ASTER: A Method to Predict Clinically Actionable Synthetic Lethal Interactions

Herty Liany[1], Anand Jeyasekharan[2,3], and Vaibhav Rajan[4]

[1] Department of Computer Science, School of Computing, National University of Singapore
e0146315@u.nus.edu
[2] Cancer Science Institute, National University of Singapore
[3] National University Hospital, Singapore
csiadj@nus.edu.sg
[4] Department of Information Systems and Analytics, School of Computing, National University of Singapore
vaibhav.rajan@nus.edu.sg

Abstract. A Synthetic Lethal (SL) interaction is a functional relationship between two genes or functional entities where the loss of either entity is viable but the loss of both is lethal. Such pairs can be used to develop targeted anticancer therapies with fewer side effects and reduced overtreatment. However, finding clinically actionable SL interactions remains challenging. Leveraging large-scale unified gene expression data of both disease-free and cancerous data, we design a new technique, based on statistical hypothesis testing, called ASTER (**A**nalysis of **S**ynthetic lethality by comparison with **T**issue-specific disease-free g**E**nomic and t**R**anscriptomic data) to identify SL pairs. For large-scale multiple hypothesis testing, we develop an extension called ASTER++ that can utilize additional input gene features within the hypothesis testing framework. Our extensive experiments demonstrate the efficacy of ASTER in accurately identifying SL pairs that are therapeutically actionable in stomach and breast cancers.

Keywords: Synthetic Lethality · Hypothesis Test · Gene Expression.

1 Introduction

Cancer is one of the leading causes of mortality and morbidity worldwide [32]. Interestingly, the mortality rate of cancer in the US has declined by 29% from 1991 to 2017, including the highest annual drop ever recorded from 2016 to 2017 [35]. These improvements in outcomes are, at least in part, due to novel immunotherapy and targeted therapies in cancers such as leukemia, melanoma and lymphoma [35]. While major challenges remain in understanding and combating cancer, these developments offer hope and impetus for advancing personalized genomics-driven cancer therapeutics. Such targeted treatments aim to design highly specific therapies with fewer adverse effects and reduced overtreatment [31]. Exploiting Synthetic Lethality is considered a promising approach to identify such targeted therapeutic targets [31, 33].

A Synthetic Lethal (SL) genetic interaction is a functional relationship between two genes or functional entities where the loss of either entity is viable but the loss of both is lethal. The key idea used in targeted cancer therapeutics is that in a malignant cell, functionally disruptive mutation in one of the two genes (say, A) of an SL pair (A,B) leads to dependency on B for survival and cancer cells can be selectively killed by inhibiting B. Non-cancerous cells, that have A, survive even when B is inhibited. See fig. 1 for a schematic. For example, mutations causing functional loss of BRCA1/2 genes leads to deficiency of DNA Damage Response mechanism and dependence on the protein PARP1/2 [9]. Drugs based on PARP inhibitors are found to be effective for breast and ovarian cancers [39, 1].

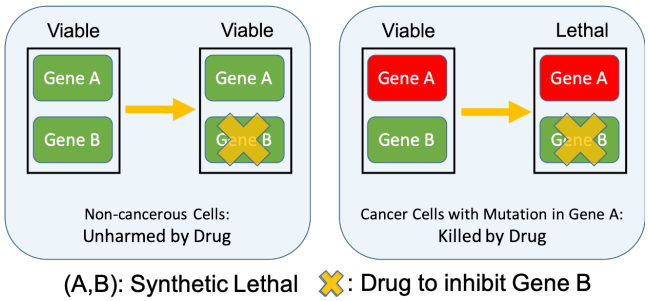

Fig. 1: Synthetic Lethality: gene A or gene B can ensure cell survival, but loss of both is lethal.

Despite the concept being decades old (from genetic studies in fruit fly and yeast, e.g., [29, 4]), SL interactions in humans remain largely unknown. Genome-wide screens have been developed, e.g., RNA interference screens and CRISPR screens, to identify potential SL pairs but they are costly and labour-intensive. Several challenges remain: first, since these genetic interactions are lethal, mutant recovery and identification become difficult; second, many SL pairs are conditionally dependent and may not be conserved in all genetic backgrounds or in different cellular conditions and third, large number of gene pairs need to be queried to identify SL interactions [31].

Computational methods have been developed to identify and prioritize potential SL pairs, as candidates that can be functionally analyzed through genome-wide screens. Broadly they can be divided into machine learning and statistical methods. Machine learning methods, including those based on network analytics, rely on labelled data to predict SL pairs, e.g., [22, 6, 26, 10]. However, these methods face the challenge of scarce labels since very few human SL pairs are experimentally confirmed. As a result, some approaches have developed models that can incorporate information from yeast SL pairs [11, 36, 45]. Many machine learning models use

binary classifiers trained on data where the evidence of negative labels, i.e., those pairs that are *not* SL are not confirmed through screens. In general, such positive-unlabelled learning tasks, with scarce positive labels can be challenging [3].

Statistical approaches, that do not rely on labelled data, such as DAISY [23] and ISLE [25], are popular alternatives based on well-defined biological hypotheses. For an input gene pair, DAISY applies three statistical inference procedures and the pair is considered SL if all three criteria are satisfied. The first test, uses Somatic Copy Number Alteration (SCNA) and gene expression data to detect gene pairs that are infrequently co-inactivated. The second test uses shRNA essentiality screens, SCNA and gene expression profiles, to identify pairs where underexpression and low copy number of a gene induces essentiality of the partner gene. The third test checks for significant co-expression in transcriptomic data, with the assumption that SL pairs, participating in related biological processes, are likely to be coexpressed.

ISLE is designed to obtain clinically relevant SL pairs, from an initial (larger) collection of potential SL pairs. They also apply three statistical procedures, but unlike DAISY, the tests are done in a sequential manner. In the first test, gene expression and SCNA data is used to identify candidate gene pairs with significantly infrequent co-inactivations, representing under-represented negative selection. Second, from the selected gene pairs in the first test, a gene pair is selected if its co-inactivation leads to better predicted patient survival, using the Cox proportional hazards model, compared to when it is not co-inactivated. In the final step, from the selected gene pairs in the second test, only pairs with high phylogenetic similarity are retained, assuming co-evolution of functionally interacting genes.

In this paper, we design a new technique, based on statistical hypothesis testing, called ASTER (**A**nalysis of **S**ynthetic lethality by comparison with **T**issue-specific disease-free g**E**nomic and t**R**anscriptomic data) to identify a potentially SL gene pairs. Unlike previous statistical methods that utilize data from only cancerous tissues (mainly from The Cancer Genome Atlas (TCGA)) for their analysis, ASTER leverages RNA-Seq expression data from disease-free tissues in the Genotype Tissue Expression (GTEx) project [42]. Data in GTex has been unified with cancer tissues from TCGA, after successful correction for study-specific biases [42, 17]. Thus, GTEx provides reference expression levels across various tissues for comparison with the expression levels found in cancer. We find that the use of tissue-specific, disease-free samples in ASTER results in a considerably simple and effective method that uses only SCNA and RNA-Seq data. For large-scale multiple hypothesis testing, we develop an extension called ASTER++ that uses AdaFDR [47] to adaptively find a decision threshold based on additional input gene features. Similar to machine learning based methods, ASTER++ can utilize gene-specific features in its predictions, but without their limitation of requiring labelled data to learn from. Moreover, it retains the interpretability of statistical hypothesis testing, while leveraging on AdaFDR's scalability and flexibility. Our extensive experiments demonstrate the efficacy of ASTER in accurately identifying SL pairs that are therapeutically actionable in stomach and breast cancers.

2 Our Method

Our method, called ASTER (**A**nalysis of **S**ynthetic lethality by comparison with **T**issue-specific disease-free g**E**nomic and t**R**anscriptomic data), consists of sequential application of three tests for a candidate gene pair (A,B).

Let $S_{(A\uparrow)}$ be the set of tissue-specific samples (from TCGA) with high copy number (SCNA > 1) for gene A. Let $S_{(B\downarrow\in A\uparrow)} \subset S_{(A\uparrow)}$ be the set of samples in $S_{(A\uparrow)}$ with low copy number

(SCNA < 1) for gene B. Let N denote non-cancerous samples of the same tissue type (from GTex).

T1: We test if the expression levels of gene A in $S_{(A\uparrow)}$ is significantly higher than the expression levels of A in N.
T2: We then test if the expression level of gene B in $S_{(B\downarrow\in A\uparrow)}$ is significantly lower than the expression levels of B in N.
T3: Finally, we test if the expression levels of gene A in $S_{(A\uparrow)}$ is significantly higher than the expression levels of gene B in $S_{(B\downarrow\in A\uparrow)}$.

We use the non-parametric Wilcoxon Rank Sum Test for each of the three tests. Fisher's method [14] is used to obtain a single p-value by combining the p-values from the three independent tests. Note that due to the sequential manner of testing, the application of ASTER on gene pairs (A,B) and (B,A) may yield different results. ASTER explicitly tests for up-regulation and amplification in the first gene and simultaneous down-regulation and deletion in the second gene; to highlight this we denote the gene pair by $(A\uparrow, B\downarrow)$. Figure 2 shows a schematic.

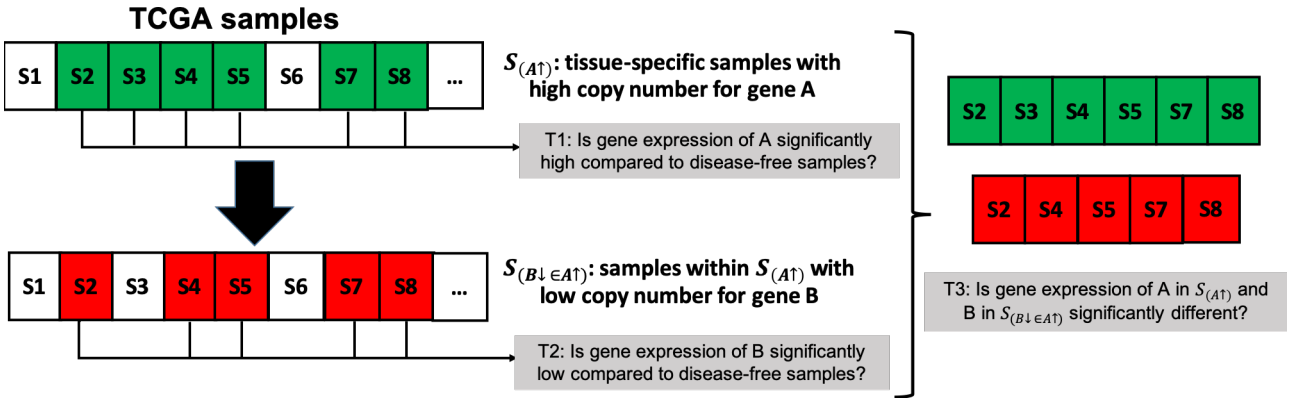

Fig. 2: Overview of ASTER. T1: Green color indicates samples in $S_{(A\uparrow)}$ where gene A is significantly up-regulated (compared to disease-free GTex samples). T2: Red color indicates samples in $S_{(B\downarrow\in A\uparrow)}$ where gene B is significantly down-regulated (compared to disease-free GTex samples). T3: Gene expression values of selected samples are compared to conduct test T3.

2.1 ASTER++

To enable large-scale multiple hypothesis testing and the use of additional known covariates about the gene pairs, we combine ASTER with AdaFDR [47]. AdaFDR adaptively finds a decision threshold from covariates at a user-specified False Discovery Proportion. For a candidate list of gene pairs, their features and corresponding p-values (obtained from ASTER in our case), AdaFDR learns a decision threshold that depends on the covariates. Thus, instead of a fixed threshold used in previous hypothesis testing based methods (e.g., DAISY), through the use of AdaFDR, we can obtain a covariate-dependent threshold that may be different for each gene pair. While the adaptive threshold can be directly used to predict SL for a specific gene pair, it poses a problem in ranking the input gene pairs.

Let p_1^n, p_2^n, p_3^n be the p-values from the three tests of ASTER for the nth gene pair. With a fixed threshold (e.g., $t = 0.01$) we can choose all the pairs with $p_i^n < t$ for $i = 1, 2, 3$ as the predicted SL pairs. They can be ranked using the single p-value obtained after applying Fisher's method. When an adaptive covariate-dependent threshold is learnt from AdaFDR, the nth gene pair has three different thresholds t_i^n, one for each p-value p_i^n for $i = 1, 2, 3$. We can consider $s_i^n = p_i^n - t_i^n$ to be a score which indicates how far each p-value is from its own threshold, with a lower value indicating higher significance. For the nth gene pair and the ith test there is a decision value $d_i^n \in \{0, 1\}$ that is set to 1 if $p_i^n < t_i^n$. Only those gene pairs are selected that pass all three tests, i.e., when $\prod_i d_i^n = 1$. The selected gene pairs can be ranked using the score $\sum_i s_i^n$ with a lower value indicating higher probability of being SL.

We call this combined method of using ASTER, AdaFDR and re-scoring as described above ASTER++. Similar to machine learning based methods, ASTER++ can utilize gene-specific features in its predictions, without their limitation of requiring labelled data to learn from. Moreover, it retains the interpretability of ASTER's statistical hypothesis testing, while leveraging on AdaFDR's scalability and flexibility of large-scale multiple hypothesis testing.

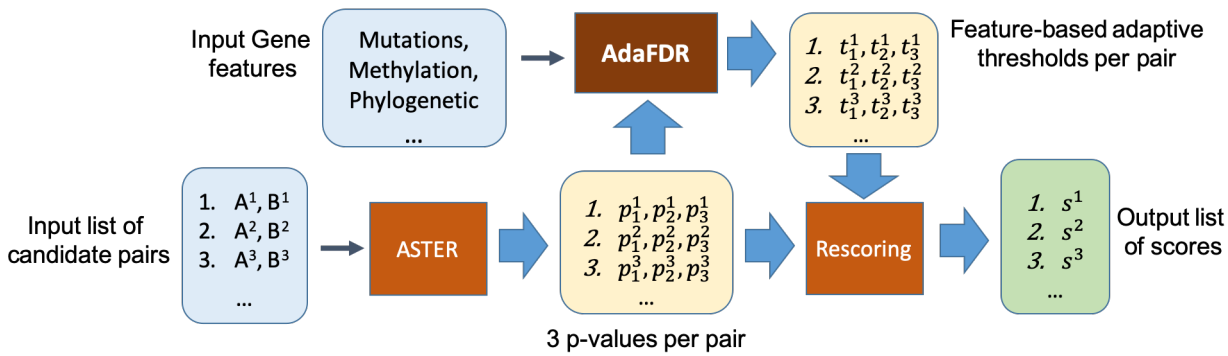

Fig. 3: ASTER++ pipeline for large-scale multiple testing and use of additional features.

3 Experiments

3.1 Validating ASTER: Prognostic Value and Functional Annotations

We validate the approach adopted for ASTER in two ways: (1) we measure the prognostic value of the mutual exclusivity pattern that was used to predict SL by comparing the survival rate of patients who exhibit the pattern with those who do not; and (2) we compare the functional annotations of the pairs of genes that are most likely to be SL with those that are least likely to be SL based on predictions from ASTER.

Prognostic Value of Predicted SL Pairs We use 16,916 gene pairs listed in SynLethDB [19] as input to ASTER. SL pairs in Breast and Stomach cancer are identified using a p-value threshold of 0.01 for each test in ASTER. For a predicted gene pair $(A\uparrow, B\downarrow)$, we consider samples in the set $I = S_{(B\downarrow \in A\uparrow)}$ and construct another set of 30 samples J that does not exhibit the pattern of simultaneous low copy number of gene B and high copy number of gene A. Samples in J are chosen to be such that normalized SCNA value is 0 for both genes A, B. We compare

the Kaplan-Meier survival curves for patients in set I with those in set J. We also compare the survival through the stratified Cox Proportional Hazards Model with two covariates: (i) RNASeq expression values of the genes A, B and (ii) Genomic instability index (GII), calculated as the proportion of amplified or deleted genomic loci, as described in [7].

Results Tables 1 and 2 list the pairs of genes, in Breast and Stomach cancer respectively, selected by ASTER. Note that the pair PARP1-BRCA2, a well-known clinically validated SL pair [28] is identified by ASTER. Table 3 shows the results of fitting the stratified Cox Proportional Hazards model, and Fig. 4 and 5 show the survival plots for the top 4 gene pairs in Breast and Stomach cancer respectively. The plots show that samples with alterations identified by ASTER have significantly lower survival rates compared to samples without the alterations. A similar trend is seen through the Kaplan-Meier survival plots in Appendix A.

Table 1: Breast Cancer gene pairs selected by ASTER from candidates in SynLethDB with all 3 p-values < 0.01. The number of TCGA samples exhibiting up-regulation (↑) or down-regulation (↓) is shown in parentheses. Samples in state B are a subset of samples in state A.

Gene A		**Gene B**		**P1**	**P2**	**P3**
PARP1	↑(95)	BRCA2	↓(5)	2.61E-39	1.33e-04	1.88e-04
BRCA2	↓(14)	PARP1	↑(5)	5.86E-08	1.33e-04	1.19e-03
DHX9	↑(75)	ESCO2	↓(4)	1.23E-34	6.17e-04	8.32e-04
ESCO2	↓(43)	DHX9	↑(4)	9.21E-22	6.17e-04	1.17e-03
MED4	↓(14)	SRP9	↑(4)	1.29E-06	6.17e-04	2.94e-03
TNFRSF10B	↓(50)	FADD	↑(8)	1.16e-03	6.84E-06	4.50e-04
ILF2	↑(91)	RFC3	↓(3)	2.59E-40	3.04e-03	3.39e-03
S100A2	↑(91)	GPX8	↓(4)	8.04E-19	2.00e-03	6.15e-03
CAPN2	↑(95)	GPX8	↓(3)	2.92E-15	5.34e-03	3.41e-03
IKBKB	↑(82)	TNFRSF10A	↓(7)	1.66E-10	1.41e-03	3.58e-04
PTPN14	↑(84)	CTSB	↓(3)	4.54e-04	3.54e-03	3.48e-03

Table 2: Stomach Cancer gene pairs selected by ASTER from candidates in SynLethDB with all 3 p-values < 0.01. The number of TCGA samples exhibiting up-regulation (↑) or down-regulation (↓) is shown in parentheses. Samples in state B are a subset of samples in state A.

Gene A		**Gene B**		**P-Value1**	**P-Value2**	**P-Value3**
ABCC10	↑(28)	CDKN2A	↓(5)	1.12e-14	3.19e-04	2.01e-03
CDKN2A	↓(50)	ABCC10	↑(5)	4.60e-13	1.36e-04	2.56e-04
CDKN2A	↓(50)	SLC29A1	↑(6)	4.60e-13	2.21e-03	8.09e-05
SLC29A1	↑(27)	CDKN2A	↓(6)	1.01e-09	1.89e-04	1.89e-04
KRAS	↑(35)	MACROD2	↓(8)	4.38e-16	4.12e-03	5.05e-04
MACROD2	↓(46)	KRAS	↑(8)	3.00e-13	1.36e-04	3.27e-04
MYC	↑(53)	RAD51B	↓(3)	5.04e-17	3.76e-03	4.11e-03

Table 3: Cox proportional hazard model results for top 4 predicted SL pairs by ASTER. LR: Likelihood Ratio

Breast Cancer					Stomach Cancer				
Gene Pair	LR	P-Value	$\|I\|$	$\|J\|$	Gene Pair	LR	P-Value	$\|I\|$	$\|J\|$
PARP1/BRCA2	10.35	1.58e-02	14	30	ABCC10/CDKN2A	2.9	4.07e-01	5	30
DHX9/ESCO2	5.86	1.18e-01	4	30	CDKN2A/SLC29A1	3.91	2.71e-01	6	30
MED/SRP9	5.84	3.51e-02	4	30	KRAS/MACROD2	1.31	7.26e-01	7	30
TNFRSF10B/FADD	8.5	3.67e-02	8	30	MYC/RAD51B	1.42	7.00e-01	3	30

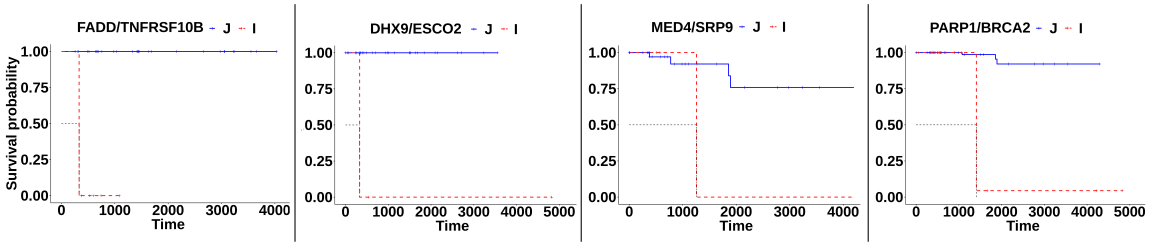

Fig. 4: Survival plots using Cox Proportional Hazards model stratified by group of samples belong to set I (those with the alterations) and J (those without the alterations) for top 4 predicted SL pairs in Breast Cancer Dataset by ASTER in Table 1.

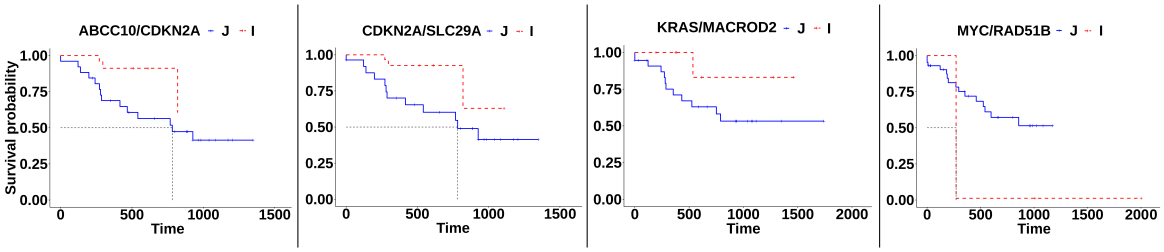

Fig. 5: Survival plots using Cox Proportional Hazards model stratified by group of samples belong to set I (those with the alterations) and J (those without the alterations) for top 4 predicted SL pairs in Stomach Cancer Dataset by ASTER in Table 2.

Functional Annotation Analysis The results of our functional annotation analysis, with respect to KEGG pathway and Gene Ontology – Biological Processes are detailed in Appendix B. We compare the most significant gene pairs (highest p-values) with the least significant gene pairs (lowest p-values) as identified by ASTER. We find that all the most significant pairs are found in pathways and biological processes that are known to be associated with cancer from previous studies. In contrast, pathways and biological processes that are enriched by the least significant pairs are not known to be directly associated with cancer.

3.2 Predictive Accuracy on Benchmark Datasets

We evaluate the efficacy of ASTER and ASTER++ in identifying SL pairs and compare their performance with state-of-the-art statistical methods DAISY and ISLE.

Data We use 3 benchmark datasets wherein SL interactions have been validated using CRISPR and/or shRNA screens. The first dataset consists of Breast Cancer SL pairs from SynLethDB [19] where we select those pairs that have evidence of SL from multiple sources including text mining, genomeRNAi and shRNA screens. The second dataset has 197 SL pairs from the functional study performed by [34] and [48], from three cell lines: lung, cervical and kidney cancer. The third dataset has 15,313 SL pairs from the functional study performed by [20] from leukemia cell line.

These datasets do not have negative samples (i.e., pairs that are not SL). To form a negative set, we randomly select genes from the HGNC database [8] after excluding genes reported in any SL interaction in SynLethDB and those reported to be essential in [41, 30]. We use 1000, 197 and 15,313 gene pairs as negative samples in the three datasets respectively.

Experiment Settings In ASTER, for comparing the expression levels of genes between normal and cancer tissues, the RNA-seq data of TCGA and GTEx as RSEM values that have been processed together (unifying process) with a consistent pipeline that helps to remove batch effects were used from RNASeqDB [42] and UCSC Xena [17].

For ASTER++, the covariates used are (1) loss of function mutation counts, (2) phylogenetic score, (3) methylation (HM450) beta-values and (4) the number of samples that exhibit up-/down-regulation for each gene based on the output results from ASTER. Loss of function mutation counts is the sum of non-synonymous mutations in each gene (excluding synonymous mutations such as 'Silent', 'Intron', '3'UTR', '5'UTR', 'IGR', 'lincRNA' and 'RNA'). Methylation value refers to human's gene methylation (HM450) beta-values (aberrant DNA methylation – hyper- or hypo-methylation – has been implicated in many disease processes, including cancer). Both loss of function mutations and methylation value were retrieved from cBioportal [15]. The phylogenetic score of a gene describes the relative sequence conservation or divergence of orthologous proteins across a set of reference genomes and has been used in several tasks such as gene annotation and function prediction. The phylogenetic score was retrieved from phylogenetic profile database by [37].

For the three tests in DAISY, we obtained SCNA, mRNA gene expression data and mutation profiles for cancer patients (BRCA, LUAD, CESC, KIRC, KIRP, KICH, AML) in TCGA [44] using cBioPortal [15] and Firehose. Essentiality profiles are based on those curated in [30]. DAISY uses a p-value cutoff of 0.05 after Bonferroni correction for multiple hypotheses testing. For ISLE, we use the software and data provided by them, using related cancer type data. We obtained phylogenetic similarity for 86 species using the phylogenetic profile database [37]. ISLE uses FDR < 0.2 based on Benjamini-Hochberg and a cut-off of 0.5 to determine phylogenetically linked pairs. ASTER, DAISY and ISLE, each yield three p-values per gene pair, that are combined using Fisher's method [14].

Evaluation Metric There is evidence of SL through experimental screens like CRISPR for the positive pairs in our datasets. However, for the randomly selected negative pairs in our data there is no evidence of their *not* being SL. Hence, they should be considered untested or unlabelled. This is similar to many other applications that have positive-unlabelled data, such as

recommendation systems. In such cases, metrics that use True Negatives or False Positives, e.g., Precision or AUC, are not reliable. A standard approach is to evaluate predictive algorithms through Recall@N, which is defined as the proportion of True Positives correctly identified in a ranked list containing the best scoring N items. In our experiments, the p-values are used as scores, with lower values indicating better scores.

Results Fig. 6 shows the performance of ASTER, DAISY, ISLE and ASTER++ on three benchmark datasets. On two datasets, ASTER outperforms both DAISY and ISLE while on the third, its performance is comparable to DAISY, the best performing method. When covariates are available, ASTER++ boosts the performance of ASTER in SynLethDB (left) but is indistinguishable, in performance, from ASTER in the other two experiments. Appendix C shows a table of the true positive counts for each dataset.

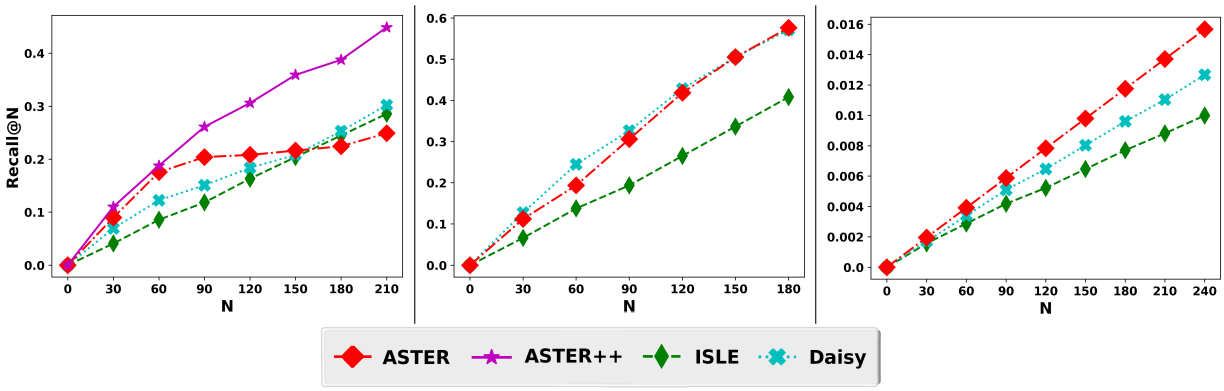

Fig. 6: Recall@N by ASTER, ASTER++, ISLE and Daisy in 3 benchmark datasets containing: (Left) 245 SL pairs from breast cancer cell lines, (Middle) 197 SL pairs from kidney, lung and cervical cell lines, (Right) 15,313 SL pairs from leukemia cell lines.

3.3 Therapeutic Actionability of Predicted SL Pairs

We evaluate the efficacy of ASTER in predicting clinically relevant SL pairs through two experiments.

Clinical Relevance of Predicted SL Pairs We consider 26 target genes in the in-vitro drug response screen covering 24 drugs in CCLE downloaded from UCSC Xena [17]. We combine these 26 genes with 32,018 genes from the human genome to form a total of 8,64,486 candidate gene pairs. These candidate pairs are used as input to ASTER, DAISY and ISLE and we evaluate the predicted pairs. To validate the methods we find the number of (target-partner) gene pairs in the predictions such that the loss of function in the partner gene is associated with sensitivity to the drug for the target gene. This information is given in the Genomics of Drug Sensitivity in Cancer (GDSC) drug response database [46], for breast cancer and stomach cancer types. Drug response is measured using IC50 and the drug IC50 effect size shows the effect of genomic features on sensitivity to the drugs, with a value < 1 indicating sensitivity to the drug [16].

Results We find that in all the predictions from ASTER, the partner genes have low copy number and are thus associated with loss of function. Table 4 shows the number of predicted pairs for which loss of function in the partner gene is associated with drug sensitivity for the target gene. We observe that ASTER finds many more such pairs compared to DAISY and ISLE. Appendix D has the complete list of identified pairs and the drugs that are effective on the target genes.

Table 4: Number of pairs of drugs and target genes in SL pairs from ASTER, DAISY and ISLE.

Method	Breast Cancer	Stomach Cancer
ASTER	52	133
DAISY	26	5
ISLE	0	0

Drug Efficacy Prediction We consider the task of drug efficacy prediction experiment following the experiments done in [23, 22]. The principle behind the experiment is that in a SL gene pair ($A \uparrow, B \downarrow$), if an inhibiting drug targets a gene (A), the expression level of the partner (B) is expected to correlate negatively with drug efficacy measured using IC50. This is because under-expression of B implies essentiality of A since they are SL. This allows us to indirectly evaluate our predicted SL interactions by analyzing their ability to predict drug efficacy.

We obtained drug efficacy profiles, measured using IC50 values, of 24 drugs from UCSC Xena [17] for 500 human cancer cell lines. For each method (ASTER, DAISY and ISLE) we perform the following test. For each drug, we find the target gene's top 5 and 10 SL partner genes predicted by each method, based on the most significant p-values. Let I be the cancer cell lines, out of 500, where the selected partner genes have less than the median level of expression (we consider these to be under-expressed). Let J be the drug efficacy (IC50) values for the drug inhibiting the target genes in the I cell lines. The two-sided Spearman correlation p-value between I and J is used as a measure of prediction accuracy. We compare the number of drugs selected for each method, when a cutoff of 0.05 is used for each method. In addition, we use the Benjamini-Hochberg false discovery rate (FDR) controlling procedure [5] with varying FDR thresholds (10, 20, 30, 40 and 50%) to find the number of unique drugs significantly correlated.

Results Fig. 7 shows that for Breast Cancer, ASTER outperforms all the baselines, at all 5 FDR values, in the number of significantly predicted drugs. In the case of Stomach Cancer, ASTER has the best performance when only gene expression is considered for the experiment and DAISY has the best performance when both SCNA and gene expression are considered. In Appendix E we list the number of drugs correlated with the top 5 and 10 predicted SL pairs for all the methods.

4 Discussion and Conclusion

We developed ASTER, a technique based on hypothesis testing, to identify SL pairs that leverages unified data from GTEx and TCGA. We also discussed how it can be extended, through the use of AdaFDR [47], for large-scale multiple hypothesis testing and to adaptively find a decision threshold based on additional input gene features. ASTER identifies SL in an input gene

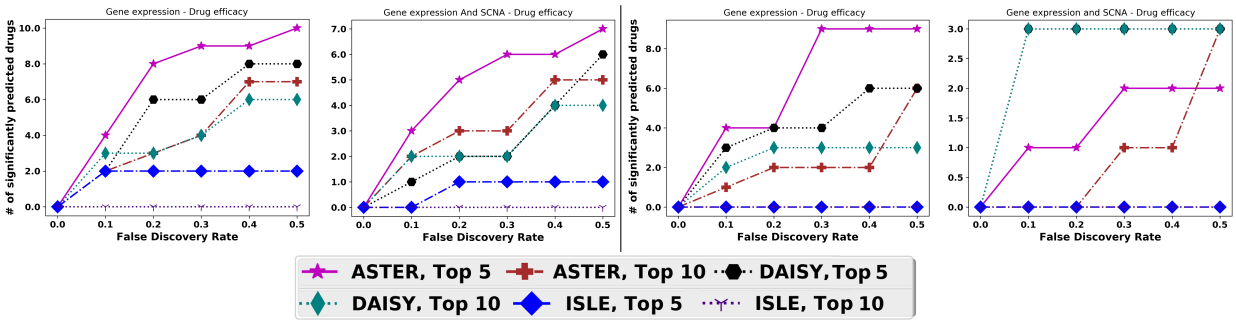

Fig. 7: Number of single-target drugs in the 24 drugs with 500 human cancer cell lines [17], whose efficacy is predicted with statistical significance at varying FDR levels based on gene expression and SCNA data from predicted SL interactions by ASTER, DAISY and ISLE, for top 5 and 10 candidate interactions for each drug target. Left (1-2): Breast Cancer; Right(3-4): Stomach Cancer.

pair through the application of 3 simple tests using only RNA-Seq and SCNA data. ASTER++ enables the use of additional gene features, when available, within the interpretable hypothesis testing framework.

We conducted three sets of experiments to evaluate ASTER on stomach and breast cancer data. Notably, ASTER was able to identify the well-known BRCA-PARP pair that DAISY and ISLE could not. Our first set of experiments showed that SL pairs identified by ASTER are associated with cancer-related pathways and cancer samples with ASTER-identified pairs have significantly lower survival rates compared to those without these alterations. In our second set of experiments, we evaluated the predictive accuracy of ASTER and ASTER++ on three benchmark datasets. On two datasets, ASTER outperforms both DAISY and ISLE while on the third, its performance is comparable to the best performing method. In the third set of experiments, we evaluated the clinical relevance of the SL pairs identified by ASTER. The number of clinically actionable pairs identified by ASTER were considerably higher than those identified by DAISY and ISLE. An implementation of ASTER is available at https://github.com/lianyh/ASTER.

We intentionally avoid the use of mutation-based data in ASTER. Due to intra- and inter-tumor heterogeneity of cancer samples, mutations in cancer samples are hard to validate and prone to high error rate, with higher false-positive rates for driver genes [2]. In contrast, gene expression signatures, especially through comparison with disease-free expression levels, provide robust signal for SL detection. A limitation of ASTER, similar to other approaches based on hypothesis testing, is its dependence on data from TCGA and GTex; the results are not reliable if sample sizes in the cancer type being tested are low. In the future, we plan to validate our predictions for previously untested gene pairs through CRISPR screens. Our framework can also be extended to incorporate additional drug-related information to improve its clinical actionability.

References

1. Audeh, M.W., Carmichael, J., Penson, R.T., Friedlander, M., Powell, B., Bell-McGuinn, K.M., Scott, C., Weitzel, J.N., Oaknin, A., Loman, N., et al.: Oral poly (adp-ribose) polymerase inhibitor olaparib in patients with brca1 or brca2 mutations and recurrent ovarian cancer: a proof-of-concept trial. The Lancet **376**(9737), 245–251 (2010)

2. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.: Comprehensive Characterization Of Cancer Driver Genes And Mutations. Cell **173**(2), 371–385 (2018)
3. Bekker, J., Davis, J.: Learning from positive and unlabeled data: a survey. Mach. Learn. **109**(4), 719–760 (2020)
4. Bender, A., Pringle, J.R.: Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in saccharomyces cerevisiae. Molecular and cellular biology **11**(3), 1295–1305 (1991)
5. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) **57**(1), 289–300 (1995)
6. Benstead-Hume, G., Chen, X., Hopkins, S.R., Lane, K.A., Downs, J.A., Pearl, F.M.: Predicting synthetic lethal interactions using conserved patterns in protein interaction networks. PLoS computational biology **15**(4), e1006888 (2019)
7. Bilal, E., Dutkowski, J., Guinney, J., Jang, I.S., Logsdon, B.A., Pandey, G., Sauerwine, B.A., Shimoni, Y., Vollan, H.K.M., Mecham, B.H., et al.: Improving breast cancer survival analysis through competition-based multidimensional modeling. PLoS computational biology **9**(5) (2013)
8. Bruford, E.A., Lush, M.J., Wright, M.W., Sneddon, T.P., Povey, S., Birney, E.: The HGNC Database in 2008: a resource for the human genome. Nucleic Acids Research **36**(suppl_1), D445–D448 (2007)
9. Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., Helleday, T.: Specific killing of brca2-deficient tumours with inhibitors of poly (adp-ribose) polymerase. Nature **434**(7035), 913–917 (2005)
10. Cai, R., Chen, X., Fang, Y., Wu, M., Hao, Y.: Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers. Bioinformatics (2020)
11. Conde-Pueyo, N., Munteanu, A., Solé, R.V., Rodríguez-Caso, C.: Human synthetic lethal inference as potential anti-cancer target gene detection. BMC Systems Biology **3**(1), 116 (2009)
12. Database, G.O.: Quickgoterm, https://www.ebi.ac.uk/QuickGO/term/GO:0008285
13. Farhan, M., Wang, H., Gaur, U., Little, P.J., Xu, J., Zheng, W.: Foxo signaling pathways as therapeutic targets in cancer. International journal of biological sciences **13**(7), 815 (2017)
14. Fisher, R.A.: 224a: Answer To Question 14 On Combining Independent Tests Of Significance. The American Statistician **2**(30) (1948)
15. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al.: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. **6**(269), pl1–pl1 (2013)
16. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al.: Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature **483**(7391), 570–575 (2012)
17. Goldman, M., Craft, B., Brooks, A., Zhu, J., Haussler, D.: The ucsc xena platform for cancer genomics data visualization and interpretation. BioRxiv p. 326470 (2018)
18. Guo, G.s., Zhang, F.m., Gao, R.j., Delsite, R., Feng, Z.h., Powell, S.N.: Dna repair and synthetic lethality. International journal of oral science **3**(4), 176–179 (2011)
19. Guo, J., Liu, H., Zheng, J.: Synlethdb: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. Nucleic acids research **44**(D1), D1011–D1017 (2016)
20. Horlbeck, M.A., Xu, A., Wang, M., Bennett, N.K., Park, C.Y., Bogdanoff, D., Adamson, B., Chow, E.D., Kampmann, M., Peterson, T.R., et al.: Mapping the genetic landscape of human cells. Cell **174**(4), 953–967 (2018)
21. Huang, D.W., Sherman, B.T., Lempicki, R.A.: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research **37**(1), 1–13 (2009)
22. Hyunghoon Cho, Bonnie Berger, J.P.: Compact integration of multi-network topology for functional analysis of genes. Cell Systems **3**(6), 540–548 (2016)
23. Jerby-Arnon, L., Pfetzer, N., Waldman, Y.Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P.A., et al.: Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell **158**(5), 1199–1209 (2014)
24. Kerr, J.F., Winterford, C.M., Harmon, B.V.: Apoptosis. its significance in cancer and cancer therapy. Cancer **73**(8), 2013–2026 (1994)
25. Lee, J.S., Das, A., Jerby-Arnon, L., Arafeh, R., Auslander, N., Davidson, M., McGarry, L., James, D., Amzallag, A., Park, S.G., et al.: Harnessing synthetic lethality to predict the response to cancer treatment. Nature Communications **9**(1), 2546 (2018)

26. Liany, H., Jeyasekharan, A., Rajan, V.: Predicting synthetic lethal interactions using heterogeneous data sources. Bioinformatics **36**(7), 2209–2216 (2020)
27. Liu, Y., Ao, X., Ding, W., Ponnusamy, M., Wu, W., Hao, X., Yu, W., Wang, Y., Li, P., Wang, J.: Critical role of foxo3a in carcinogenesis. Molecular cancer **17**(1), 104 (2018)
28. Lord, C.J., Ashworth, A.: Parp inhibitors: Synthetic lethality in the clinic. Science **355**(6330), 1152–1158 (2017)
29. Lucchesi, J.C.: Synthetic lethality and semi-lethality among functionally related mutants of drosophila melanogaster. Genetics **59**(1), 37 (1968)
30. Marcotte, R., Brown, K., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P., Sircoulomb, F., Medrano, M., Fedyshyn, Y., Koh, J., et al.: Essential gene profiles in breast, pancreatic, and ovarian cancer cells. Cancer Discovery 2: 172–189. doi: 10.1158/2159-8290 (2012)
31. O'Neil, N.J., Bailey, M.L., Hieter, P.: Synthetic lethality and cancer. Nature Reviews Genetics **18**(10), 613–623 (2017)
32. Organization, W.H.: Cancer fact sheet. https://www.who.int/news-room/fact-sheets/detail/cancer (2018)
33. Senft, D., Leiserson, M.D., Ruppin, E., Ze'ev, A.R.: Precision oncology: the road ahead. Trends in Molecular Medicine (2017)
34. Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al.: Combinatorial crispr–cas9 screens for de novo mapping of genetic interactions. Nature methods **14**(6), 573 (2017)
35. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2020. CA: a cancer journal for clinicians **70**(1), 7–30 (2020)
36. Srivas, R., Shen, J.P., Yang, C.C., Sun, S.M., Li, J., Gross, A.M., Jensen, J., Licon, K., Bojorquez-Gomez, A., Klepper, K., et al.: A network of conserved synthetic lethal interactions for exploration of precision cancer therapy. Molecular Cell **63**(3), 514–525 (2016)
37. Tabach, Y., Golan, T., Hernández-Hernández, A., Messer, A.R., Fukuda, T., Kouznetsova, A., Liu, J.G., Lilienthal, I., Levy, C., Ruvkun, G.: Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. Molecular systems biology **9**(1) (2013)
38. Technology, C.S.: Derbb / her signaling, https://www.cellsignal.com/contents/science-cst-pathways-kinase-signaling/erbb-her-signaling/pathways-erbb
39. Tutt, A., Robson, M., Garber, J., Domchek, S., Audeh, M., Weitzel, J., Friedlander, M., Carmichael, J.: Phase II trial of the oral PARP inhibitor olaparib in BRCA-deficient advanced breast cancer. Journal of Clinical Oncology **27**(18_suppl), CRA501–CRA501 (2009)
40. Vara, J.Á.F., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., González-Barón, M.: Pi3k/akt signalling pathway and cancer. Cancer treatment reviews **30**(2), 193–204 (2004)
41. Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., Lawo, S., et al.: A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. Molecular Systems Biology **9**(1) (2013)
42. Wang, Q., Armenia, J., Zhang, C., Penson, A.V., Reznik, E., Zhang, L., Minet, T., Ochoa, A., Gross, B.E., Iacobuzio-Donahue, C.A., et al.: Unifying cancer and normal rna sequencing data from different sources. Scientific data **5**, 180061 (2018)
43. Wang, X., Simon, R.: Identification of potential synthetic lethal genes to p53 using a computational biology approach. BMC medical genomics **6**(1), 30 (2013)
44. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Network, C.G.A.R., et al.: The cancer genome atlas pan-cancer analysis project. Nature genetics **45**(10), 1113 (2013)
45. Wu, M., Li, X., Zhang, F., Li, X., Kwoh, C.K., Zheng, J.: In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. Cancer Informatics **13**, CIN–S14026 (2014)
46. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al.: Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research **41**(D1), D955–D961 (2012)
47. Zhang, M.J., Xia, F., Zou, J.: Fast and covariate-adaptive method amplifies detection power in large-scale multiple hypothesis testing. Nature communications **10**(1), 1–11 (2019)
48. Zhao, D., Badur, M.G., Luebeck, J., Magaña, J.H., Birmingham, A., Sasik, R., Ahn, C.S., Ideker, T., Metallo, C.M., Mali, P.: Combinatorial crispr-cas9 metabolic screens reveal critical redox control points dependent on the keap1-nrf2 regulatory axis. Molecular cell **69**(4), 699–708 (2018)

Appendix: ASTER: A Method to Predict Clinically Actionable Synthetic Lethal Interactions

Herty Liany, Anand Jeyasekharan, Vaibhav Rajan

National University of Singapore

A Prognostic Value of Predicted SL Pairs

Kaplan-Meier survival plots for BRCA, stratified by group of samples belong to set I (those with the alterations) and J (those without the alterations) for top 4 predicted SL pairs in Table 1

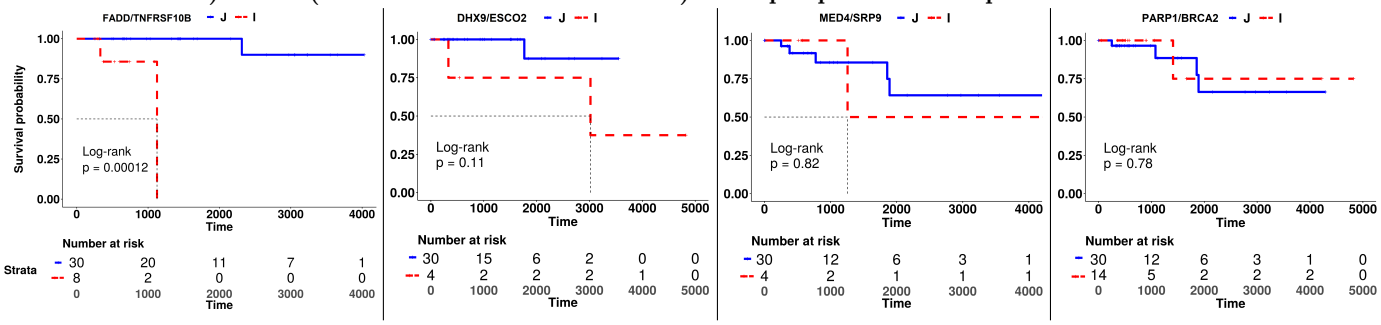

Fig. A.1: Kaplan-Meier survival plots for top 4 predicted SL gene pairs by ASTER

Kaplan-Meier survival plots for STAD, stratified by group of samples belong to set I (those with the alterations) and J (those without the alterations) for top 4 predicted SL pairs in in Table 2

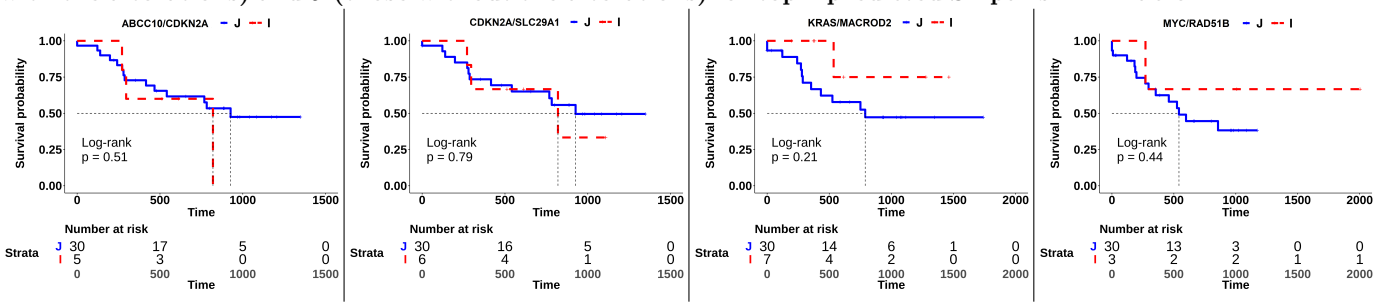

Fig. A.2: Kaplan-Meier survival plots for top 4 predicted SL gene pairs by ASTER

B Functional Annotation of Predicted SL Pairs

Functional Annotation of Genes in Predicted SL Pairs We use 16,916 gene pairs listed in SynLethDB [19] as input to ASTER. Separately for Breast and Stomach Cancers, we order the predictions based on the p-values and consider two sets of genes: (i) Set I: genes in the most significant predicted SL pairs (ii) Set J: genes in the least significant predicted SL pairs. For each set, we consider 90 genes for Breast Cancer and 24 genes for Stomach Cancer (with p-value < 0.05). We compare the functional annotation of these two sets of genes using DAVID version 6.8 [21] with respect to KEGG pathway and Gene Ontology (GO) - Biological Processes (BP).

Results

Table B.1: DAVID 6.8 Pathway Enrichment Analysis (KEGG) for predicted Top 90 SL pairs (BRCA) from SynLethDB by ASTER with p-value < 0.05

Pathway Name	**Genes**	**Gene** Count	**%**	**P-value**	**Benjamini**
Pathways in cancer	FGFR1, MYC, HRAS, ERBB2, IKBKB, TP53, BRCA2, FADD, RAD51, AKT1, CASP3, CDKN2A, KRAS, FGF2, RB1, CSF1R, FZD3, CKS1B, CYCS	19	19.4	5.8E-10	9.3E-8
Apoptosis	TNFRSF10A, AKT1, CASP3, TNFRSF10B, TP53, FADD, IKBKB, CAPN2, CYCS,	9	9.2	2.4E-8	1.9E-6
p53 signaling pathway	CDK1, CASP3, CDKN2A, CYCS, TP53, CHEK1	6	6.1	1.8E-4	2.0E-3
PI3K-Akt signaling pathway	FGFR1, HRAS, KRAS, TP53, MYC, RPS6KB1, IKBKB, BRCA1, EPHA2, CSF1R	10	10.2	1.5E-3	1.1E-2
Cell cycle	CDK1, CDKN2A, TP53, CHEK1, RB1, MYC	6	6.1	3.0E-3	2.0E-2
MAPK signaling pathway	FGFR1, CASP3, HRAS, KRAS, NTRK1, TP53, IKBKB, MYC	8	8.2	3.8E-3	2.4E-2
ErbB signaling pathway	HRAS, KRAS, ERBB2, RPS6KB1, MYC	5	5.1	5.0E-3	3.0E-2
Thyroid hormone signaling pathway	MED4, HRAS, KRAS, TP53, MYC	5	5.1	1.3E-2	6.8E-2
Neurotrophin signaling pathway	AKT1, HRAS, KRAS, TP53, IKBKB	5	5.1	1.5E-2	7.5E-2
Natural killer cell mediated cytotoxicity	TNFRSF10A, HRAS, CASP3, TNFRSF10B, KRAS	5	5.1	1.6E-2	7.7E-2
Homologous recombination	POLD4, BRCA2, RAD51	3	3.1	2.2E-2	1.0E-1
Ras signaling pathway	FGFR1, HRAS, KRAS, IKBKB, EPHA2, CSF1R	6	6.1	3.4E-2	1.4E-1
T cell receptor signaling pathway	PRKCQ, HRAS, KRAS, IKBKB	4	4.1	4.5E-2	1.8E-1

B.1: It has been known that ErbB receptors signal through MAPK, Akt and other pathways to regulate cell proliferation, migration, differentiation, apoptosis and cell motility [38]. ErbB family member genes are often over-expressed or highly amplified in cancer [38]. It has also been shown that MAPK signaling pathway plays important role in control of cell cycle and the control of cell cycle is p53 dependent [43]. It has also been reported that MAPK pathway related genes are potentially synthetic lethal to p53 [43]. ASTER'stop predicted SL genes show enrichment in these pathways: MAPK signaling, cell cycle and p53 signaling pathways. In addition, homologous recombination pathway is a major repair pathway for double-strand breaks (DSBs). HR–defective tumor cells are vulnerable to synthetic lethality if other DNA repair mechanisms in the HR–defective cells are inhibited [18].

Table B.2: DAVID 6.8 Pathway Enrichment Analysis (KEGG) for Top most insignificant predicted top 90 SL pairs (BRCA) from SynLethDB by ASTER ($\sim$ 90 gene pairs)

Pathway Name	Genes	Gene Count	%	P-value	Benjamini
Pathways in cancer	WNT5A, CBLC, VEGFC, HRAS, BCL2, NTRK1, PDGFRB, RARA, ZBTB16, MYC	10	16.1	7.2E-4	2.8E-2
Metabolism of xenobiotics by cytochrome P450	UGT1A7, AKR1C2, CYP1B1 , CYP2C9	4	6.5	7.7E-3	2.3E-1
Chemical carcinogenesis	UGT1A7, CYP1B1, CYP2C19, CYP2C9	4	6.5	9.5E-3	2.2E-1
Focal adhesion	VEGFC, HRAS, BCL2, PDGFRB, ACTN1	5	8.1	2.6E-2	3.2E-1
Neurotrophin signaling pathway	HRAS, BCL2, NTRK1, CALM1	4	6.5	2.8E-2	3.1E-1
PI3K-Akt signaling pathway	VEGFC, HRAS, CSH2, BCL2, PDGFRB, MYC	6	9.7	4.0E-2	3.9E-1
Insulin signaling pathway	CBLC, HRAS, PPP1R3A, CALM1	4	6.5	4.0E-2	3.6E-1
Steroid hormone biosynthesis	UGT1A7, AKR1C2, CYP1B1	3	4.8	4.0E-2	3.4E-1

Pathways in Table B.2: metabolism of xenobiotics, neurotrophin signaling, insulin signaling and steroid hormone biosynthesis pathways are not known to be directly involved in cancer. However, it has been reported that PI3K-Akt signaling pathway is associated with regulation of cell cycle progression and its alterations are frequent in human cancer [40]. PI3K-Akt signaling pathway is also found in the top most significant predicted SL gene pairs in Table B.1.

Table B.3: DAVID 6.8 Pathway Enrichment Analysis (KEGG) for predicted Top 24 SL pairs (STAD) from SynLethDB by ASTER with p-value < 0.05

Pathway Name	Genes	Gene Count	%	P-value	Benjamini
Pathways in cancer	CASP3, CDKN2A, KRAS, SMAD4, PIK3CA, RARA, PTEN, MYC	8	30.8	1.4E-5	7.6E-4
Homologous recombination	POLD4, RAD51B, XRCC2	3	11.5	2.0E-3	2.0E-2
Foxo signaling pathway	KRAS, SMAD4, PIK3CA, PTEN	4	15.4	3.4E-3	3.0E-2
MicroRNAs in cancer	CASP3, CDKN2A, KRAS, PTEN, MYC	5	19.2	3.6E-3	3.0E-2
Signaling pathways regulating pluripotency of stem cells	KRAS, SMAD4, PIK3CA, MYC	4	15.4	3.8E-3	3.0E-2
p53 signaling pathway	CASP3, CDKN2A, PTEN	3	11.5	1.0E-2	6.1E-2
ErbB signaling pathway	KRAS, PIK3CA, MYC	3	11.5	1.7E-2	8.5E-2
Cell cycle	CDKN2A, SMAD4, MYC	3	11.5	3.3E-2	1.3E-1

B.3: It has been reported that Foxo signaling pathway is a potential therapeutic target in gastric cancer [13]. FOXOs genes are involved in both pro- and anti-angiogenic factors. FOXO1 inactivation promotes angiogenesis in gastric cancer[13]. FOXO3 regulates vessel formation in the postnatal stage and interacts with the tumor suppressor p53 at different levels and it is required, at least partially, for p53-induced apoptosis [13]. In addition, it has been known that Micro-RNAs regulate FOXO levels in cancers [13]. Some micro-RNAs including miR-183, miR-182 and miR-96 act as regulators of FOXO expression in various cancer types. Micro-RNAs which down-regulate FOXO1 and support cancer cell proliferation and cell survival for example, miR-135b in osteosarcoma cells, miR-370 in prostate cancer, miR-411 in lung cancer and miR-1269 in hepatocellular carcinoma [13]. FOXO3a cooperates with RUNX3 to induce apoptosis by activating Bim in gastric cancer cells [27]. FOXO3a has the ability to suppress cancer cell proliferation by down-regulating the expression of several ER-relates genes, which are involved in cell cycle progression [27]. ASTER's top predicted genes show enrichment in these pathways: Foxo signaling pathway, Micro-RNAs, p53 signaling pathway and cell cycle.

Pathways in Table B.4: steroid hormone biosynthesis, retinol metabolism, cytochrome P450 and metabolism of xenobiotics, neurotrophin signaling pathways are not known or reported to be directly involved in cancer.

Table B.4: DAVID 6.8 Pathway Enrichment Analysis (KEGG) for Top most insignificant predicted gene pairs (STAD) from SynLethDB by ASTER ($\sim$ 24 gene pairs)

Pathway Name	Genes	Gene Count	%	P-value	Benjamini
Steroid hormone biosynthesis	CYP3A4, CYP3A5, CYP3A7	3	14.3	2.4E-3	9.0E-2
Retinol metabolism	CYP3A4, CYP3A5, CYP3A7	3	14.3	2.9E-3	5.6E-2
Chemical carcinogenesis	CYP3A4, CYP3A5, CYP3A7	3	14.3	4.6E-3	5.8E-2
Drug metabolism - cytochrome P450	CYP3A4, CYP3A5	2	9.5	8.6E-2	5.8E-1
Metabolism of xenobiotics by cytochrome P450	CYP3A4, CYP3A5	2	9.5	9.3E-2	5.3E-1

DAVID 6.8 GO:Biological Process Functional Analysis for predicted Top 90 SL pairs (BRCA) from SynLethDB by ASTER with p-value < 0.05

GO:BP Term	Gene Count	%	P-value	Benjamini
Positive regulation of transcription	15	15.3	1.0E-6	1.1E-3
Response to drug	12	12.2 7	1.0E-6	5.9E-4
Negative regulation of extrinsic apoptotic signaling pathway via death domain receptors	5	5.1	3.3E-5	1.2E-2
Response to lipopolysaccharide	8	8.2	3.7E-5	1.1E-2
DNA synthesis involved in DNA repair	5	5.1	4.2E-5	9.5E-3
DNA damage response	4	4.1	9.0E-5	1.7E-2
Activation of cysteine-type endopeptidase activity involved in apoptotic process	6	6.1	1.0E-4	1.6E-2
Regulation of extrinsic apoptotic signaling pathway via death domain receptors	4	4.1	1.1E-4	1.5E-2
Regulation of apoptotic process	8	8.2	1.9E-4	2.4E-2

Table B.5: GO:Biological Process Functional Analysis for BRCA - DAVID.

GO biological processes (BP) in Table B.5 are involved in regulation of apoptotic process, transcription, DNA damage response and reparation. These BP processes are highly involved in the regulation of cell cycle and cancer cell progression, especially the apoptotic process plays a critical role in cancer [24].

DAVID 6.8 GO:Biological Process Functional Analysis for the top most insignificant predicted SL pairs (BRCA) from SynLethDB by ASTER ($\sim$ 90 gene pairs)

GO:BP Term	Gene Count	%	P-value	Benjamini
Negative regulation of cell proliferation	9	14.5	5.4E-5	4.5E-2
Response to drug	8	12.9	7.2E-5	3.1E-2
Omega-hydroxylase P450 pathway	3	4.8	4.0E-4	1.1E-1
Steroid metabolic process	4	6.5	4.2E-4	8.6E-2
Positive regulation of fibroblast proliferation	4	6.5	8.1E-4	1.3E-1

Table B.6: GO:Biological Process Functional Analysis for BRCA - DAVID.

None of the biological processes (BP) in Table B.6 are known or reported to be directly involved in cancer progression. Negative regulation of cell proliferation is a process that stops, prevents or reduces the rate or extent of cell proliferation. Positive regulation of fibroblast proliferation is a process that activates or increases the frequency, rate or extent of multiplication or reproduction of fibroblast cells [12].

GO biological processes (BP) in Table B.7 are involved in positive regulation of cell proliferation, cell growth, receptor signaling pathway, epidermal growth factor and DNA synthesis involved in DNA repair.

DAVID 6.8 GO:Biological Process Functional Analysis for top 24 predicted SL pairs (STAD) from SynLethDB by ASTER with p-value < 0.05

GO:BP Term	Gene Count	%	P-value	Benjamini
Response to estradiol	4	15.4	3.2E-4	1.7E-1
Positive regulation of cell proliferation	6	23.1	5.4E-4	1.4E-1
Platelet-derived growth factor receptor signaling pathway	3	11.5	8.4E-4	1.5E-1
Response to gamma radiation	3	11.5	9.6E-4	1.3E-1
DNA synthesis involved in DNA repair	3	11.5	1.2E-3	1.3E-1
Cell proliferation	5	19.2	2.0E-3	1.7E-1
Epidermal growth factor receptor signaling pathway	3	11.5	3.1E-3	2.2E-1
Response to glucocorticoid	3	11.5	4.2E-3	2.5E-1
Nucleoside transport	2	7.7	4.5E-3	2.4E-1
Regulation of protein stability	3	11.5	4.8E-3	2.4E-1

Table B.7: GO:Biological Process Functional Analysis for STAD - DAVID.

DAVID 6.8 GO:Biological Process Functional Enrichment Analysis for the top most insignificant predicted gene pairs (STAD) from SynLethDB by GTexMeSL (∼ 24 gene pairs)

GO:BP Term	Gene Count	%	P-value	Benjamini
Lipid hydroxylation	3	14.3	1.4E-5	2.6E-3
Xenobiotic metabolic process	4	19.0	6.3E-5	5.7E-3
Steroid metabolic process	3	14.3	8.5E-4	5.1E-2
Alkaloid catabolic process	2	9.5	3.0E-3	1.3E-1
Drug catabolic process	2	9.5	6.1E-3	2.0E-1
Oxidative demethylation	2	9.5	1.2E-2	3.1E-1

Table B.8: GO:Biological Process Functional Analysis for STAD - DAVID.

None of the biological processes (BP) in Table B.8 (lipid hyroxylation, xenobiotic metabolic process, steroid and alkaloid metabolic processes) are known or reported to be directly involved in cancer progression.

C Predictive Accuracy on Benchmark Datasets

Table of TP counts for 245 (BC) SynlethDB (fig. 6)

N	ISLE	DAISY	ASTER	ASTER++
30	10	17	22	27
60	11	13	21	19
90	8	7	7	18
120	11	8	1	11
150	10	6	2	13
180	10	11	2	7
210	10	12	6	15

Table C.1: Recall@N for 245 known SL pairs in SynLethDB(BC).

Table of TP counts for Shen et.al's 197 SL pairs benchmark (kidney, lung and cervical cell lines) (fig. 6)

N	ISLE	DAISY	ASTER
30	13	25	22
60	14	23	16
90	11	16	22
120	14	20	22
150	14	15	17
180	14	13	14

Table C.2: Recall@N for 197 SL pairs (kidney, lung and cervical cell lines) by [34] and [48].

Table of TP counts for Horlbeck et. al.'s benchmark (leukemia cell lines) (fig. 6)

N	ISLE	DAISY	ASTER
30	24	26	30
60	20	26	30
90	20	26	30
120	16	21	30
150	19	24	30
180	19	24	30
210	17	22	30
230	18	25	30

Table C.3: Recall@N for 15,313 known SL pairs in Horlbeck dataset(leukemia cell line).

D Therapeutic Actionability of Predicted SL Pairs

Table D.1: Clinically relevant SL pairs predicted by ASTER in Breast Cancer (drug sensitivity of partner genes reported)

Method	Drug Name	Gene A (Target Gene)	Gene B (Partner Gene)	IC50 < 1 (Drug Sensitivity of Gene B)
ASTER	Paclitaxel	BCL2	PTEN	0.005126289
ASTER	Sorafenib	BRAF	ATAD1	0.607454911
ASTER	Sorafenib	BRAF	KLLN	0.607454911
ASTER	Sorafenib	BRAF	PTEN	0.607454911
ASTER	Sorafenib	BRAF	ANKRD6	0.76744379
ASTER	Sorafenib	BRAF	ATAD1	0.607454911
ASTER	Sorafenib	BRAF	KLLN	0.607454911
ASTER	Lapatinib	ERBB2	MAP2K4	0.219279138
ASTER	Lapatinib	ERBB2	CASP3	0.455555636
ASTER	Lapatinib	ERBB2	MICU2	0.241855679
ASTER	Lapatinib	ERBB2	CENPU	0.455555636
ASTER	Lapatinib	ERBB2	FRMD1	0.128991496
ASTER	Lapatinib	ERBB2	FAM149A	0.455555636
ASTER	Lapatinib	ERBB2	WDR27	0.128991496
ASTER	Lapatinib	ERBB2	STOX2	0.455555636
ASTER	Lapatinib	ERBB2	SNX25	0.455555636
ASTER	Lapatinib	ERBB2	DNAH9	0.219279138
ASTER	Lapatinib	ERBB2	THBS2	0.128991496
ASTER	Lapatinib	ERBB2	MRPL18	0.128991496
ASTER	Lapatinib	ERBB2	CCDC110	0.455555636
ASTER	Lapatinib	ERBB2	ING2	0.455555636
ASTER	Lapatinib	ERBB2	CLDN24	0.455555636
ASTER	Lapatinib	ERBB2	FRMD1	0.128991496
ASTER	Lapatinib	ERBB2	CENPU	0.455555636
ASTER	Lapatinib	ERBB2	CASP3	0.455555636
ASTER	Lapatinib	ERBB2	WDR27	0.128991496
ASTER	Lapatinib	ERBB2	FAM149A	0.455555636
ASTER	Lapatinib	ERBB2	MRPL18	0.128991496
ASTER	Lapatinib	ERBB2	KLKB1	0.455555636
ASTER	Lapatinib	ERBB2	GPR31	0.128991496
ASTER	Lapatinib	ERBB2	ING2	0.455555636
ASTER	Lapatinib	ERBB2	TCP1	0.128991496
ASTER	Lapatinib	ERBB2	DNAH9	0.219279138
ASTER	Lapatinib	ERBB2	CYP4V2	0.455555636
ASTER	Lapatinib	ERBB2	DACT2	0.128991496
ASTER	Lapatinib	ERBB2	ACAT2	0.128991496
ASTER	Lapatinib	ERBB2	RWDD4	0.455555636
ASTER	Lapatinib	ERBB2	MAP2K4	0.219279138
ASTER	Lapatinib	ERBB2	PDLIM3	0.455555636
ASTER	Lapatinib	ERBB2	ACSL1	0.455555636
ASTER	Lapatinib	ERBB2	IRF2	0.455555636
ASTER	Lapatinib	ERBB2	CCDC110	0.455555636
ASTER	Lapatinib	ERBB2	SMOC2	0.128991496
ASTER	Lapatinib	ERBB2	THBS2	0.128991496
ASTER	Lapatinib	ERBB2	LRP2BP	0.455555636
ASTER	Lapatinib	ERBB2	KIF25	0.128991496
ASTER	Lapatinib	ERBB2	SNX25	0.455555636
ASTER	Paclitaxel	TUBB1	ATAD1	0.005126289
ASTER	Paclitaxel	TUBB1	PTEN	0.005126289
ASTER	Paclitaxel	TUBB1	SNX9	0.799398286
ASTER	Paclitaxel	TUBB1	MAP2K4	0.260690274
ASTER	Paclitaxel	TUBB1	MAP2K4	0.260690274

Table D.2: Clinically relevant SL pairs predicted by DAISY in Breast Cancer (drug sensitivity of partner genes reported)

Method	**Drug Name**	**Gene A (Target Gene)**	**Gene B (Partner Gene)**	**IC50 < 1 (Drug Sensitivity of Gene B)**
DAISY	Nilotinib	ABL1	DLL1	0.415992145
DAISY	Paclitaxel	BCL2	SPCS3	0.144489103
DAISY	Paclitaxel	BCL2	DCTD	0.144489103
DAISY	Paclitaxel	BCL2	TRAPPC11	0.144489103
DAISY	Sorafenib	BRAF	MAP3K7	0.76744379
DAISY	Sorafenib	BRAF	MDN1	0.76744379
DAISY	Sorafenib	BRAF	ASCC3	0.76744379
DAISY	Sorafenib	BRAF	FBXL4	0.76744379
DAISY	Erlotinib	EGFR	LCA5	0.07286252
DAISY	Lapatinib	ERBB2	SLC25A4	0.455555636
DAISY	Selumetinib	ERBB2	SLC25A4	0.013386235
DAISY	Erlotinib	ERBB2	SLC25A4	0.534151326
DAISY	Panobinostat	HDAC9	ZDHHC20	0.25487677
DAISY	Sorafenib	KIT	NT5E	0.76744379
DAISY	Sorafenib	KIT	CCNC	0.76744379
DAISY	Sorafenib	KIT	ANKRD6	0.76744379
DAISY	Selumetinib	MAP2K1	TIAM2	0.14122658
DAISY	Selumetinib	MAP2K2	GPS2	0.001641818
DAISY	PHA-665752	MET	NT5E	0.272778084
DAISY	PHA-665752	MET	LCA5	0.272778084
DAISY	PHA-665752	MET	AKIRIN2	0.272778084
DAISY	Sorafenib	PDGFRB	PRSS35	0.76744379
DAISY	Sorafenib	PDGFRB	FHL5	0.76744379
DAISY	Irinotecan	TOP1	IGF2R	0.02216292
DAISY	Topotecan	TOP1	IGF2R	0.07563743
DAISY	Paclitaxel	TUBB1	MPC1	0.799398286

Table D.3: Clinically relevant SL pairs predicted by ASTER in Stomach Cancer (drug sensitivity of partner genes reported)

Method	Drug Name	Gene A (Target Gene)	Gene B (Partner Gene)	IC50 < 1 (Drug Sensitivity of Gene B)
ASTER	Erlotinib	EGFR	CCSER1	0.221546426
ASTER	Erlotinib	EGFR	CENPU	0.272876742
ASTER	Erlotinib	EGFR	SLC25A4	0.272876742
ASTER	Erlotinib	EGFR	STOX2	0.272876742
ASTER	Erlotinib	EGFR	SORBS2	0.272876742
ASTER	Erlotinib	EGFR	KLKB1	0.272876742
ASTER	Erlotinib	EGFR	WDR17	0.272876742
ASTER	Erlotinib	EGFR	MTNR1A	0.272876742
ASTER	Erlotinib	EGFR	NEIL3	0.272876742
ASTER	Erlotinib	EGFR	FAM149A	0.272876742
ASTER	Erlotinib	EGFR	CCDC110	0.272876742
ASTER	Erlotinib	EGFR	CCSER1	0.221546426
ASTER	Erlotinib	EGFR	IRF2	0.272876742
ASTER	Erlotinib	EGFR	AGA	0.272876742
ASTER	Erlotinib	EGFR	CASP3	0.272876742
ASTER	Erlotinib	EGFR	CDKN2B	0.214282813
ASTER	Erlotinib	EGFR	VEGFC	0.272876742
ASTER	Erlotinib	EGFR	CENPU	0.272876742
ASTER	Erlotinib	EGFR	HMGB2	0.272876742
ASTER	Erlotinib	EGFR	FAT1	0.272876742
ASTER	Erlotinib	EGFR	CDKN2A	0.214282813
ASTER	Erlotinib	EGFR	DMRTA1	0.214282813
ASTER	Erlotinib	EGFR	DMRTA1	0.214282813
ASTER	Erlotinib	EGFR	CDKN2A	0.214282813
ASTER	Erlotinib	EGFR	SLC25A4	0.272876742
ASTER	Erlotinib	EGFR	FAT1	0.272876742
ASTER	Erlotinib	EGFR	CASP3	0.272876742
ASTER	Erlotinib	EGFR	CDKN2B	0.214282813
ASTER	Erlotinib	EGFR	WDR17	0.272876742
ASTER	Erlotinib	EGFR	IRF2	0.272876742
ASTER	Erlotinib	EGFR	TLR3	0.272876742
ASTER	Erlotinib	EGFR	FAM149A	0.272876742
ASTER	Erlotinib	EGFR	RWDD4	0.272876742
ASTER	Erlotinib	EGFR	MTNR1A	0.272876742
ASTER	Erlotinib	EGFR	NEIL3	0.272876742
ASTER	Erlotinib	EGFR	CCDC110	0.272876742
ASTER	Erlotinib	EGFR	ING2	0.272876742
ASTER	Erlotinib	EGFR	KLKB1	0.272876742
ASTER	Erlotinib	EGFR	DCTD	0.272876742
ASTER	Erlotinib	EGFR	STOX2	0.272876742
ASTER	Erlotinib	EGFR	TENM3	0.272876742
ASTER	Erlotinib	EGFR	SORBS2	0.272876742
ASTER	Erlotinib	EGFR	TRAPPC11	0.272876742
ASTER	Erlotinib	EGFR	SNX25	0.272876742
ASTER	Erlotinib	EGFR	IFNE	0.214282813
ASTER	Erlotinib	EGFR	VEGFC	0.272876742
ASTER	Erlotinib	EGFR	AGA	0.272876742
ASTER	Erlotinib	EGFR	HELT	0.272876742
ASTER	Erlotinib	EGFR	GPM6A	0.272876742
ASTER	Erlotinib	EGFR	F11	0.272876742
ASTER	Erlotinib	EGFR	CLDN22	0.272876742
ASTER	Erlotinib	EGFR	HMGB2	0.272876742
ASTER	Erlotinib	EGFR	CDKN2AIP	0.272876742
ASTER	Erlotinib	EGFR	CYP4V2	0.272876742
ASTER	Erlotinib	EGFR	LRP2BP	0.272876742
ASTER	Erlotinib	EGFR	SPATA4	0.272876742
ASTER	Erlotinib	EGFR	GALNT7	0.272876742
ASTER	Erlotinib	EGFR	ASB5	0.272876742
ASTER	Erlotinib	EGFR	CEP44	0.272876742
ASTER	Erlotinib	EGFR	IFNA5	0.214282813
ASTER	Lapatinib	EGFR	CCSER1	0.310873204
ASTER	Lapatinib	EGFR	CENPU	0.149959523
ASTER	Lapatinib	EGFR	SLC25A4	0.149959523
ASTER	Lapatinib	EGFR	STOX2	0.149959523
ASTER	Lapatinib	EGFR	SORBS2	0.149959523
ASTER	Lapatinib	EGFR	KLKB1	0.149959523

Method	Drug Name	Gene A (Target Gene)	Gene B (Partner Gene)	IC50 < 1 (Drug Sensitivity of Gene B)
ASTER	Lapatinib	EGFR	WDR17	0.149959523
ASTER	Lapatinib	EGFR	MTNR1A	0.149959523
ASTER	Lapatinib	EGFR	NEIL3	0.149959523
ASTER	Lapatinib	EGFR	FAM149A	0.149959523
ASTER	Lapatinib	EGFR	CCDC110	0.149959523
ASTER	Lapatinib	EGFR	CCSER1	0.310873204
ASTER	Lapatinib	EGFR	IRF2	0.149959523
ASTER	Lapatinib	EGFR	AGA	0.149959523
ASTER	Lapatinib	EGFR	CASP3	0.149959523
ASTER	Lapatinib	EGFR	CDKN2B	0.308125525
ASTER	Lapatinib	EGFR	VEGFC	0.149959523
ASTER	Lapatinib	EGFR	CENPU	0.149959523
ASTER	Lapatinib	EGFR	HMGB2	0.149959523
ASTER	Lapatinib	EGFR	FAT1	0.149959523
ASTER	Lapatinib	EGFR	CDKN2A	0.308125525
ASTER	Lapatinib	EGFR	DMRTA1	0.308125525
ASTER	Lapatinib	EGFR	DMRTA1	0.308125525
ASTER	Lapatinib	EGFR	CDKN2A	0.308125525
ASTER	Lapatinib	EGFR	SLC25A4	0.149959523
ASTER	Lapatinib	EGFR	FAT1	0.149959523
ASTER	Lapatinib	EGFR	CASP3	0.149959523
ASTER	Lapatinib	EGFR	CDKN2B	0.308125525
ASTER	Lapatinib	EGFR	WDR17	0.149959523
ASTER	Lapatinib	EGFR	IRF2	0.149959523
ASTER	Lapatinib	EGFR	TLR3	0.149959523
ASTER	Lapatinib	EGFR	FAM149A	0.149959523
ASTER	Lapatinib	EGFR	RWDD4	0.149959523
ASTER	Lapatinib	EGFR	MTNR1A	0.149959523
ASTER	Lapatinib	EGFR	NEIL3	0.149959523
ASTER	Lapatinib	EGFR	CCDC110	0.149959523
ASTER	Lapatinib	EGFR	ING2	0.149959523
ASTER	Lapatinib	EGFR	KLKB1	0.149959523
ASTER	Lapatinib	EGFR	DCTD	0.149959523
ASTER	Lapatinib	EGFR	STOX2	0.149959523
ASTER	Lapatinib	EGFR	TENM3	0.149959523
ASTER	Lapatinib	EGFR	SORBS2	0.149959523
ASTER	Lapatinib	EGFR	TRAPPC11	0.149959523
ASTER	Lapatinib	EGFR	SNX25	0.149959523
ASTER	Lapatinib	EGFR	IFNE	0.308125525
ASTER	Lapatinib	EGFR	VEGFC	0.149959523
ASTER	Lapatinib	EGFR	AGA	0.149959523
ASTER	Lapatinib	EGFR	HELT	0.149959523
ASTER	Lapatinib	EGFR	GPM6A	0.149959523
ASTER	Lapatinib	EGFR	F11	0.149959523
ASTER	Lapatinib	EGFR	CLDN22	0.149959523
ASTER	Lapatinib	EGFR	HMGB2	0.149959523
ASTER	Lapatinib	EGFR	CDKN2AIP	0.149959523
ASTER	Lapatinib	EGFR	CYP4V2	0.149959523
ASTER	Lapatinib	EGFR	LRP2BP	0.149959523
ASTER	Lapatinib	EGFR	SPATA4	0.149959523
ASTER	Lapatinib	EGFR	GALNT7	0.149959523
ASTER	Lapatinib	EGFR	ASB5	0.149959523
ASTER	Lapatinib	EGFR	CEP44	0.149959523
ASTER	Lapatinib	EGFR	IFNA5	0.308125525
ASTER	Selumetinib	MAP2K1	CDKN2A	0.129647539
ASTER	Sorafenib	RAF1	CCSER1	0.69995009
ASTER	Irinotecan	TOP1	CCSER1	0.587624222
ASTER	Irinotecan	TOP1	CCSER1	0.587624222
ASTER	Topotecan	TOP1	CCSER1	0.42042701
ASTER	Topotecan	TOP1	CCSER1	0.42042701
ASTER	Paclitaxel	TUBB1	CCSER1	0.282704048
ASTER	Paclitaxel	TUBB1	CDKN2A	0.234907926
ASTER	Paclitaxel	TUBB1	CCSER1	0.282704048
ASTER	Paclitaxel	TUBB1	CDKN2B	0.234907926
ASTER	Paclitaxel	TUBB1	MTAP	0.234907926
ASTER	Paclitaxel	TUBB1	HMGB2	0.200799066
ASTER	Paclitaxel	TUBB1	GALNT7	0.200799066

Table D.4: Clinically relevant SL pairs predicted by DAISY in Stomach Cancer (drug sensitivity of partner genes reported)

Method	**Drug Name**	**Gene A (Target Gene)**	**Gene B (Partner Gene)**	**IC50 < 1 (Drug Sensitivity of Gene B)**
DAISY	Paclitaxel	BCL2	TRAPPC11	0.200799066
DAISY	Sorafenib	BRAF	FAT1	0.116420474
DAISY	Erlotinib	ERBB2	SLC25A4	0.272876742
DAISY	Selumetinib	ERBB2	SLC25A4	0.868505444
DAISY	Lapatinib	ERBB2	SLC25A4	0.149959523

E Drug Efficacy Prediction

Table E.1: (BC) The number of efficacy drugs correlated with top 5,10 predicted SL pairs

Predictors	Top 5	Top 10
ASTER	6	4
Daisy	5	3
ISLE	0	0

(BC) Predicted SL pairs with p-value $<$ 0.05 (Gene expression only)

Table E.2: (STAD) The number of efficacy drugs correlated with top 5,10 predicted SL pairs

Predictors	Top 5	Top 10
ASTER	6	2
Daisy	4	3
ISLE	0	0

(STAD) Predicted SL pairs with p-value $<$ 0.05 (Gene Expression only)

Table E.3: (BC) The number of efficacy drugs correlated with top 5,10 predicted SL pairs

Predictors	Top 5	Top 10
ASTER	4	3
Daisy	2	2
ISLE	0	0

(BC) Predicted SL pairs with p-value $<$ 0.05 (Gene expression and SCNA)

Table E.4: (STAD) The number of efficacy drugs correlated with top 5,10 predicted SL pairs

Predictors	Top 5	Top 10
ASTER	1	0
Daisy	3	2
ISLE	0	0

(STAD) Predicted SL pairs with p-value $<$ 0.05 (Gene expression and SCNA)