCGA analysis largely due to lung cancer samples. Similarly, signals for *APC-CTNNB1* and all the *BRAF* associated SL interactions were mostly specific to large intestine and skin samples, respectively (Fig. 2d). Thus, although a considerable number of pan-cancer hits are consistent with previous findings, these examples show the need to identify cancer type-specific SL partners.

**Enrichment of cancer type-specific SL interactions by SLiDR**

To identify cancer type-specific SL interactions, we applied SLiDR to the DRIVE data for 17 cancer types and identified a total of 839 SL pairs over 233 unique driver genes (Supplementary Data 1). Out of the 233

drivers, 66 genes were mutated in more than one cancer type (Fig. 3a). However, the mutation profiles are diverse across cancer types, with *TP53* mutations being highly prevalent and observed in 81% of the cancer types, while well-known drivers such as *BRAF*, *APC*, and *PTEN* were distinctly associated with skin, large intestine, and endometrial cancers, respectively.

Upon extensive literature survey of the SL pairs, we identified 55 established and potential pairs with literature support (Fig. 3b; Supplementary Data 2). For example, SLiDR predicted *GATA3-ESR1* in breast cancer. *GATA3* is mutated in >10% of breast cancers and directly impacts *ESR1* enhancer accessibility, thereby altering binding potential and transcriptional targets in tumor cells<sup>40</sup>. Furthermore, *GATA3*



random gene pairs and found some evidence for these hits in pancreatic, skin, stomach, and renal cancers (empirical  $p$ -value = {0.024, 0.076, 0.002, 0.075}, respectively), which reached statistical significance after multiple testing correction for stomach cancer ( $q$ -value  $\leq 0.02$ ). These results demonstrate that SLIdR is capable of finding established and potential cancer type-specific SL pairs.

### SLIdR outperforms conventional tests for SL prediction

To demonstrate the advantage of relative ranking over raw viability scores and to benchmark the performance of SLIdR, we compared SLIdR with existing SL prediction methods in a simulation study. DAISY and ISLE are two popular methods that both use the Wilcoxon rank sum test for predicting candidate SL interactions from perturbation screens<sup>29,31</sup>. While both of these methods are multi-step and use other data sources to predict the final pairs, their reliance on raw viability scores rather than relative viabilities from shRNA screens limits the full utilization of such screens. In fact, the authors of DAISY<sup>29</sup> expressed that the data obtained from shRNA screens have low statistical power and hence, DAISY uses such screens only to refine a list of highly statistically significant SL interactions obtained from other data sources. By contrast, we show here that these screens are more powerful if ranked viabilities across genes are used rather than raw viabilities. In addition, we included the  $t$ -test in our simulation study as we found it to be statistically more powerful than the Wilcoxon rank sum test. The simulation study focused on liver, ovarian, and bone cancers. For each cancer type, the ground truth comprised 30 random pairwise interactions between driver genes and perturbed genes. For the simulation study, we reused the binarized mutation matrices of the corresponding cancer types and simulated the viabilities by sampling from normal distributions. In particular, for the ground truth SL pairs, the viability distribution parameters were different between mutated and WT cell lines (see Methods). Subsequently, we compared SLIdR to the Wilcoxon rank sum test and  $t$ -test on the simulated data. SLIdR significantly outperformed these tests for predicting SL interactions from simulated data in liver, ovary, and bone cancers (Supplementary Fig. 3). SLIdR was especially advantageous in reducing the number of false positives as illustrated by both the ROC and precision-recall curves, even in rare osteosarcomas with only seven cell lines. The performance difference is more pronounced in the precision-recall curves due to the class imbalance. This imbalance stems from our assumption about the simulated data that the number of true SL pairs is much fewer than non-SL pairs. It is noteworthy that the precision-recall curves are in general more informative for performance comparison in highly imbalanced datasets<sup>46</sup>.

### SLIdR recovers experimentally identified SL interactions

Going beyond simulated data, we also evaluated the overlap of SLIdR pan-cancer predictions with experimentally identified SL interactions from (i) 17 in vitro SL screens reported by Lee et al.<sup>31</sup>, and (ii) 10 combinatorial CRISPR screens<sup>17–26</sup>. For these evaluations, we excluded the experimentally identified SL interactions that were not in the set of possible pairwise interactions in the DRIVE dataset, and focused on 978 and 1301 unique experimentally identified SL interactions in the in vitro and CRISPR screens, respectively (see Methods). The retained 1301 unique experimentally identified SL interactions included pairs from 8 of the 10 CRISPR screens. Only four interactions were shared across these 8 combinatorial CRISPR screens. In contrast, SLIdR recovered a significant fraction of established SL interactions (hypergeometric  $p$ -values  $< 10^{-9}$ ) with sensitivities of 12.3% and 11.45% in the in vitro SL screens and CRISPR screens, respectively. With accuracies of ~93% across both sets, these results comprehensively validated SLIdR predictions (Supplementary Data 4).

Overall, these results highlight the ability of SLIdR to identify well-established and potential targets in both pan-cancer and cancer type-specific settings. Exploring these SL interactions, we observed that the

overlap between pan-cancer and cancer type-specific predictions is limited, re-emphasizing the need to explore the data in both settings, as they provide complementary information. The flexibility of SLIdR to work on both small and large sample sizes makes it an attractive method for this problem.

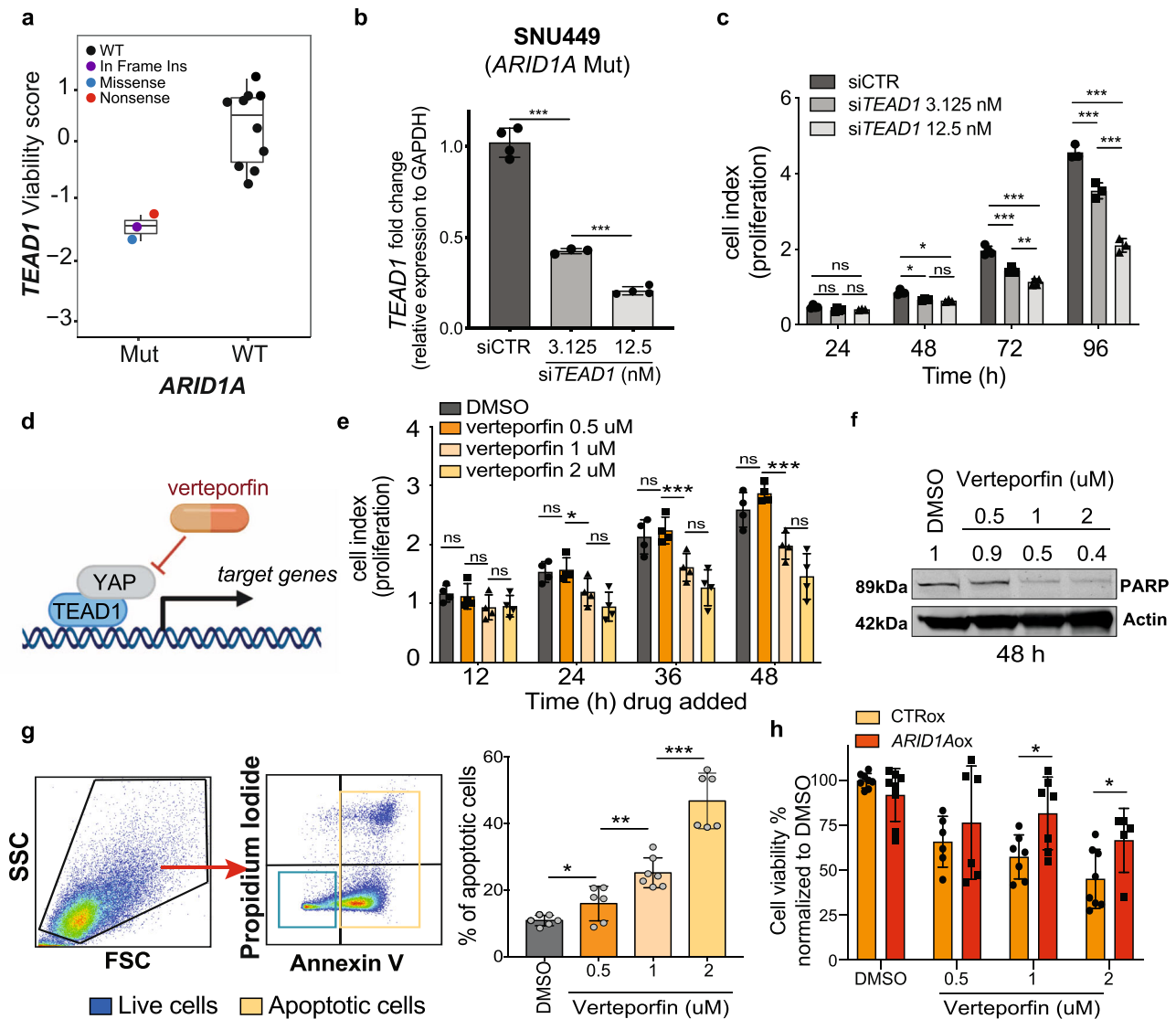
### SLIdR identified two putative targets in hepatocellular carcinoma

In hepatocellular carcinoma (HCC), we identified nine SL pairs (Fig. 3c). Identifying new potential lethal interactions is particularly relevant for the treatment of liver cancer, where few targeted therapies are available<sup>47</sup>. To demonstrate the predictive power of SLIdR, we sought to validate two of our top hits in HCC, namely *ARIDIA-TEAD1* and *AXINI-UR11* (Fig. 4a and Supplementary Fig. 5a). Both *ARIDIA* and *AXINI* loss-of-function mutations are highly prevalent in HCC patients. *ARIDIA* is a SWI/SNF chromatin remodeling gene, and it is mutated in ~9% of hepatocellular carcinomas<sup>48</sup>. *AXINI* encodes for a key Wnt signaling factor, and it is mutated in 5–15% of HCCs<sup>49</sup>. *TEAD1* and *UR11* have been shown to act as oncogenes and potential therapeutic targets in liver cancer<sup>50–52</sup>. First, we validated the SL interaction between *ARIDIA* and *TEAD1* in vitro, using SNU449, a HCC cell line carrying an *ARIDIA* somatic mutation. Silencing of *TEAD1* in SNU449 cells significantly impaired cell proliferation compared to control cells (Fig. 4b, c). Using different concentrations of siRNAs, we observed that the phenotype induced by the knock-down of *TEAD1* was dose and time-dependent, thus indicating a specific on-target effect (Fig. 4b, c).

In addition, we inhibited TEAD1 function using an orthogonal method, specifically the small molecule inhibitor verteporfin. The transcriptional activity of TEAD transcription factors relies on their binding with the Yes-associated protein (YAP) coactivator, a well-known effector of the Hippo-signaling pathway<sup>53</sup>. Verteporfin disrupts the YAP-TAZ complex, therefore inhibiting the transcription of the downstream targets<sup>54</sup> (Fig. 4d). Treatment with different dosage of verteporfin significantly impaired cell proliferation and induced cell death in *ARIDIA* mutant SNU449 cells (Fig. 4e–g), indicating that the SL interaction between *ARIDIA* and *TEAD1* requires *TEAD1* transcriptional activity and is dependent on the Hippo-signaling pathway. Furthermore, to demonstrate that indeed SNU449 sensitivity to verteporfin was dependent on presence of mutant *ARIDIA*, we rescued *ARIDIA* wild-type expression in SNU449 cells. Indeed, we observed that rescuing wild-type *ARIDIA* desensitized cells to verteporfin (Fig. 4h and Supplementary Fig. 6).

Cancer cell lines carry multiple genetic alterations. Therefore, to prove that inhibition of *TEAD1* is specifically synthetic lethal to *ARIDIA* loss-of-function, we employed two HCC-derived cell lines carrying a wild-type *ARIDIA*—Huh-7 and HLE, and modulated *ARIDIA* and *TEAD1* expression using siRNAs (Supplementary Fig. 4a, b). Both Huh-7 and HLE cells transfected with siRNAs targeting both *ARIDIA* and *TEAD1* proliferated significantly less compared to control cells or cells silenced for each gene individually (Fig. 5a and Supplementary Fig. 4c). Additionally, inhibition of TEAD1 using the small molecule verteporfin reduced proliferation in both cell lines only upon *ARIDIA* silencing, while it did not impact the viability of control cells, thus indicating that loss of *ARIDIA* sensitized cells to the treatment (Fig. 5b and Supplementary Fig. 4d). Indeed, *ARIDIA*-silenced Huh-7 and HLE cells both showed a lower IC50 for verteporfin compared to control cells (Fig. 5c and Supplementary Fig. 4e). *ARIDIA*-silenced cells treated with verteporfin additionally showed a significantly higher proportion of apoptotic cells compared to untreated cells or treated control cells in both Huh7 and HLE cell lines (Fig. 5d, e and Supplementary Fig. 4f).

To assess whether *ARIDIA* expression levels would also modulate response to verteporfin in vivo, we employed an in vivo model of the chicken chorioallantoic membrane (CAM), a densely vascularized extraembryonic tissue<sup>55,56</sup>. We treated *ARIDIA*-silenced and control Huh-7 cells with verteporfin (1  $\mu$ M) or vehicle (DMSO) for 24 h. We then



**Fig. 4 | Inhibition of *TEAD1* is deleterious in *ARID1A* mutant liver cancer cells.** **a** Viability scores of *ARID1A* mutant vs wild-type (WT) HCC cell lines with *TEAD1* knockdown from Project DRIVE dataset, where  $n = 13$  HCC cell lines subject to *TEAD1* knockdown experiment. Data are presented as boxplots: Mutant = {min = -1.666, lower (1st Qu.) = -1.5375, middle (median) = -1.409, upper (3rd Qu.) = -1.3190, max = -1.229} and WT = {min = -0.722, lower (1st Qu.) = -0.3455, middle (median) = 0.442, upper (3rd Qu.) = 0.7295, max = 1.044}. **b** RNA expression level (fold-change) of *TEAD1* relative to *GAPDH* in SNU449 cells transfected with control siRNA or with different concentrations (3.125 and 12.5 nM) of *TEAD1* siRNA. RNA levels were assessed by quantitative real-time PCR (qPCR). **c** Proliferation kinetic of SNU449 cells (*ARID1A* mutated) transfected with control siRNA or with different

concentrations (3.125 and 12.5 nM) of *TEAD1* siRNA. **d** Schematic representation of verteporfin mechanism of action. **e, f, g** Proliferation kinetic (**e**), immunoblot for PARP (**f**), and apoptosis assay using AnnexinV and propidium iodide (PI) staining (**g**) of SNU449 cells treated with vehicle (DMSO) or different dosage of verteporfin. **f** Protein lysates of SNU449 cells 48 h post-treatment with DMSO or different dosage of verteporfin up to 48 h. **h** SNU449 cells overexpressing *ARID1A* and control cells treated with vehicle (DMSO) or different dosage of verteporfin. Error bars represent mean (+/-SD) from  $n \geq 2$  replicated. For all experiments performed, statistical significance was assessed by two-sided multiple *t*-tests ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ). **d** was generated using BioRender.

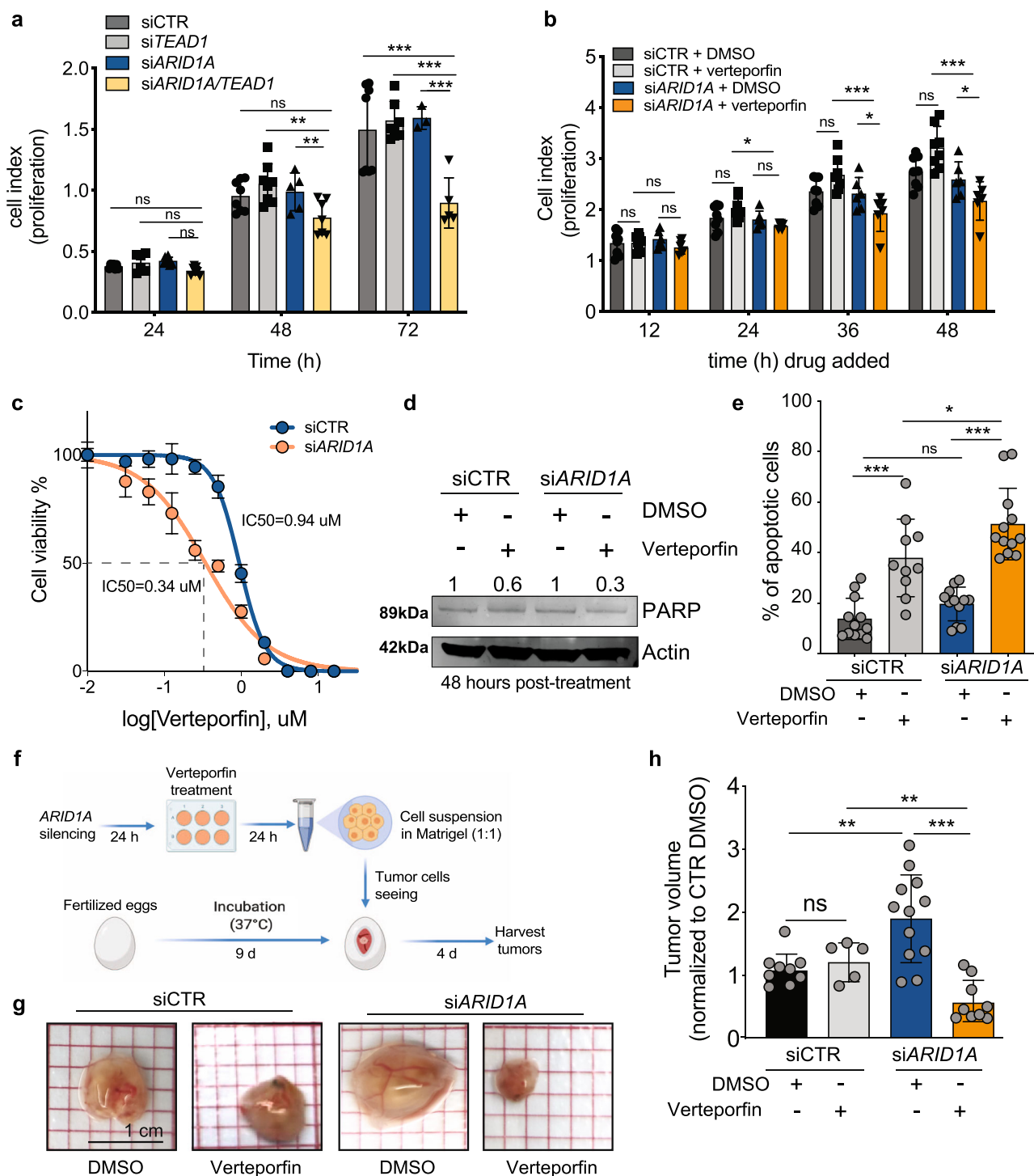
inoculated the cells into the CAMs and screened the eggs for tumor formation 4 days later (Fig. 5f). In accordance with our in vitro results, verteporfin treatment reduced the volume of tumors formed by *ARID1A*-silenced cells, but not in control cells (Fig. 5g, h), suggesting that *ARID1A* expression modulates response to verteporfin in the CAM model as well.

To further corroborate the predictive power of our method, we additionally validated the SL interaction between *AXINI* and *URII*. Similar to the previous SL pair, knockdown of *URII* affected the viability of *AXINI*-mutant and *AXINI*-silenced HCC cells (Supplementary Fig. 5). On the contrary, dual silencing of *AXINI* and *TP53*, predicted as a non-significant SL pair by SLIdR, did not result in decreased proliferation in HCC cells (Supplementary Fig. 5), thus confirming the

specificity of the synthetic lethal interactions predicted by SLIdR. Taken together, our experimental data strongly support the two HCC SL pairs predicted by SLIdR.

### Integration with CRISPR data

We applied SLIdR to CRISPR data from the DepMap consortium (Project Achilles)<sup>57</sup> to identify SL pairs. In both pan-cancer and cancer type-specific settings, we reused the mutation data from the Project DRIVE analyses, replacing the viability scores from DRIVE with the CERES<sup>58</sup> viability scores from the CRISPR screens (see Methods). Subsequently, we ran SLIdR in both settings. In the pan-cancer analysis of the CRISPR data, we found 104 SL pairs including *KEAP1-NFE2L2*, *RPL22-RPL22L1*, *RPL22-WRN*, *APC-CTNBN1*, and *MTAP-PRMT5* from the DRIVE



**Fig. 5** | *ARID1A* down-regulation sensitizes liver cancer cells to *TEAD1* inhibition.

**a** Proliferation kinetic of Huh-7 cells (*ARID1A* wild-type) transfected with control siRNA, *ARID1A* siRNA, *TEAD1* siRNA or both. **b** Proliferation kinetic of Huh-7 cells transfected with control or *ARID1A* siRNA and treated with DMSO or verteporfin (1uM). **c** Dose-response curve of verteporfin in Huh-7 cells transfected with control or *ARID1A* siRNA. **d, e** Immunoblot of PARP (**d**) and apoptosis assay using AnnexinV and propidium iodide (PI) staining (**e**) of Huh-7 cells transfected with control or *ARID1A* siRNA 48 h after treatment with DMSO or verteporfin (1uM). Immunoblot

quantification is relative to loading control (actin). **f** Schematic representation of the CAM assay. **g, h** Representative pictures (**g**) and relative volume quantification (**h**) of tumors explanted from the CAM and derived from Huh-7 cells transfected with control siRNA or *ARID1A* siRNA and treated with DMSO or verteporfin (1uM). Error bars represent mean (+/-SD) from  $n \geq 2$  replicated. For all experiments performed, statistical significance was assessed by two-sided multiple *t*-tests (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). **f** was generated using BioRender.

pan-cancer analysis. While the overlap between the two screens was significant (hypergeometric  $p$ -value  $< 10^{-16}$ ), we observed that several established and strong candidate pairs of the DRIVE analysis, such as *MTAP-MAT2A*, *MTAP-RIOK1*, and *TP53-TP53BP1*, were missed in the

CRISPR analysis by small margins. This could partly be attributed to the reduced power in the CRISPR analysis, as the data had only 266 cell lines in common with the DRIVE data, i.e., 107 cell lines from the DRIVE screen were missing in the CRISPR screen. Another factor is the



differences in pre-processing and normalization steps across the two screens. Therefore, using Fisher's method to combine the corresponding  $p$ -values from both screens for each SL pair and retaining only the significant pairs, we identified 162 robust SL pairs across both screens (see Methods). These 162 pairs retained 62% (91 SL pairs) of the original DRIVE candidate SL pairs (Supplementary Data 5), including strong pairs, such as *ACVR2A-WRN*, *RBI-E2F3*, and *RBI-SKP2*, backed by literature evidence (Supplementary Data 2).

Extending the analysis to the cancer type-specific setting, we found several robust hits in 15 different cancer types (Supplementary Data 5). Some of these were supported by evidence in the literature, including *BRAF-DUSP4*, *BRAF-MAPK1*, and *BRAF-PEA15* in skin, *GATA3-ESR1* in breast, and *CDHI-CTNNB1* in stomach cancers (Supplementary Data 2). In addition to being common between the two screens, *CD93-ALDH18A1* in pancreatic, *KEAP1-ATPIA1* in lung, *BRAF-MAPK1*, *BRAF-MDM2*, and *BRAF-CDK5* in skin, *RHOA-CTNNB1* in stomach, and *OR6CL-MDM2* in large-intestine cancers also found support in the PRISM dataset (significance level  $\alpha = 0.1$ ) before multiple testing correction (Supplementary Data 3).

Similar to the pan-cancer setting, using Fisher's method, we were able to leverage the complementary information across the two screens in the cancer type-specific setting. The method proved especially beneficial in retaining pairs which showed reduced signal or were dropped due to the  $p$ -value threshold for WT cell lines in one of the two screens. Some examples of such robust hits include *KEAP1-TAPT1* in lung cancer - a dependency recently established by Romero et al.<sup>59</sup>, *EIF3C-MTOR* in pancreatic cancer (Supplementary Data 2, 3), *ARIDIA-TEAD1*, *ARIDIA-FERMT2*, and *AXINI-UR11* in hepatocellular carcinoma (Figs. 3c, 4, 5; Supplementary Figs. 4, 5), and *GATA3-MDM2* in breast cancer. In a recent study by Bianco et al.<sup>60</sup>, *GATA3* and *MDM2* were found to be synthetic lethal in estrogen receptor-positive breast cancers. Thus, through parallel analysis and integration of the two datasets, SLIdR reports robust and complementary SL pairs.

## Discussion

With precision medicine advancing in oncology, there is a need to identify mutation-specific therapeutic options for different types of cancer. Most current methods for predicting novel drug targets perform pan-cancer multi-omics analysis and use genetic perturbation screening data, if at all, merely to refine the candidate list of SL pairs. To make better use of these data and generate improved predictions, we developed SLIdR, a rank-based statistical method for predicting mutation-specific SL partners from large-scale perturbation screens.

Following the work of McDonald et al.<sup>27</sup>, we initially applied SLIdR to the Project DRIVE dataset in the pan-cancer setting and recovered most of the SL interactions reported by them in addition to several candidate pairs. However, further analysis of these pan-cancer hits revealed that the signal for many pairs were highly specific to a particular cancer type, warranting the need to identify cancer type-specific SL pairs. Thus, we applied SLIdR to each cancer type individually and identified several well-known as well as candidate pairs. Through a detailed literature survey, systematic validation on large-scale drug-response profiles, and integration with CRISPR screening data, we found strong supporting evidence for several pairs in both the pan-cancer and cancer type-specific settings. These lists of candidate targets, while still requiring extensive validation in follow-up experiments, hold promise for drug repurposing and designing targeted therapies.

A key advantage of SLIdR is its ability to predict SL pairs even with small sample sizes while reducing the number of false positive targets. SLIdR improves the statistical power by using statistical tests based on the Irwin-Hall distribution on the ranked viabilities across all genes. In a simulation study, we have illustrated this advantage and superiority of SLIdR by comparing it to the Wilcoxon rank sum test adopted by ISLE and DAISY. In addition, SLIdR recovered experimentally identified SL

interactions from in vitro SL screens and CRISPR screens with ~93% specificity. Thus, we propose that by replacing the Wilcoxon rank sum test step with SLIdR, methods such as ISLE and DAISY could reduce the number of false positive candidate SL pairs in their initial sets and improve their overall performance.

To corroborate the predictive power of our method, we experimentally validated two synthetic lethal interactions—*ARIDIA-TEAD1* and *AXINI-UR11*, in hepatocellular carcinoma. Specifically, we proved that inhibition of *TEAD1* function, using siRNA or the TEAD-YAP inhibitor verteporfin, is lethal in *ARIDIA*-mutated or *ARIDIA*-silenced HCC cell lines. We additionally showed that treatment with verteporfin significantly reduced the growth of *ARIDIA*-silenced HCC cells in vivo. Taken together, our results show that *ARIDIA* and *TEAD1* are synthetic lethal interactors in HCC and strongly indicate that this relationship is dependent on the regulatory function of *TEAD1* in the Hippo-signaling pathway. Indeed, our results are in accordance with the recent work of Chang et al.<sup>61</sup>. Given the widespread role of the SWI/SNF complex<sup>62</sup>, the frequency of *ARIDIA* inactivation in several malignancies other than HCC<sup>62</sup>, and the availability of TEAD-YAP inhibitors<sup>53</sup>, the identification of the *ARIDIA-TEAD1* synthetic lethal pair provides an example of how SLIdR can help improve cancer therapy. We also validated another independent synthetic lethal pair, *AXINI-UR11*, in HCC, further corroborating the predictive power of our method.

While SLIdR was successful in identifying SL pairs, there are several limitations: (1) In its current scope, SLIdR primarily focuses on identifying SL partners from perturbation screening data, copy number, and mutation data. Consequently, it failed to recover the predictions based on expression and pathway reported by McDonald et al.<sup>27</sup>, emphasizing the importance of integrating other omics data. Thus, extending SLIdR to incorporate multi-omics and pathway data or using it in tandem with methods like ISLE could further improve its overall performance. (2) Occasionally, the results from SLIdR are sensitive to the WT  $p$ -value threshold, resulting in losing important SL pairs. While this threshold is important for filtering out candidates showing increased or decreased viabilities in WT cell lines, the choice of the threshold depends on the data type and processing steps, requiring the user to optimize this threshold to their specific setting. For improved versatility, an extension of SLIdR could include a score to prioritize the pairs instead of filtering based on the WT  $p$ -value threshold. (3) As cancer cells can harbor multiple mutations, SLIdR initially identified the same SL partner for different mutations mapping to the same cells, especially in the pan-cancer setting. To overcome this issue and identify the true SL pairs from several potentially confounding mutated genes, we used propensity score matching-based potential outcome models and identified significant SL pairs. Propensity-score matching tries to replicate a randomized experiment by matching samples between mutated and wild-type groups on other confounding mutations, thereby obtaining balanced covariate distributions across the two groups. Consequently, the average causal effect using matching is determined only from a subset of data. Since the mutation matrix is sparse, matching yields under-powered subsets in certain cases, inadequate for robust estimation of the average causal effect. Furthermore, our current approach for controlling confounding drivers is prone to selection bias, because we do not account for the underlying dependency structure between confounding drivers due to small sample sizes. Further research is required to improve the estimation of the causal effect of driver mutations from sparse mutation data. A possible solution would be to use information from other omics data to select the confounders to adjust for, thus improving the overall estimate.

Taken together, SLIdR provides a robust statistical framework for rapid discovery of SL interactions from large-scale perturbation screens in both pan-cancer and cancer type-specific settings. Particularly in precision oncology, SLIdR can help in developing putative mutation-specific and effective personalized therapies.

## Methods

### Primary screening data

We used viability profiles published in Project DRIVE<sup>27</sup>, as well as corresponding mutation data and copy number data from the Cancer Cell Line Encyclopedia (CCLE) collection<sup>28</sup> for 373 cell lines across 23 cancer types.

### Viability data from Project DRIVE

Viability data captures the viability of cell lines for each gene knockdown experiment. In Project DRIVE, 7837 genes were targeted by using an average of 20 pooled shRNAs per gene. The shRNA activities were defined as the quantile normalized log fold change in shRNA read counts 14 days after the start of the knockdown experiment to the shRNA abundance in the input library. The gene-level viability score of each cell line was computed by aggregation of shRNA activities using two computational methods, namely RSA<sup>63</sup> and ATARIS<sup>64</sup>. The RSA method uses all shRNA reagents targeting a gene and can be used to identify essential, inert, and active genes, while ATARIS only uses a subset of shRNAs with consistent activity across the cell lines and aims to provide a robust gene-level score by discarding shRNA reagents with off-target effects. ATARIS provides a relative score for the gene-level activity by median-centering the data for each reagent and, as a result, cannot distinguish between inert and essential genes.

To process the viability data, we removed essential genes using the RSA method as was performed in Project DRIVE<sup>27</sup>. Genes with an RSA value  $\leq -3$  in more than 50% of cancer cell lines were reported as essential genes. In total, 460 and 185 genes were reported essential in cancer type-specific and pan-cancer settings, respectively. The resulting viability matrices consisted of the ATARIS scores for the remaining perturbed genes (rows) for each cell line (columns).

### Mutation and copy number data

For the pan-cancer setting, we focused on genes with mutations or copy number aberrations in more than 30 cell lines. We downloaded mutation data and copy number data from the CCLE website, and binarized them as follows. A gene in a given cell line was assigned a value of 1 if it was subject to non-synonymous mutations, and 0 otherwise. For copy number data, we focused only on deep deletions and binarized a gene in a given cell line by assigning a value of 1 if the gene was homozygously deleted and a value of 0 otherwise. Finally, combining both these data, a driver gene in a given cell line was assigned a value of 1 if it was subject to non-synonymous mutations, deep deletions, or both; and 0 otherwise.

In the cancer type-specific setting, to define the set of driver genes, we first used the MutSig 2CV v3.1<sup>32,33</sup> MAF file from TCGA for each cancer type and focused only on significantly mutated genes ( $q$ -value  $\leq 0.05$ ). Next, we concentrated on genes with non-synonymous mutations in two or more cell lines, and excluded copy number data as it was very noisy in this setting. Thus, a gene in a given cell line was assigned a value of 1 if it was subject to non-synonymous mutations, and 0 otherwise. The resulting binarized mutation matrices described the mutation profiles for each cell line (column) across all driver genes (rows).

### SLIdR algorithm

SLIdR is a rank-based statistical framework to identify the presence of synthetic lethal dependency between a driver gene  $d$  and a perturbed gene  $g$ . For each driver gene  $d$ , we divided the cell lines into two groups according to the mutation status of  $d$ , namely wild-type cell lines (WT) and mutated cell lines (Mut). Further, we ranked the perturbed genes by their ATARIS scores, for each mutated and WT cell line, and normalized it between 0 and 1. Due to a large number of perturbed genes, the normalized ranks have many distinct levels and are highly fine-grained. Hence, we assumed the normalized ranks to be continuous.

Let  $(d, g)$  be a fixed pair of driver and perturbed gene and  $C_d$  be the set of cell lines mutated in  $d$  of cardinality  $n$ . If  $(d, g)$  is an SL pair, based on the aforementioned definition (Fig. 1a), a mutation in driver gene  $d$  in combination with knockdown of gene  $g$ , results in low viabilities in mutated cell lines  $C_d$ . We used a one-sided statistical test based on the Irwin-Hall distribution to test whether the viabilities of mutated cell lines  $C_d$  from knockdown of gene  $g$  are lower than expected by chance. We defined the null hypothesis  $H_0$  as the knockdown of gene  $g$  having no impact on the viability of the cell lines in  $C_d$ . For each cell line  $c \in C_d$ , we computed the normalized rank of the viability of  $c$  from knockdown of gene  $g$  across all other gene knockdowns in cell line  $c$ , and denoted this rank as  $r_{c|g}$ . Under the null hypothesis, the normalized ranks take uniform random values in the interval  $[0, 1]$ ,  $r_{c|g} \sim U(0, 1)$ . The test statistic  $T$  for the pair  $(d, g)$  is then defined as the sum of normalized viability ranks of mutated cell lines  $C_d$  perturbed in gene  $g$ ,  $T = \sum_{c \in C_d} r_{c|g}$ . Under  $H_0$ , the test statistic  $T$  is the sum of  $n$  independent uniform random variables on the unit interval and hence it follows the Irwin-Hall distribution of order  $n$ . The resulting  $p$ -value was computed as the lower tail probability  $P(T < t_{obs})$ , where  $t_{obs}$  is the observed test statistic. For large  $n$ , computation of the Irwin-Hall probability distribution is either computationally expensive or numerically unstable. Therefore, we used the approximation  $T \sim N(n/2, n/12)$  for  $n > 20$ .

Conversely, based on the definition of synthetic lethality (Fig. 1a), wild-type cell lines with respect to driver gene  $d$  are expected to behave similar to healthy cells when perturbed in gene  $g$ . Therefore, it is important to filter out genes which upon knockdown adversely alter the viabilities of WT cell lines. We used a two-sided Irwin-Hall test to filter out any pair  $(d, g)$  that reached statistical significance ( $\alpha = 0.1$ ) in the WT cell lines. However, we did not use this filter for the pan-cancer setting due to the diverse nature of the cell types and cancer types. Since knockdown data of several driver genes is unavailable in the Project DRIVE screen, we tested for SL interactions only in one direction and were unable to test the effect on viability of the cell lines when the driver genes alone are knocked down.

### Multiple testing

We reduced the number of false positives arising from multiple testing by choosing a significance level of  $1/(M \times N)$ , where  $M$  is the number of knockdowns and  $N$  is the number of driver genes. Therefore, we expect on average one false positive hit among all reported SL hits, for each cancer. Our method is computationally inexpensive as it avoids performing all  $M \times N$  tests. For each driver, we compute the test statistic for all perturbed genes and sort them in ascending order. The pre-ordering of the test statistics enables us to test for genes until the corresponding  $p$ -value is less than the chosen significance level. Further, we note that this approach was in good agreement with controlling the false discovery rate at 10% (Supplementary Fig. 1c).

### Causal inference

Cell lines are often subject to mutations or aberrations in multiple driver genes and, as a result, different driver genes pair with the same SL partner gene (Fig. 2b; Supplementary Fig. 1a). This is typically not an issue in the cancer type-specific setting but is prevalent in the pan-cancer setting. In order to identify the most likely SL pairs from the many confounding driver genes, we used matching-based potential outcome models. The main goal of the matching method is to emulate a randomized experiment by matching samples of treated and control groups according to covariates, thereby obtaining similar covariate distributions across the two groups.

Let  $S = \{d_1, \dots, d_k\}$  be a set of  $k$  driver genes pairing with the same SL partner gene  $g$ . For a given driver gene  $d_i \in S$ , the cell lines mutated and wild-type in  $d_i$  constitute the treated and control groups, respectively, and the viability from knockdown of the SL partner gene  $g$  is used as the response or outcome variable.  $S_{-d_i} = S \setminus \{d_i\}$  is the set of all the driver genes in  $S$  excluding  $d_i$  and constitutes the set of

confounding covariates. We used the *Matching* R package<sup>65</sup> and performed propensity-score matching with a caliper of 0.1 to match the treated and control groups based on confounding covariates, i.e., the mutation status of all the driver genes in  $S_{-d}$ . Since matching is dependent on the order of the samples, we reshuffled and repeated matching 50 times. After each run, we recorded the standardized mean difference (smd) between the two groups for all covariates and chose the run with the lowest sum of smd across all covariates. Finally, for the chosen run, we performed a paired *t*-test between the responses of treated and control groups. We repeated these steps for all  $k$  driver genes in  $S$ .

We repeated the above described process for all such sets of driver genes pairing with the same SL partner gene and reported the pairs that reached statistical significance ( $\alpha = 0.05$ ) in the paired *t*-test (Fig. 2c; Supplementary Fig. S1b).

### Simulation Study

We performed a simulation study focusing on both prevalent and rare cancer types, including liver, ovarian, and bone cancers with 13, 14, and 7 cell lines, respectively. For each cancer type, we reused the binarized mutation matrix from the cancer type-specific analysis on the Project DRIVE dataset and simulated the corresponding viabilities. For a given cancer type, the ground truth comprised 30 random pairwise interactions between driver genes and perturbed genes. The viability matrix for each cancer type was simulated in two steps. First, the viabilities of all the cell lines across all perturbations were simulated by sampling from  $N(\mu_s, \sigma_s^2)$ , where  $\mu_s$  and  $\sigma_s$  are the mean and standard deviation of viabilities from the Project DRIVE data for the primary site  $s$ , respectively. Next, for each true SL pair ( $d, g$ ), we stratified the cell lines into WT and mutated based on the mutation status of the driver gene  $d$ . Since true SL pairs should exhibit differential viability profiles between mutated and WT cell lines, the viabilities of mutated cell lines from perturbation of the true SL partner gene  $g$  were sampled from another distribution  $N(\mu_s^{mut}, \sigma_s^{mut2})$ , where  $\mu_s^{mut} = (\min_s + \mu_s)/2$  and  $\sigma_s^{mut} = 1.2 * \sigma_s$ , and  $\min_s$  is the minimum of viabilities from the Project DRIVE data for the primary site  $s$ . Using these parameters ensured that (i) the perturbation of gene  $g$  resulted in reduced viabilities in mutated cell lines than in WT cell lines, and (ii) the simulated viabilities were congruent with the Project DRIVE data. This was repeated for all 30 true SL pairs in each cancer type.

Finally, we applied SLIdR to the simulated data and compared it to Wilcoxon rank sum test and *t*-test applied to the raw simulated viabilities. For each possible pairwise interaction in a given cancer type, we stratified the cell lines into mutated and WT cell lines based on the mutation status of the driver gene constituting the pair. Then, we tested whether the raw viabilities of the mutated cell lines from the perturbation of the partner gene was less than zero using a one-sided Wilcoxon rank sum test and *t*-test. The resulting *p*-values from SLIdR, Wilcoxon rank sum test, and *t*-test were compared to the true labels in order to plot the ROC and Precision-Recall curves. The true labels comprised a binary vector of all possible pairwise interactions with 1 assigned to true SL pairs and 0 otherwise. This was repeated for all three cancer types.

### Comparison of experimentally identified SL interactions

To compare SLIdR's predictions with established SL interactions, we focused on (i) 6,033 experimentally identified SL interactions from 17 in vitro focused SL screens reported by Lee et al.<sup>31</sup>, and (ii) 24,651 experimentally identified SL interactions from 10 combinatorial CRISPR screens<sup>17–26</sup> (details in Supplementary Note 1). Since we applied SLIdR to the Project DRIVE dataset, we excluded the experimentally identified SL interactions that were not in the set of possible pairwise interactions in the DRIVE dataset (9,443,304), yielding 978 and 1301 unique experimentally identified SL interactions in the in vitro screens and CRISPR screens, respectively. The retained 1301 unique

experimentally identified SL interactions included pairs from 8 of the 10 CRISPR screens and four shared SL interactions (Supplementary Data 4). Finally, we re-ran SLIdR in the pan-cancer setting by relaxing the significance level to 5% and used a hypergeometric test to assess whether the overlap between predicted and experimentally identified SL pairs was significant. We also computed the sensitivity, specificity, and accuracy in both sets.

### Validation on PRISM dataset

We validated the predicted pan-cancer and cancer type-specific SL pairs on the primary PRISM drug-response screen containing the growth inhibitory activity of 4686 drugs tested across 578 human cancer cell lines<sup>37</sup>. For both settings, we first filtered out all the pairs exhibiting oncogene-addiction and retained only those pairs with a targetable SL partner gene from the drugs in the screen. As previously described in the SLIdR algorithm, for each predicted SL pair, we stratified the cell lines based on the mutation status of the driver gene into WT cell lines and mutated cell lines. Subsequently, we tested if the drug compound targeting the corresponding SL partner gene resulted in reduced viabilities of mutated cell lines in comparison to WT cell lines using a one-sided *t*-test. Since multiple drug compounds could target the same SL partner gene, for a given SL pair we tested the differential drug response for each drug compound. We repeated this for all SL pairs and reported the SL pairs with a significant difference in drug response at a significance level of  $\alpha = 0.1$  along with their multiple testing corrected *q*-values.

### Control experiments on PRISM screen

To assess the robustness of the PRISM validation results of SLIdR hits, we compared them to the PRISM results obtained from 1000 sets of random gene pairs. The same process was followed for both pan-cancer and cancer type-specific settings. We first retained only those perturbed genes with at least one matching drug compound in the PRISM screen for each cancer. We then generated a set of all possible pairwise interactions ( $U$ ) between all the driver genes and druggable perturbed genes and filtered out all the pairs exhibiting oncogene-addiction. Subsequently, we ran permutation tests on random sets of gene pairs. For each run, we sampled  $K$  gene pairs from  $U$  at random, where  $K$  is the number of SL pairs predicted by SLIdR and tested on PRISM screens. For each sampled gene pair, we stratified the cell lines based on the mutation status of the driver gene into WT and mutated cell lines. Then, we tested whether the drug compound targeting the corresponding partner gene reduced the viabilities of mutated cell lines compared to WT cell lines using a one-sided *t*-test. Since multiple drug compounds could target the same partner gene, we tested the differential drug response for each drug compound for a given gene pair. We repeated this for all random gene pairs and counted the gene pairs with a significant difference in drug response (significance level of  $\alpha = 0.1$ ). Finally, we computed the empirical *p*-values by comparing the number of significant drug responses of SLIdR hits and those of random sets.

### SLIdR on CRISPR dataset

With Project DRIVE as our primary screen, we extended SLIdR to CRISPR data from the DepMap consortium (Project Achilles 20Q2)<sup>57</sup> to identify robust hits. This secondary screen consisted of CERES<sup>58</sup> viability scores of 769 cell lines across 18,119 CRISPR knockout experiments. However, only 266 cell lines were common with the primary screen and were subsequently used for the downstream analysis.

In both pan-cancer and cancer type-specific settings, we reused the mutation data from the Project DRIVE analyses, replacing only the viability scores from DRIVE with the CRISPR viability scores. However, it should be noted that the reduced overlap of cell lines in the CRISPR screen resulted in dropping a few driver genes in a few cancer types. Specifically, a few driver genes were dropped as they lacked non-

synonymous mutations in at least two cell lines in the cancer type-specific setting. For the viability matrices, we focused only on the common perturbations between both the screens across common cell lines. Subsequently, the CRISPR viability scores were median centered across cell lines.

We re-ran SLiDR with the same threshold parameters as with Project DRIVE. To identify the robust hits, we took the union of the sets of SL pairs predicted by SLiDR on both the screens. For the SL pairs unique to the DRIVE screen, we determined the significance of the pair in the CRISPR screen, and vice-versa. Consequently, we combined the *p*-values corresponding to the mutated cell lines from the two screens using Fisher's method, and filtered those with combined *p*-value < 1/(*M* × *N*), where *M* is the number of knockdowns and *N* is the number of driver genes in the Project DRIVE analysis. Finally, we filtered out all the candidate pairs that reached statistical significance ( $\alpha = 0.1$ ) in the WT cell lines in both screens. However, we did not use this final filter for the pan-cancer setting due to the diverse nature of the cell types and cancer types.

### Cell lines maintenance

Liver cancer-derived cell lines SNU449, Huh-7, and HLE were obtained from the Laboratory of Experimental Carcinogenesis (Bethesda, MD, USA), authenticated by short tandem repeat profiling as previously described<sup>66</sup>, and tested for mycoplasma infection using a PCR-based test (ATCC). All cell lines were maintained under the conditions recommended by the provider. Briefly, all cell lines were cultured in DMEM supplemented with 5% Fetal Bovine Serum (FBS), non-essential amino-acids (NEAA) and antibiotics (Penicillin/Streptomycin). The cells were incubated at 37 °C in a humidified atmosphere containing 5% CO<sub>2</sub>. Exponentially growing cells were used for all in vitro studies.

### Transient gene knockdown by siRNAs

Transient gene knockdown was conducted using ON-TARGET plus siRNA transfection. ON-TARGET plus SMARTpool siRNAs against human *TEAD1*, *ARID1A*, *UR11*, *AXINI* and *TP53*, ON-TARGET plus SMARTpool non-targeting control, and DharmaFECT reagent were all purchased from GE Dharmacon (Supplementary Data 6). Transfection was performed according to the manufacturer's protocol. Briefly, log-phase liver cancer cells were seeded at ~60% confluence. Since residual serum affects the knockdown efficiency of ON-TARGET plus siRNAs, growth medium was removed as much as possible and replaced by serum-free medium (Opti-MEM; Supplementary Data 6). siRNAs were added to a final concentration of 25 nM. siRNAs targeting different genes can be multiplexed. Cells were incubated at 37 °C in 5% CO<sub>2</sub> for 24–48–72 h (for mRNA analysis) or for 48–72 h (for protein analysis). In order to avoid cytotoxicity, the transfection medium was replaced with a complete medium after 24 h.

### Transient *ARID1A* overexpression

For *ARID1A* overexpression, pcDNA6-ARID1A (#39311)<sup>67</sup> was obtained from Addgene (<https://www.addgene.org>), and the empty control vector was pcDNA6-V5/His.b (ThermoFisher Scientific). The expression vectors were transiently transfected using the Lipofectamine 3000 transfection reagent (ThermoFisher Scientific) following the manufacturer's instructions.

### RNA extraction and relative expression by qRT-PCR

Total RNA and proteins were extracted from cells at 75% confluence using TRIZOL (Supplementary Data 6) according to manufacturer's guidelines. cDNA was synthesized from 1 µg of total RNA using SuperScript™ VILO™ cDNA Synthesis Kit (Supplementary Data 6). All reverse transcriptase reactions, including no-template controls, were run on an Applied Biosystem 7900HT thermocycler. Gene expression was assessed by using FastStart Universal SYBR Green Master Mix (Supplementary Data 6) and all qPCR were performed at 50 °C for

2 min, 95 °C for 10 min, and then 40 cycles of 95 °C for 15 s and 60 °C for 1 min on a QuantStudio 3 Real-Time PCR System (Applied Biosystems). The specificity of the reaction was verified by melting curve analysis. Measurements were normalized using GAPDH level as the reference. The fold change in gene expression was calculated using the standard  $\Delta\Delta C_t$  method as previously described<sup>68</sup>. All samples were analyzed in triplicates.

### Proliferation assay

Cell proliferation was assayed using the xCELLigence system (RTCA, ACEA Biosciences, San Diego, CA, USA). Background impedance of the xCELLigence system was measured for 12 s using 50 µl of room temperature cell culture media in each well of E-plate 16. Cells were grown and expanded in tissue culture flasks as previously described. After reaching 75% confluence, the cells were washed with PBS and detached from the flasks using a short treatment with trypsin/EDTA. 5000 cells were dispensed into each well of an E-plate 16. Growth and proliferation of the cells were monitored every 15 min up to 120 h via the incorporated sensor electrode arrays of the xCELLigence system, using the RTCA-integrated software according to the manufacturer's parameters. In the case of transient siRNA transfection, cells were detached and plated on xCELLigence 24 h post-transfection.

### Drug treatment

5 × 10<sup>3</sup> exponentially growing cells were plated in a 96-well plate. After 24 h, cells were treated with serial dilution of verteporfin (Supplementary Data 6) or dimethyl sulfoxide (DMSO). DMSO served as the drug vehicle, and its final concentration was no >0.1%. Cell viability was measured after 72 h using CellTiter-Glo Luminescent Cell Viability Assay reagent (Supplementary Data 6). Luminescence was measured on Varioskan Microplate Reader (ThermoFisher Scientific). Results were normalized to DMSO control. All experiments were performed in triplicates. Results are shown as mean ± SD. Curve fitting was performed using Prism (GraphPad) software and the nonlinear regression equation.

### Apoptosis analysis by flow cytometry

Cells were collected 72 h post siRNA transfection, stained with annexinV (FITC conjugate; Supplementary Data 6) and propidium iodide (PI), and analyzed by flow cytometry using the BD FACSCanto II cytometer (BD Biosciences, USA). Briefly, cells were harvested after the incubation period and washed twice by centrifugation (1200 g, 5 min) in cold phosphate-buffered saline (DPBS; Supplementary Data 6). After washing, cells were resuspended in 0.1 mL AnnV binding buffer 1X (ABB 5X, Supplementary Data 6) containing fluorochrome-conjugated AnnV and PI (PI to a final concentration of 1 µg/mL) and incubated in darkness at room temperature for 15 min. Following immediately, cells were analyzed by flow cytometry, measuring the fluorescence emission at 530 nm and >575 nm. Data were analyzed by FlowJo software version 10.5.3 (<https://www.flowjo.com/>).

### Immunoblot

Total proteins were extracted by directly lysing the cells in Co-IP lysis buffer (100 mmol/L NaCl, 50 mmol/L Tris pH 7.5, 1 mmol/L EDTA, 0.1% Triton X-100) supplemented with protease and phosphatase inhibitors. Cell lysates were then treated with a reducing agent, boiled and loaded onto neutral pH, pre-cast, discontinuous SDS-PAGE mini-gel system. After electrophoresis, proteins were transferred to nitrocellulose membranes using the Trans-Blot Turbo Transfer System (Bio-Rad). The transblotted membranes were blocked for 1 h in TBST 3% SureBlock (LubioScience) and then probed with appropriate primary antibodies (1:1000) overnight at 4 °C. List of antibodies and working concentrations are available in (Supplementary Data 6). Next, the membranes were incubated for 1 h at room temperature with fluorescent secondary goat anti-mouse (IRDye 680) or anti-rabbit (IRDye

800) antibodies (both from LI-COR Biosciences). Blots were scanned using the Odyssey Infrared Imaging System (LI-COR Biosciences) and band intensity was quantified using ImageJ software. The ratio of proteins of interest/loading control in verteporfin-treated samples were normalized to their DMSO-treated control counterparts. All experiments were performed and analyzed in duplicates.

### Chorioallantoic membrane (CAM)

Fertilized chicken eggs were obtained from Gepro Geflügelzucht AG at day 1 of gestation and were maintained at 37 °C in a humidified (60%) incubator for 9 days<sup>69</sup>. At this time, an artificial air sac was formed using the following procedure: a small hole was drilled through the eggshell into the air sac and a second hole near the allantoic vein that penetrates the eggshell membrane. Mild vacuum was applied to the hole over the air sac in order to drop the CAM. Subsequently, a square 1 cm window encompassing the hole near the allantoic vein was cut to expose the underlying CAM<sup>69</sup>. After the artificial air sac was formed, Huh-7 cells growing in tissue culture were inoculated on CAMs at  $1.5 \times 10^6$  cells per CAM, on three to six CAMs each. Specifically, 48 h post-siRNA transfection, *ARID1A*-silenced and control Huh-7 cells were treated with verteporfin (1  $\mu$ M). 24 h post-treatment, cells were detached from the culture dish with Trypsin, counted, suspended in 20  $\mu$ l of medium (DMEM) and mixed with an equal volume of Matrigel. To prevent leaking and spreading of cells, an 8 mm (inner diameter) sterile teflon ring (removed from 1.8 ml freezing vials, Nunc, Denmark) was placed on the CAMs and the final mixture was grafted onto the chorioallantoic membranes inoculating the cells with a pipette inside the ring<sup>70</sup>. Embryos were maintained at 37 °C for 4 days after which tumors at the site of inoculation were excised using surgical forceps. Images of each tumor were acquired with a Canon EOS 1100D digital camera. Surface measurements were performed by averaging the volume (height\*width\*width) of each tumor using ImageJ, as previously described<sup>71</sup>. Total  $n \geq 5$  tumors for each condition were analyzed over three independent experiments. Experiments in fertilized eggs were performed before the third period of the embryonation period (before day 14). Experiments in fertilized eggs in this period do not require ethical approval.

### Statistics and reproducibility

No statistical method was used to predetermine sample size. For the in vivo experiments, CAMs were allocated randomly to each condition. For the in vivo model, CAMs were screened for tumor formation blindly by two independent scientists.

For the functional experiment reported in Figs. 4, 5 and Supplementary Figs. 4, 5, 6, statistical analyses were conducted using Prism software v8.0 (GraphPad Software, La Jolla, CA, USA). For in vitro experimental studies, statistical significance was determined by the two-tailed unpaired Student's *t*-test. For comparison involving multiple time points, statistical significance was determined by multiple Student's *t*-test corrected for multiple comparisons with the Holm-Sidak method. A *p*-value < 0.05 was considered statistically significant. For all figures, 'ns' indicates not reaching the significance level. For the CAM assay, a two-tailed unpaired Student's *t*-test was used. Unless otherwise indicated, all data represent the mean  $\pm$  standard deviation from at least three independent experiments.

For details on the reagents used, please refer to Supplementary Data 6.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw shRNA data has already been published as a part of Project DRIVE (<https://data.mendeley.com/datasets/y3ds55n88r/4>) and all the

mutation and copy number data from CCLE is available at <https://portals.broadinstitute.org/ccle>. The MutSig 2CV v3.1<sup>32,33</sup> MAF file for each cancer type is available at <http://firebrowse.org/>. The processed project DRIVE data for running SLiDR in pan-cancer and cancer type-specific settings are available at <https://doi.org/10.6084/m9.figshare.21508065.v4><sup>72</sup>. The PRISM drug-response dataset and the CRISPR dataset (Project Achilles 20Q2) are available at <https://depmap.org/portal/download/>. All these datasets are publicly available. The validation experimental data generated in this study are provided in the Source data file. Source data are provided with this paper.

### Code availability

The latest version of SLiDR package is available at <https://github.com/cbg-ethz/slidr><sup>73</sup> along with the scripts used to process and generate the results for the paper. Other R packages used, include, Matching (v4.9-3), tableone (v0.9.3), dplyr (v0.7.6), ggplot2 (v3.2.0.9), cowplot (v0.9.2), circlize (v0.4.5). For the validation experiments the following softwares were used—BD, FACSDiva and FlowJo (10.5.3), Prism for MacOS v8.2.1.

### References

- Miki, Y. et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
- Wooster, R. et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–792 (1995).
- Powell, S. N. & Kachnic, L. A. Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* **22**, 5784–5791 (2003).
- Metzger, M. J., Stoddard, B. L. & Monnat, R. J. PARP-mediated repair, homologous recombination, and back-up non-homologous end joining-like repair of single-strand nicks. *DNA Repair* **12**, 529–534 (2013).
- Ying, S., Hamdy, F. C. & Helleday, T. Mre11-dependent degradation of stalled DNA replication forks is prevented by BRCA2 and PARP1. *Cancer Res.* **72**, 2814–2821 (2012).
- Shaheen, M., Allen, C., Nickoloff, J. A. & Hromas, R. Synthetic lethality: exploiting the addiction of cancer to DNA repair. *Blood* **117**, 6074–6082 (2011).
- Bryant, H. E. et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
- Szczurek, E. & Beerenwinkel, N. Modeling mutual exclusivity of cancer mutations. *PLoS Comput. Biol.* **10**, e1003503 (2014).
- Vandin, F., Upfal, E. & Raphael, B. J. De novo discovery of mutated driver pathways in cancer. *Genome Res.* **22**, 375–385 (2012).
- Leiserson, M. D. M., Blokh, D., Sharan, R. & Raphael, B. J. Simultaneous identification of multiple driver pathways in cancer. *PLoS Comput. Biol.* **9**, e1003054 (2013).
- Constantinescu, S., Szczurek, E., Mohammadi, P., Rahnenführer, J. & Beerenwinkel, N. TiMEx: a waiting time model for mutually exclusive cancer alterations. *Bioinformatics* **32**, 968–975 (2016).
- Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res* **22**, 398–406 (2012).
- Deshpande, R. et al. A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res.* **73**, 6128–6136 (2013).
- Jacunski, A., Dixon, S. J. & Tatonetti, N. P. Connectivity homology enables inter-species network models of synthetic lethality. *PLoS Comput. Biol.* **11**, e1004506 (2015).
- Liu, L. et al. Synthetic lethality-based identification of targets for anticancer drugs in the human signaling network. *Sci. Rep.* **8**, 8440 (2018).
- Astsaturov, I. et al. Synthetic lethal screen of an EGFR-centered network to improve targeted therapies. *Sci. Signal.* **3**, ra67 (2010).

17. Najm, F. J. et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189 (2018).
18. Han, K. et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463–474 (2017).
19. Hortbeck, M. A. et al. Mapping the genetic landscape of human cells. *Cell* **174**, 953–967.e22 (2018).
20. Shen, J. P. et al. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14**, 573–576 (2017).
21. Thompson, N. A. et al. Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat. Commun.* **12**, 1302 (2021).
22. Norman, T. M. et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786–793 (2019).
23. Zhao, D. et al. Combinatorial CRISPR–Cas9 metabolic screens reveal critical redox control points dependent on the KEAP1–NRF2 regulatory axis. *Mol. Cell* **69**, 699–708.e7 (2018).
24. 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.* **21**, 262 (2020).
25. Zhao, Y. et al. A one-step tRNA–CRISPR system for genome-wide genetic interaction mapping in mammalian cells. *Sci. Rep.* **9**, 14499 (2019).
26. DeWeirdt, P. C. et al. Genetic screens in isogenic mammalian cell lines without single cell cloning. *Nat. Commun.* **11**, 752 (2020).
27. McDonald, E. R. 3rd et al. Project DRIVE: A compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell* **170**, 577–592.e10 (2017).
28. Barretina, J. et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
29. Jerby-Arnon, L. et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* **158**, 1199–1209 (2014).
30. Sinha, S. et al. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. *Nat. Commun.* **8**, 15580 (2017).
31. Lee, J. S. et al. Harnessing synthetic lethality to predict the response to cancer treatment. *Nat. Commun.* **9**, 2546 (2018).
32. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
33. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
34. Marjon, K. et al. MTAP deletions in cancer create vulnerability to targeting of the MAT2A/PRMT5/RIOK1 Axis. *Cell Rep.* **15**, 574–587 (2016).
35. Tanizaki, J. et al. Roles of BIM induction and survivin downregulation in lapatinib-induced apoptosis in breast cancer cells with HER2 amplification. *Oncogene* **30**, 4097–4106 (2011).
36. Zhao, P. et al. Regulation of survivin by PI3K/Akt/p70S6K1 pathway. *Biochem. Biophys. Res. Commun.* **395**, 219–224 (2010).
37. Corsello, S. M. et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **1**, 235–248 (2020).
38. Goldstein, L. D. et al. Recurrent loss of NFE2L2 exon 2 is a mechanism for Nrf2 pathway activation in human cancers. *Cell Rep.* **16**, 2605–2617 (2016).
39. Leiserson, M. D. M. et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
40. Theodorou, V., Stark, R., Menon, S. & Carroll, J. S. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* **23**, 12–22 (2013).
41. Kansara, M. & Thomas, D. M. Molecular pathogenesis of osteosarcoma. *DNA Cell Biol.* **26**, 1–18 (2007).
42. Chen, X. et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112 (2014).
43. Pignochino, Y. et al. Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a mechanism potentially involving the inhibition of ERK1/2, MCL-1 and ezrin pathways. *Mol. Cancer* **8**, 118 (2009).
44. Doldan, A. et al. Loss of the eukaryotic initiation factor 3f in pancreatic cancer. *Mol. Carcinog.* **47**, 235–244 (2008).
45. Marchione, R., Leibovitch, S. A. & Lenormand, J.-L. The translational factor eIF3f: the ambivalent eIF3 subunit. *Cell. Mol. Life Sci.* **70**, 3603–3616 (2013).
46. Davis, J. & Goadrich, M. The relationship between Precision-Recall and ROC curves. *Proc. 23rd International Conference on Machine Learning—ICML '06* <https://doi.org/10.1145/1143844.1143874> (2006).
47. Llovet, J. M., Villanueva, A., Lachenmayer, A. & Finn, R. S. Advances in targeted therapies for hepatocellular carcinoma in the genomic era. *Nat. Rev. Clin. Oncol.* **12**, 436 (2015).
48. Peng, Y. et al. Hepatocellular carcinoma with ARID1A mutation is associated with higher TMB and poor survival. *J. Clin. Oncol.* **38**, e16667–e16667 (2020).
49. Abitbol, S. et al. AXIN deficiency in human and mouse hepatocytes induces hepatocellular carcinoma in the absence of  $\beta$ -catenin activation. *J. Hepatol.* **68**, 1203–1213 (2018).
50. Zhang, J. et al. RMP promotes venous metastases of hepatocellular carcinoma through promoting IL-6 transcription. *Oncogene* **34**, 1575–1583 (2015).
51. Tummala, K. S. et al. Inhibition of de novo NAD(+) synthesis by oncogenic URI causes liver tumorigenesis through DNA damage. *Cancer Cell* **26**, 826–839 (2014).
52. Zhou, Z. et al. Targeting Hippo pathway by specific interruption of YAP–TEAD interaction using cyclic YAP-like peptides. *FASEB J.* **29**, 724–732 (2015).
53. Pobbati, A. V. & Hong, W. A combat with the YAP/TAZ–TEAD oncoproteins for cancer therapy. *Theranostics* **10**, 3622–3635 (2020).
54. Giraud, J. et al. Verteporfin targeting YAP1/TAZ–TEAD transcriptional activity inhibits the tumorigenic properties of gastric cancer stem cells. *Int. J. Cancer* **146**, 2255–2267 (2020).
55. Hagedorn, M. et al. Accessing key steps of human tumor progression in vivo by using an avian embryo model. *Proc. Natl Acad. Sci. USA.* **102**, 1643–1648 (2005).
56. Fluegen, G. et al. Phenotypic heterogeneity of disseminated tumour cells is preset by primary tumour hypoxic microenvironments. *Nat. Cell Biol.* **19**, 120–132 (2017).
57. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017).
58. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
59. Romero, R. et al. Keap1 mutation renders lung adenocarcinomas dependent on Slc33a1. *Nat. Cancer* **1**, 589–602 (2020).
60. Bianco, G. et al. GATA3 and MDM2 are synthetic lethal in estrogen receptor-positive breast cancers. *Commun. Biol.* **5**, 1–15 (2022).
61. Chang, L. et al. The SWI/SNF complex is a mechanoregulated inhibitor of YAP and TAZ. *Nature* **563**, 265–269 (2018).
62. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
63. König, R. et al. A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* **4**, 847–849 (2007).

64. Shao, D. D. et al. ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res* **23**, 665–678 (2013).
65. Sekhon, J. S. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Statistical Software* <https://doi.org/10.18637/jss.v042.i07> (2011).
66. Reid, Y., Storts, D., Riss, T. & Minor, L. *Authentication of Human Cell Lines by STR DNA Profiling Analysis* (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2013).
67. Guan, B., Wang, T.-L. & Shih, I.-M. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Res.* **71**, 6718–6727 (2011).
68. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**, 402–408 (2001).
69. Zijlstra, A. et al. A quantitative analysis of rate-limiting steps in the metastatic cascade using human-specific real-time polymerase chain reaction. *Cancer Res.* **62**, 7083–7092 (2002).
70. Kim, J., Yu, W., Kovalski, K. & Ossowski, L. Requirement for specific proteases in cancer cell intravasation as revealed by a novel semi-quantitative PCR-based assay. *Cell* **94**, 353–362 (1998).
71. Lauzier, A. et al. Colorectal cancer cells respond differentially to autophagy inhibition in vivo. *Sci. Rep.* **9**, 11316 (2019).
72. Srivatsa, S. et al. *Processed DRIVE Data for Running SLiDR*. <https://doi.org/10.6084/m9.figshare.21508065.v4> (2022).
73. Srivatsa, S. et al. *Discovery of Synthetic Lethal Interactions From Large-Scale Pan-Cancer Perturbation Screens, cbg-ethz/slidr*. Zenodo <https://doi.org/10.5281/ZENODO.7302071> (2022)
- in vitro and in vivo experiments. N.B., S.P. and C.K.Y.N. critically discussed the results. S.S., H.M., G.B., M.C.-L., C.K.Y.N., S.P., and N.B. interpreted the results and wrote the manuscript. All authors agreed to the final version of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-35378-z>.

**Correspondence** and requests for materials should be addressed to Salvatore Piscuoglio or Niko Beerenwinkel.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

### Acknowledgements

This work was partly supported by ERC Synergy Grant 609883 to N.B., by the Swiss Cancer League [KFS-4543-08-2018 to C.K.Y.N. and KFS-4988-02-2020-R to S.P.], and by The Professor Dr Max Cloëtta Foundation (Medical research position) to S.P.

### Author contributions

S.S. and N.B. conceived the presented idea. S.S. and H.M. developed and implemented the method and built the R package. N.B. supervised the computational study. S.P. and C.K.Y.N. supervised the experimental study. S.S. performed the downstream bioinformatics analyses and contributed to the visualizations. G.B., M.C.-L. and M.M. performed