Computational and Structural Biotechnology Journal 20 (2022) 5287–5295

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

# Literature-based translation from synthetic lethality screening into therapeutics targets: CD82 is a novel target for *KRAS* mutation in colon cancer

Hsih-Te Yang [a], Ming-Yu Chien [b], Jung-Hsien Chiang [b,c], Peng-Chan Lin [d,e,*]

[a] Levine Cancer Institute, Atrium Health, Charlotte, NC 28204, USA
[b] Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University, Tainan 704, Taiwan
[c] Institute of Medical Informatics, National Cheng Kung University, Tainan 704, Taiwan
[d] Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan
[e] Department of Genomic Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan

ARTICLE INFO

*Article history:*
Received 15 June 2022
Received in revised form 19 September 2022
Accepted 19 September 2022
Available online 21 September 2022

*Keywords:*
*CD82*
*KRAS* mutation
Colorectal cancer
Text-mining
High-throughput screening

ABSTRACT

Synthetic lethality (SL) is an emerging therapeutic paradigm in cancer. We introduced a different approach to prioritize SL gene pairs through literature mining and *RAS*-mutant high-throughput screening (HTS) data. We matched essential genes from text-mining and mutant genes from the COSMIC and CCLE HTS datasets to build a prediction model of SL gene pairs. CCLE gene expression data were used to enrich the essential-mutant SL gene pairs using Spearman's correlation coefficient and literature mining. In total, 223 essential trigger terms were extracted and ranked. The threshold of the essential gene score ($S_g$) was set to 10. We identified 586 genes essential for the SL prediction model of colon cancer. Seven essential *RAS*-mutant SL gene pairs were identified in our model, including *CD82-KRAS/NRAS*, *PEBP1-NRAS*, *MT-CO2-HRAS*, *IFI27-NRAS/KRAS*, and *SUMO1-HRAS* gene pairs. Using *RAS*-mutant HTS data validation, we identified two potential SL gene pairs, including the *CD82* (essential gene)–*KRAS* (mutant gene) pair and *CD82–NRAS* pair in the DLD-1 colon cancer cell line (Spearman's correlation *p*-values = 0.004786 and 0.00249, respectively). Based on further annotations by PubChem, we observed that digitonin targeted the complex comprising *CD82*, especially in *KRAS*-mutated HCT116 cancer cells. Moreover, we experimentally demonstrated that *CD82* exhibited selective vulnerability in *KRAS*-mutant colorectal cancer. We used literature mining and HTS data to identify candidates for SL targets for *RAS*-mutant colon cancer.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

*KRAS* is the most frequently mutated oncogene in human cancers. Most *KRAS*-mutant cancers depend on the sustained expression and signaling of *KRAS*, making this gene a high-priority therapeutic target. Nevertheless, developing small-molecule direct inhibitors of the *KRAS* function remains an ongoing challenge [1]. An alternative therapeutic strategy for *KRAS*-mutant malignancies involves targeting codependent vulnerabilities or synthetic lethal partners essential for oncogenic *KRAS* [2]. Synthetic interactions between mutations in two different genes were first identified in *Drosophila* by Dobzhansky [3] in 1946. Synthetic lethality (SL) is defined as the biological consequence for a pair of viable genes if a cell with a mutation of either gene remains alive, but mutations or malfunctions of both genes would lead to cell death. SL has recently emerged as a novel strategy for the treatment of cancer. The interpretation of SL stipulates the downstream effects of two mutually expressed genes capable of performing the same essential function. In this regard, inhibiting a single gene is viable, whereas inhibiting both is lethal. The first clinical trial of SL-based treatment investigated *BRCA1/2*-deficient cancers, which

*Abbreviations:* BC, breast cancer; DAISY, data mining synthetic lethality identification pipeline; HTS, high-throughput screening; MSI-H, microsatellite instability; NLP, natural language processing; PARP, poly (ADP-ribose) polymerase; $S_g$, essential gene score; $S_t$, trigger term score; SL, synthetic lethality; WRN, Werner syndrome helicase.

* Corresponding author at: Department of Oncology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan 704, Taiwan.

*E-mail address:* pengchan@mail.ncku.edu.tw (P.-C. Lin).

https://doi.org/10.1016/j.csbj.2022.09.025
2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

have an SL interaction with poly (ADP-ribose) polymerase (PARP) [4,5]. Small PARP inhibitors are currently being designed for breast cancer treatment and are undergoing evaluation in clinical trials for *BRCA1/2*-deficient cancers [6]. Another example is Werner syndrome helicase (WRN) and high microsatellite instability (MSI-H) in cancer. Defects in DNA mismatch repair promote a hypermutable state referred to as microsatellites, which are unstable. Microsatellite instability contributes to the occurrence of several cancers, including colon (15 %), gastric (22 %), endometrial (20 %–30 %), and ovarian cancers [7]. WRN, a synthetic lethal partner of MSI-H, was identified using the DepMap database [8] and in an *in vitro* study [9].

*KRAS* activates numerous effector pathways that contribute to many potential synthetic lethal partners. A major challenge is the selection of appropriate candidate targets from thousands of SL gene pairs to determine the mechanism and potency of anti-cancer drugs. Many public SL databases have been employed to identify appropriate SL gene pairs. The dominant method, high-throughput screening (HTS) data surveys [10], enables the identification of many SL interactions in human cancer. However, the screening-based approach in human cancer is fraught with technical issues, including inconsistent cell lines and mechanistic interpretations, leading to false-positive SL candidates. In addition to these obstacles, most known SL cases have been discovered in yeast cells, but only a few SL gene pairs have been identified in humans. Indeed, it is challenging to select drug target candidates from these HTS-derived SL gene pairs simply based on their priority by ranking the score or signal of a reporter gene. Although more than 12 000 HTS-derived human SL gene pairs have been listed in the SynLethDB database [11], very few cases have become drug targets for cancer therapeutic discovery. Instead of identifying new SL gene pairs, we adopted a different approach by translating HTS-derived SL gene pairs into potential anti-cancer drug targets at the cellular level.

In this study, we aimed to identify novel therapeutic targets of essential genes for *RAS*-mutant colon cancer based on the concept of SL. We designed and implemented a text-mining model and experimental biological data to predict the critical SL gene pairs for specific cancer types. Moreover, we validated the anti-cancer potential of SL-based therapeutic targets through *RAS*-mutant colon cell line testing and a bioassay database [12,13].

## 2. Materials and methods

### 2.1. Study design and workflow

Our study comprised four main themes based on literature mining and HTS data. We designed an SL prediction model based on a text-mining method, which was modified in our previous study [14]. Fig. 1 depicts the proposed model architecture, which consists of four parts. (A) Essential genes were first identified by text-mining and HTS. We extracted a candidate list of essential genes using databases and biomedical literature mining from PubMed. (B) Essential SL gene pairs for cancer-specific types were matched and predicted. Since SL was identified by linking phenotype with genotype [10], we hypothesized that cell death could be mediated by an essential gene that was dysregulated at conduction, whereby another gene was mutated. Therefore, an SL gene pair could be modeled by building an association between an essential gene and a particular mutant gene. Cell lines were selected in the SL experimental screening approach based on whether they had one mutated gene (gene X). Next, the importance of other genes in these cell lines was tested. A potential SL (gene X and gene Y) gene pair could be predicted if one gene (gene Y) was essential in the

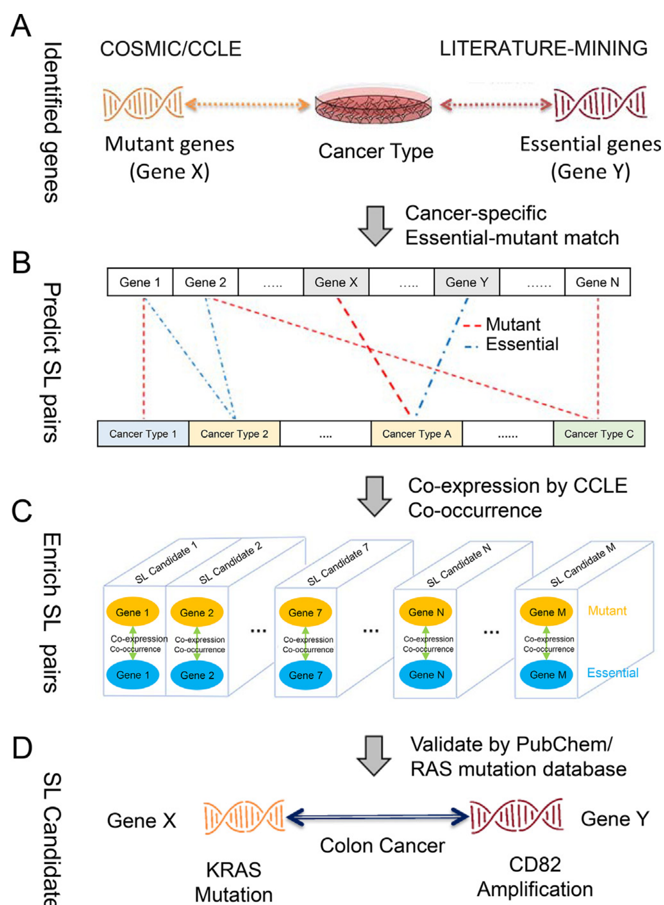



**Fig. 1.** Overview of the synthetic lethality (SL) prediction model. A. Identification of essential and mutant SL genes using text-mining and high-throughput screening (HTS) data (COSMIC and CCLE). B. Matching and prediction of essential-mutant SL gene pairs for cancer-specific types. C. Enrichment and filtering of SL gene pairs by gene co-expression and co-occurrences. D. Selection of candidate (*CD82-KRAS*) SL gene pairs using *RAS*-mutant HTS and PubChem bioassay data.

cancer cell lines that had mutations in a specific gene (gene X) but was not essential in other cell lines. (C) SL gene pairs were enriched and filtered using CCLE data and literature. (D) Appropriate candidate (i.e., *CD82-KRAS*) SL gene pairs were selected using *RAS*-mutant HTS and PubChem bioassay data [13].

### 2.2. Trigger term mining and ranking

To extract the biological relationships between genes and diseases from the literature, we identified essential genes using trigger terms as the main concept behind our text-mining method. First, we extracted trigger terms regarding "essentiality." A word was considered a trigger term if it was a common ancestor of two entities (referring to gene and disease herewith) in the dependency parse tree. We used the Stanford natural language processing (NLP) tool [15] to parse the dependency tree in the present study. COLT-Cancer [16], including breast, ovarian, and pancreatic cancers, was used as our initial essential gene database.

To rank the trigger terms, we divided sentences containing one gene name and one disease name into two sets: (i) essential relations, comprising sentences in which the gene was annotated as an essential gene in the disease; and (ii) other relations, comprising sentences in which the gene was not annotated as an essential gene in the disease. Subsequently, we used the following equation to compute the reliability score of each trigger term ($S_t$):

$$S_t(t) = \sum_{d \in D} \frac{E(t,d)}{MaxE(d)} * \frac{MaxO(d)+1}{O(t,d)+1}$$

$D$ represents the set of all cancer types in the essential database, $E(t,d)$ is the number of co-occurrences of trigger term $t$ and cancer type $d$ in essential relations, $O(t,d)$ is the number of co-occurrences of trigger term $t$ and cancer type $d$ in other relations, $MaxE(d)$ is the maximum number of co-occurrences in essential relations for disease $d$, and $MaxO(d)$ is the maximum number of co-occurrences in other relations for disease $d$.

### 2.3. Essential gene extraction and scoring

After extracting the essential trigger terms, we extracted potential essential genes with the essential trigger terms. Each potential essential gene had a common essential trigger term with other diseases. To filter out false essential genes, we computed and assigned a score ($S_g$) to each essential gene candidate using the following equation:

$$S_g(g,d) = \sum_{t \in T} S_t(t) * C(t,g,d)$$

where $S_g(g,d)$ denotes the score of the essential gene candidate $g$ in disease $d$, $T$ denotes the set of trigger terms, $S_t(t)$ is the score of trigger term $t$, and $C(t,g,d)$ is the number of co-occurrences when trigger term $t$ is the common ancestor of gene $g$ and disease $d$.

### 2.4. Matching essential-mutant SL gene pairs for colon cancer

We matched essential-mutant SL gene pairs in colon cancer after identifying cancer-specific essential genes using the text-mining method. We derived our mutant gene data from the well-known COSMIC [17] and CCLE datasets [18]. Using the COSMIC dataset, we selected gene names and tumor locations (e.g., the large intestine). We then identified each tumor location corresponding to the cancer type (e.g., colon cancer). Using the CCLE dataset, we obtained the pre-processed dataset containing mutant genes in specific cells using the cell line named entity recognition tool [19]. We then identified each cell line name corresponding to the cancer type and matched the essential-mutant SL gene pairs.

### 2.5. Enrichment and filtering of SL gene pairs

To extract more reliable predicted SL gene pairs, we used three criteria to rank the predicted SL gene pairs: (i) essential gene candidate score ($S_g$), (ii) gene co-expression, and (iii) number of co-occurrences (Fig. 2). We ranked each SL gene pair according to its essential gene candidate score ($S_g$) for the first criterion. We enriched each SL gene pair for the second criterion according to the gene co-expression obtained from the CCLE gene expression data; Spearman's correlation coefficient and $p$-value were used. For the third criterion, we filtered each SL gene pair according to the number of co-occurrences of two genes in one article.

### 2.6. RAS-mutant HTS database

Two *RAS*-mutant HTS datasets, including the HCT-116 cell line (*KRAS* mutation) with the *NAE1* gene study [20] and the DLD-1 cell line with *RAS* genes (*HRAS*, *NRAS*, and *KRAS*) study [21], were used to validate the SL gene pairs. The DLD-1 cell line is a genome-wide RNAi screen that permits the identification of multiple synthetic lethal interactions with the *RAS* oncogene. HTS experimental scores with log2 ratio differences were calculated to determine the correlation with *RAS* gene status [22].

### 2.7. PubChem bioassay database

PubChem is an open database of chemical and biological functional assays maintained by the National Institute of Health (NIH) that allows users to search for chemicals by name and identify chemical and physical properties biological activities safety and toxicity information. The literature search identified compounds targeting proteins (reporters) of the pathway or phenotype in a cell line under a particular culture condition. We annotated the compounds for SL (i.e. *CD82-KRAS*) candidate targets in the PubChem database [13] using the search keywords "digitonin" and "colon cancer." Using the COSMIC database we further identified the type of mutations in the *KRAS* gene in colon cancer cell lines.

### 2.8. Cell culture and cell viability assay

The isogenic colon cancer cell lines DLD-1 W/− (W: *RAS* wild-type, HD 105–002) and DLD-1 W/M (M: *KRAS* p.Gly13Asp mutation, HD PAR-086) [22] were obtained from Horizon Discovery Ltd. (Cambridge, UK). Cells were seeded in 96-well plates at a density of 100 cells/well in a complete medium and incubated overnight. A series of 5-FU concentrations (0 μM, 0.5 μM, 1 μM, and 10 μM) was added to the cultures, and the cells were incubated for a further 72 h. Cell viability was analyzed using the MTT assay and ATP bioluminescence assay kit (Promega, Madison, WI, USA).

## 3. Results

### 3.1. Identification of essential genes by trigger term mining in colon cancer

For the literature collection, we first downloaded literature published before 1 February 2016 from PubMed using Entrez Programming Utilities (E-utilities). We identified 783 975 articles that mentioned genes and 637 932 articles that mentioned specific cancers.

Our approach involved mining essential genes with trigger terms and parsing the dependency tree using the Stanford NLP tool [15]. Supplementary Fig. S1A presents an example sentence, "Her-2/neu gene amplification in familial vs sporadic breast cancer" (PMID: 14671981). In this example, "neu" is an essential gene in "breast cancer," and "amplification" is the common ancestor in the dependency parse tree. A total of 223 essential trigger terms were extracted and ranked (Supplementary Table S1). In total, 57, 52, and 99 essential trigger terms were extracted and ranked for breast, ovarian, and pancreatic cancers, respectively (Supplementary Tables S2, S3, and S4, respectively). Supplementary Fig. S1B presents the top seven trigger terms for breast, ovarian, and pancreatic cancers.

The $S_t$ was considered high if the trigger term appeared more in essential relations than in other relations. The frequency of the trigger term was considered higher in other relations than in essential relations when$S_t < 1$. This study used $S_t > 1$ to reduce any false-positive errors.

Our essential gene database consisted of three cancer types: breast, pancreatic, and ovarian. We selected two cancers as the training data and the third as the test data in this experiment. Supplementary Fig. S1B presents the precision/recall/F-score results for pancreatic, ovarian, and breast cancers. Supplementary Fig. S1B depicts the results tested on breast cancer and trained on two other cancers with different thresholds of the $S_g$, whereby a higher threshold reflects higher precision but a substantially lower recall. To reduce the number of false essential genes extracted by trigger terms and remove nonsignificant genes, we set the threshold of the $S_g$ to 10 (Fig. 2).

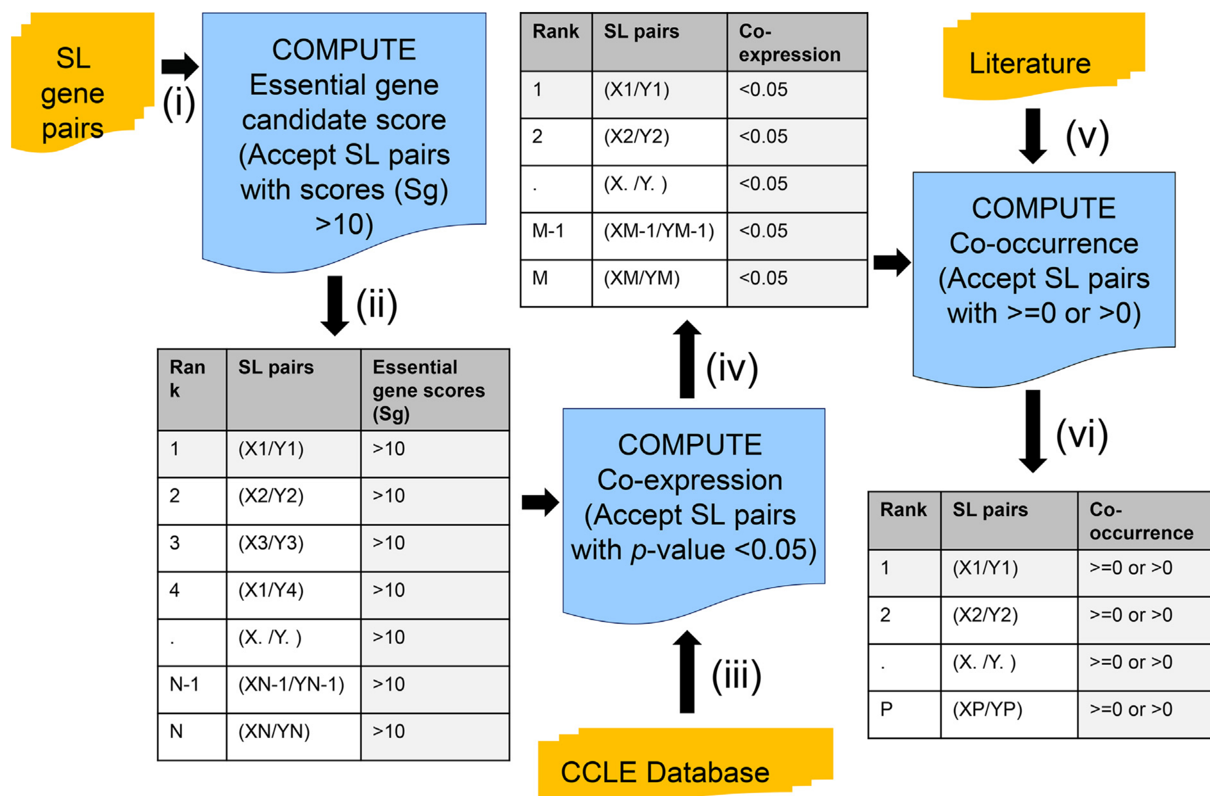



**Fig. 2.** Workflow of enrichment and filtering of synthetic lethality (SL) gene pairs. To extract and rank more reliable SL gene pairs, we used three criteria: (i) Essential gene candidate score ($S_g$), gene co-expression, and co-occurrences to enrich and filter SL gene pairs. In total, 586 essential genes were identified based on the threshold of the (ii) essential gene candidate score ($S_g$) > 10. Subsequently, we enriched and set Spearman's correlation (iv) $p$-value < 0.05 according to gene co-expression obtained from (iii) CCLE gene expression data. Next, we filtered each SL gene pair by co-occurrences (v) in the literature. The threshold of (vi) co-occurrences of two genes in one article was > 0 or $\geq$ 0. $S_g$, essential gene score.

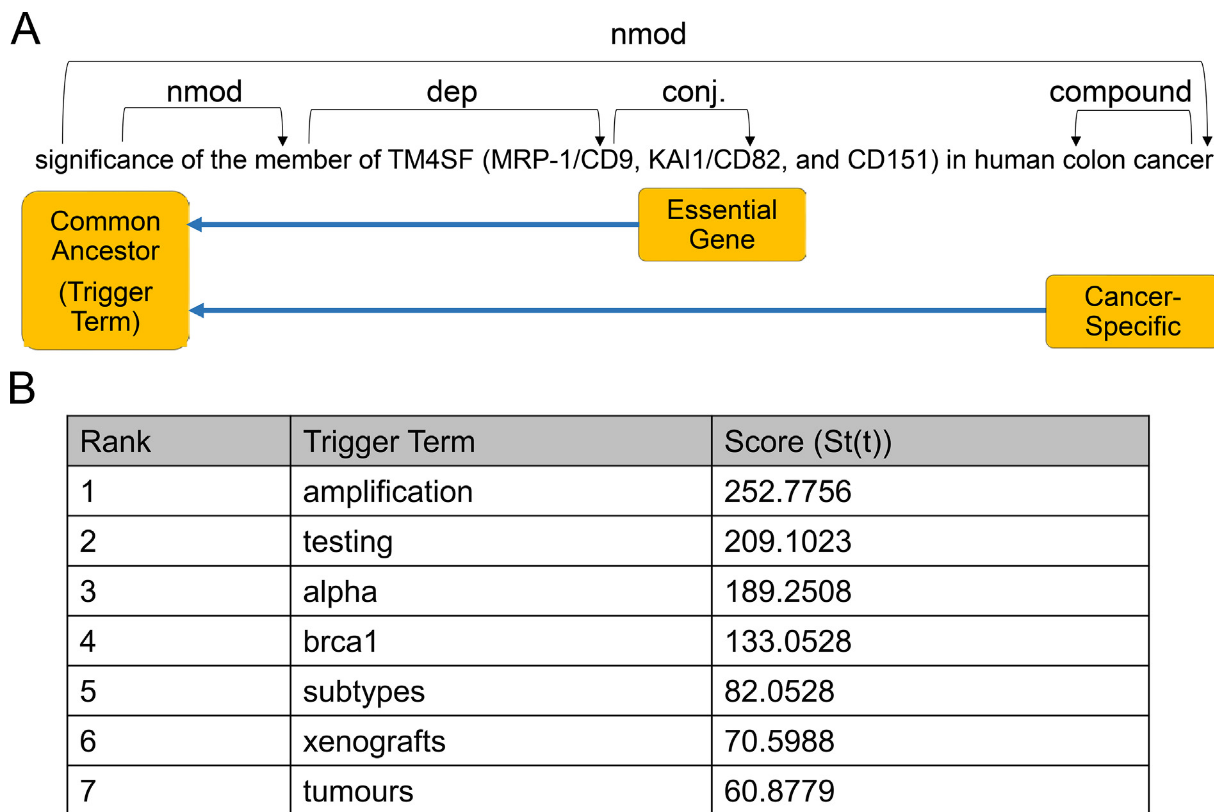
For colon cancer-specific text-mining, Fig. 3A presents the dependency parse tree for "significance of the member of TM4SF (MRP-1/CD9, KAI1/CD82, and CD151) in human colon cancer" (PMID: 12838318). In this example, "KAI1/CD82" is an essential gene in "colon cancer," and "significance" is the common ancestor in the dependency parse tree. Fig. 3B depicts the top seven trigger terms in colon cancer, including *amplification*, *testing*, *alpha*, *brca1*, *evaluation*, *xenografts*, and *tumors*. We identified 586 essential genes for SL gene pairs with $S_g$ > 10 in colon cancer (Supplementary Table S5).

### 3.2. Enrichment and filtering of matched essential-mutant SL gene pairs by gene co-expression and co-occurrence in colon cancer

We selected the appropriate essential-mutant SL gene pairs from our SL prediction model and matched the COSMIC/CCLE genetic mutation datasets with specific cancer types (Fig. 1). After selecting a suitable threshold, we used the CCLE dataset to enrich the SL gene pairs by co-expression. The SL gene pairs were identified by a cutoff Spearman's $p$-value < 0.05 (Fig. 2), and each SL pair was ranked by co-occurrence in the literature. Fig. 4 presents the number of gene pairs between our prediction model and the screening data in Venn diagram format with different thresholds. For colon cancer, we set a Spearman's correlation $p$-value < 0.05 and the threshold of the number of co-occurrences of two genes in one article $\geq$ 0 (Fig. 2), as shown in Fig. 4A (left). Our SL model predicted 16 SL pairs. However, we only set Spearman's correlation $p$-value < 0.05 and the threshold of the number of co-occurrences of two genes in one article > 0, as shown in Fig. 4B. We identified seven appropriate candidates for SL gene pairs in our model (Supplementary Table S6), including *CD82-KRAS/NRAS*, *PEBP1-NRAS*, *MT-CO2-HRAS*, *IFI27-NRAS/KRAS*, and *SUMO1-HRAS* SL gene pairs. The appropriate SL candidates were selected by enriching and filtering.

### 3.3. Validation of CD82-KRAS/NRAS SL gene pairs using RAS-mutant HTS data

Focusing on *RAS* mutations, we investigated SL in colon cancer using HTS screening data from the HCT116 cell line (*KRAS* mutation) with the *NAE1* gene study [20] and the DLD-1 cell line with *RAS* genes (*HRAS*, *NRAS*, and *KRAS*) study [20]. We validated notable cases in the intersection between our prediction model and the *RAS*-mutant HTS data. We identified two potential SL gene pairs, including the *CD82* (essential gene)–*KRAS* (mutation gene) and *CD82* (essential gene)–*NRAS* (mutation gene) pairs in colon cancer (Spearman correlation $p$-values = 0.004786 and 0.00249, respectively) (Fig. 4C and Supplementary Table S6). For gene co-expression, Spearman's correlation coefficients between *CD82* and *KRAS* mutations and between *CD82* and *NRAS* mutations were −0.329 and −0.355, respectively. The HTS experimental score for *CD82* and *KRAS/NRAS* was −1.23 in *RAS*-mutant HTS data [21]. The data revealed negative co-expression of *CD82* and *KRAS/NRAS* mutation genes (Fig. 4C and Supplementary Table S6). As *KRAS* and *NRAS* are oncogenes and *CD82* is a tumor suppressor gene, the results imply that the *CD82-KRAS/NRAS* mutation is a synthetic dosage-lethality pair. In this regard, the *CD82-KRAS/NRAS* genetic interaction, involving an under-expression of *CD82* combined with *KRAS/NRAS* mutations, could lead to cancer cell death.




| Rank | Trigger Term | Score (St(t)) |
|---|---|---|
| 1 | amplification | 252.7756 |
| 2 | testing | 209.1023 |
| 3 | alpha | 189.2508 |
| 4 | brca1 | 133.0528 |
| 5 | subtypes | 82.0528 |
| 6 | xenografts | 70.5988 |
| 7 | tumours | 60.8779 |

**Fig. 3.** Dependency parse tree of an example sentence and trigger term extraction. A. The dependency parse tree for "significance of the member of TM4SF (MRP-1/CD9, KAI1/CD82, and CD151) in human colon cancer." "KAI1/CD82" is an essential gene in "colon cancer," and "significance" is the common ancestor in the dependency parse tree. B. The top seven trigger terms in colon cancer. The nmod, dep, and conj. are Universal Stanford Dependencies representing grammatical relations between words. Abbreviations. nomd, nominal modifier; the nmod relation is used for nominal dependents of another noun or noun phrase and functionally corresponds to an attribute or genitive complement. Dep, unspecified dependency; a dependency can be labeled as dep when it is impossible to determine a more precise relation. Conj, conjunct; a conjunct is a relation between two elements connected by a coordinating conjunction, such as and, or, etc. The links for the main organizing principles of the Universal Dependencies taxonomy were as follows: https://universaldependencies.org/u/dep/all.html (accessed on 2 Feb 2022). $S_t$, trigger term score.

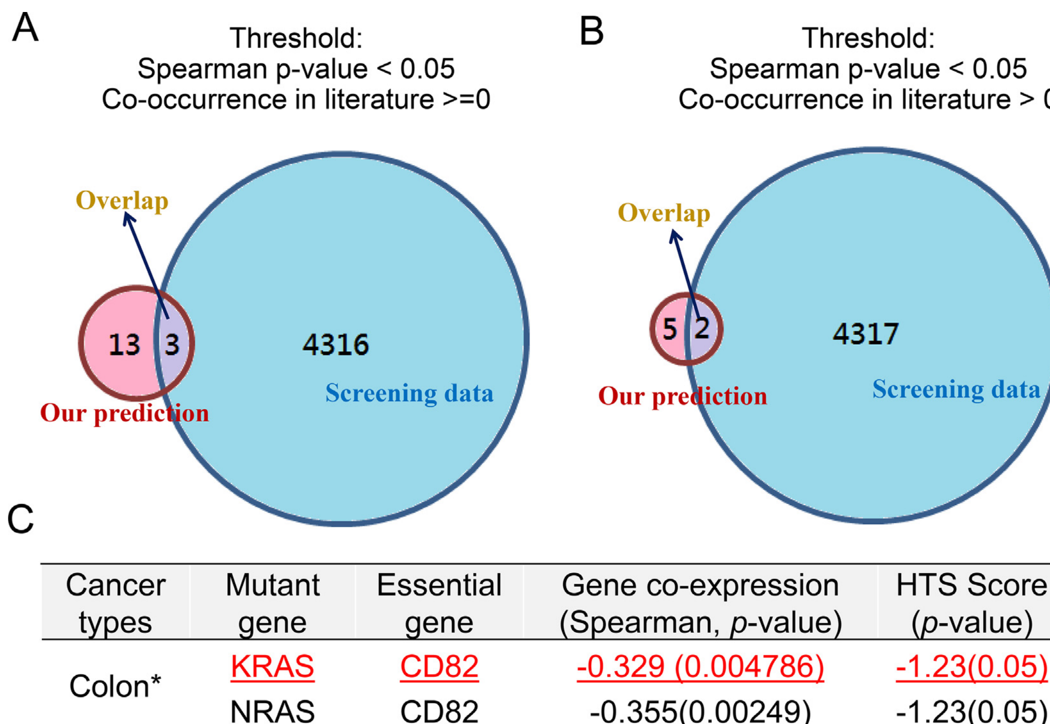
Our approach enabled matching essential-mutant SL gene pairs from the biomedical literature and COSMIC and CCLE datasets. For instance, *KRAS* is a mutant gene in colon cancer, and *CD82* is an essential gene extracted using the text-mining method. Thus, we predicted and verified that *KRAS* and *CD82* constituted an SL pair in colon cancer. Various unknown SLs were extracted in screening experiments. We selected key SLs from the screening data, and our prediction system permitted the selection of the most important SL gene pairs.

### 3.4. Identification of KRAS mutation drug candidates using a bioassay database

To investigate whether *CD82-KRAS* SL gene pairs had previously been reported in bioassay studies, we searched the bioassay database in PubChem (Fig. 5A), with *CD82* as the target. The search revealed a tetraspanin-cholesterol interaction via digitonin [12]. "Digitonin" was identified as an active substance in three bioassays, in which HT29 (AID: 417513, IC50 (half maximal inhibitory concentration) = 5.6 µM), CC20 (AID: 417511, IC50 = 9.6 µM), and HCT116 (AID: 417512, IC50 = 8.7 µM) were the target cell lines [13,23] (Fig. 5B). Among the cell lines, HCT116 carries the *KRAS*-mutant (G13D) gene (reported in the COSMIC database) (Fig. 5B), and HT29 carries the *BRAF* mutation (V600E), a downstream gene of *KRAS* within the same pathway. These results implied that depletion of tetraspanin family members (such as *CD9*, *CD82*, and *CD151*) by digitonin suppressed *KRAS/BRAF*-mutant colon cancer cell lines and that *CD82* was a target of digitonin, especially in *KRAS*-mutated HCT116 cancer cells.

### 3.5. Knockdown of CD82 expression in KRAS-mutant isogenic cell lines

The *RAS*-mutant DLD-1 isogenic cell line was used to validate the biological relevance of *CD82* in SL. *CD82*, a tetraspanin family member, contributes to the chemoresistant phenotype of cancers [24]. To experimentally corroborate the *CD82* dependency of *KRAS*-mutant cancer cells, we applied siRNA-mediated knockdown of *CD82* and assessed the interaction of *CD82* and *KRAS* based on MEK activation in CRC cell lines (Fig. 6). Knockdown of *CD82* resulted in a modest reduction in MEK phosphorylation in *KRAS*-mutant cells compared to that in *KRAS* wild-type and control cells. CRC cell lines were transfected with specific siRNA targeting *CD82*, and the impact on cell viability following 5-FU chemotherapy was determined using the MTT assay. Treatment with *CD82* siRNA significantly reduced the level of p-MEK in DLD-1 W/M cells compared to that in cells treated with scrambled siRNA (Fig. 6A). DLD-1 W/− cells transfected with *CD82* siRNA exhibited a significant increase in p-MEK levels. A greater reduction in cell viability was observed in *CD82* siRNA-treated DLD-1 W/M cells than in DLD-1 W/− cells (Fig. 6B and Supplementary Table S7). In the bar graph, siRNA targeting CD82 affects the viability of isogenic CRC cells following 5-FU chemotherapy using the MTT assay. A series of doses of 5-FU was administered to cancer cells (0.5 µM, 1 µM, and 10 µM). The viability of DLD-1 W/M cells treated with CD82 siRNA was significantly lower than that of scrambled siRNA-treated cells. The *p*-values for 5FU concentrations of 0.5 µM and 1 µM are 1.06E-07 and 1.63E-05, respectively (Supplementary Table S7). There was no significant difference in cell viability at high doses of 5-FU (10 µM). *CD82* depletion by siRNA




| Cancer types | Mutant gene | Essential gene | Gene co-expression (Spearman, *p*-value) | HTS Score (*p*-value) |
| --- | --- | --- | --- | --- |
| Colon* | KRAS | CD82 | -0.329 (0.004786) | -1.23(0.05) |
| | NRAS | CD82 | -0.355(0.00249) | -1.23(0.05) |

**Fig. 4.** Identification of *KRAS*-mutant synthetic lethality (SL) gene pairs. The synthetic lethal relationships of these genes were identified using *RAS*-mutant high-throughput screening (HTS) data and text-mining analysis. Predicted SL gene pairs were compared with HTS data using a Venn diagram with the threshold of gene co-expression and co-occurrences in colon cancer (A and B). The left circle (red) denotes the number of predicted SL gene pairs; the right circle (blue) represents the number of SL gene pairs recorded in the screening data. A. Stricter thresholds were set for gene co-expression (Spearman correlation *p*-value < 0.05) and co-occurrence (number of co-occurrences of two genes in the literature ≥ 0). B. Stricter thresholds were set for gene co-expression (Spearman correlation *p*-value < 0.05) and co-occurrence (number of co-occurrences of two genes in the literature > 0). C. Two potential SL gene pairs in colon cancer, including *CD82* (essential)-*KRAS* (mutant) and *CD82* (essential)-*NRAS* (mutant) gene pairs, were identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

selectively impaired the viability of *KRAS*-mutant cancer cell lines. These results indicated that CD82 dependency was associated with *KRAS* mutation status in cancer cells.

## 4. Discussion

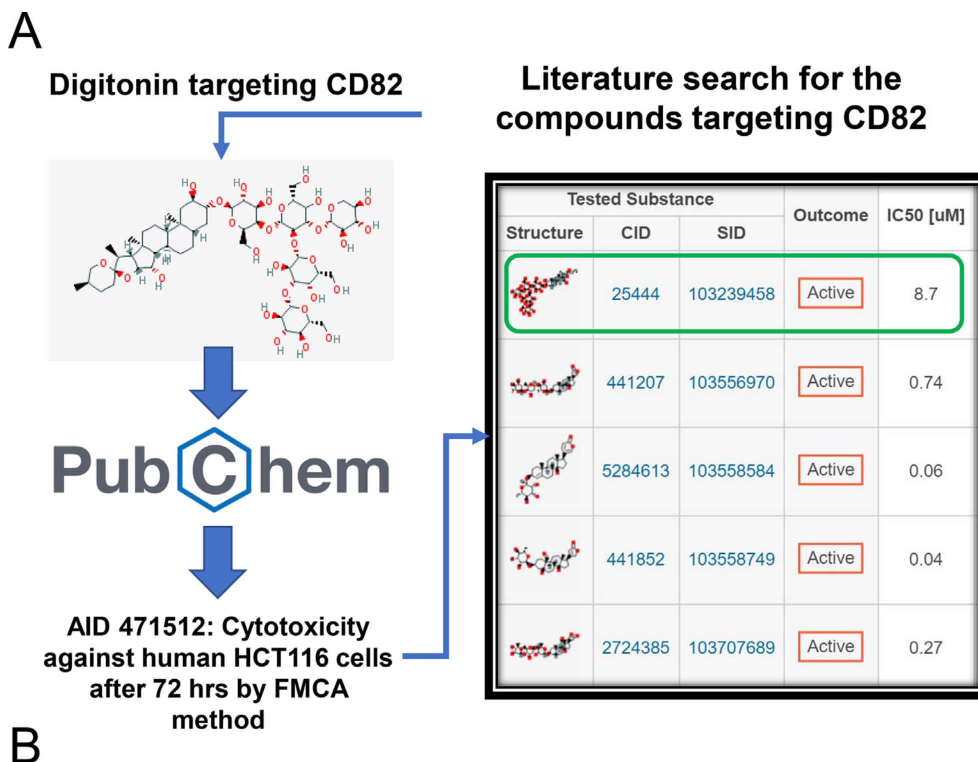
We conducted a comprehensive SL prediction model and identified *CD82-KRAS* SL gene pairs in colon cancer. Our results highlighted the following important points: (i) We designed and implemented a text-mining model to predict more potential SL gene pairs in cancer-specific types. (ii) We compared potential SL gene pairs with screening data and identified the *CD82* gene as a novel target for colon cancer with *KRAS*/*NRAS* mutation by synthetic lethality. (iii) We demonstrated digitonin as a potential therapeutic agent via the bioassay database for *KRAS*-mutant colon cancer cell lines. (iv) Finally, the *CD82-KRAS* is an essential SL gene pair validated *in vivo*.

*KRAS* and *NRAS* are well-known mutated genes in most cancers, including leukemia and colon, pancreatic, and lung cancers. However, no studies to date have provided direct evidence that these *CD82* mutations constitute SL gene pairs. Functional genetic screening approaches, including RNAi or CRISPR-Cas9, have been used for *KRAS* synthetic lethal targets [2]. Previous studies used this approach to identify therapeutic targets for a cancer type with a specific mutant gene using HTS, resulting in SL databases and information resources. Several screening technologies have been developed to detect SL interactions in yeast cells [25], human cell lines [26,27], and malignant tissues [28]. Several published algorithms use cancer genomic and multiomic data to predict SL interactions [29,30]. However, their scope remains insufficient to include the comprehensive functional and experimental studies of genetic interactions that need to be evaluated. In this regard, a literature-based translation from SL screening to therapeutic targets may provide more functional and experimental data for SL candidates.

In order to identify and find relationships between essential genes and specific cancers, we used text-mining to analyze their co-occurrence frequency in biomedical literature.

The literature mining and screening-based approach for identifying SL gene pairs in human cancer have several issues that should be addressed, such as false-positive results and inconsistencies for different cancer cell types. Many SL gene pairs in HTS experimental data have been identified and ranked; however, there is a paucity of appropriate SL candidates for cancer therapeutic discovery. As an alternative to the HTS experimental score, we introduced literature mining as a distinct approach to prioritize SL candidates and identified three SL gene pairs. With further annotations based on public data sources of gene expression and bioactivity, a *CD82-KRAS* SL gene pair in colon cancer was identified. This permitted the selection of novel SL gene pairs, although they were not included in the top-ranking candidates in the HTS experimental data. Moreover, we experimentally demonstrated that the knockdown of *CD82* increased the sensitivity of *KRAS*-mutant colon cells in response to regular cancer drug treatment with 5-FU. Our approach will enable researchers to narrow the size of notable SL pairs. Indeed, our literature-based identification of anti-cancer targets from HTS data and bioassays provides an alternative method to unveil novel modes of action for modern cancer therapeutics.

Compared with the data mining SL identification pipeline (DAISY), another data-driven SL prediction system [31], our approach permitted the identification of more SL gene pairs in




| Tested Substance | | | Outcome | IC50 [uM] |
|---|---|---|---|---|
| Structure | CID | SID | | |
| | 25444 | 103239458 | Active | 8.7 |
| | 441207 | 103556970 | Active | 0.74 |
| | 5284613 | 103558584 | Active | 0.06 |
| | 441852 | 103558749 | Active | 0.04 |
| | 2724385 | 103707689 | Active | 0.27 |

| Colon cancer cell line | PubChem AID | Activity | PubChem Standard Value, uM ($IC_{50}$) | *KRAS* mutation (COSMIC database) |
|---|---|---|---|---|
| HCT116 | AID 471512 | Active | 8.7 | ***KRAS***:p.G13D |
| CC20 | AID 471511 | Active | 9.6 | None |
| HT-29 | AID 471513 | Active | 5.6 | ***BRAF***:p.Val600Glu** |

** A downstream gene of ***KRAS*** within the same pathway

**Fig. 5.** PubChem bioassay of the *CD82-KRAS* synthetic lethality (SL) gene pair in colon cancer. A. A PubChem database search revealed that a chemical compound screen assay (fluorometric microculture cytotoxicity assay (FMCA)) demonstrated that digitonin was used to precipitate tetraspanins (a protein complex including *CD82*) and significantly suppressed the growth of a *KRAS/BRAF*-mutant colon cancer cell line (https://pubchem.ncbi.nlm.nih.gov/bioassay/471512). B. The half-maximal inhibitory concentration (IC50) values for three cancer cell lines are presented.

different cancers because we extended essential genes with diseases using text-mining in the published literature. Limited experimental data were available on essential genes (147 cell lines) compared to mutant genes (approximately 1600 cell lines). We filtered with a stricter threshold regarding co-occurrence in the literature to remove incorrect SL gene pairs and enriched with gene co-expression using the CCLE database. With advances in treatable *KRAS* mutations for cancer therapy, cancer-specific SL should be considered in SL prediction [32]. DAISY predicted potential SL using mutant genes, essential genes, and gene expression in non-specific cancers. In addition to the quantity of experimental data, only three cancer types (breast, ovarian, and pancreatic cancers) have been examined in experiments on essential genes. Owing to the lack of experimental data on essential genes in colon cancer, we utilized the trigger term-based method to extend the quantity of potential essential genes and identified two notable SL gene pairs at the intersection of *RAS*-mutant screening data in colon cancer. Furthermore, this method used the co-occurrence of entities and trigger terms to identify additional relationships between essential genes and specific cancers.

Recently, several drugs that directly target *KRAS* have been investigated in clinical trials. Sotorasib can be used to treat patients with non-small cell lung cancer with a particular *KRAS* mutation subtype, referred to as the G12C mutation [33]. However, a single-arm phase II trial reported a lack of efficacy in CRC patients with the *KRAS* (G12C) mutation [34]. In this regard, the same mutations in cancer cells in the colon or lung may lead to different therapeutic outcomes. The spectrum and distribution of *KRAS* mutation subtypes also differ for different cancer types. Despite the clinical success of targeting *KRAS* mutations in patients with lung cancer, additional efforts are warranted to identify drugs to treat CRC patients with *KRAS* mutations.

In this study, we identified seven appropriate candidates for *RAS*-mutant SL gene pairs and verified that *CD82-KRAS/NRAS* gene pairs had SL in a bioassay database of colon cancer cell lines. Digitonin (CID: 25444), a cholesterol-precipitating reagent, can inhibit tetraspanins *CD9*, *CD81*, and *CD82* [11] and be used to verify *CD82-KRAS* as an SL gene pair via bioassay. We searched the PubChem database with the keywords "digitonin," "*KRAS*," and "colon cancer," leading to three digitalis glycosides candidate bioassays

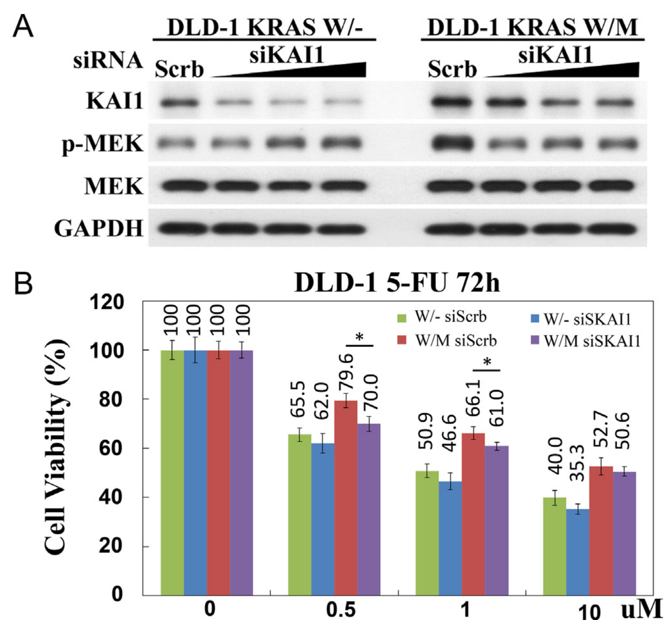



**Fig. 6.** Effects of siRNA-mediated *CD82* depletion on cell viability of *KRAS*-mutant cancer cell lines. siRNA-mediated knockdown of *CD82* resulted in the downregulation of p-MEK and decreased cell viability following chemotherapy. A. Western blot (WB) analysis of DLD-1 W/M (*KRAS* mutant) and DLD-1 W/− (*KRAS* wild-type). *CD82* expression was higher in DLD-1 W/M cells than in DLD-1 W/− cells. The colon cancer cell lines were transiently transfected with control siRNA or *CD82* siRNA. The WB results of siRNA-*CD82* in DLD-1 W/− and W/M cells. B. Results of the MTT assay after *CD82* knockdown and treatment with 5-FU for 72.

as the output. The three digitalis glycosides include digitonin, digitoxin (CID: 441207), and digoxin (CID: 2724385) (Fig. 5A) [35,36]. Digitonin, digitoxin, and digoxin were identified as active substances in HCT116 cancer cells (IC50 = 8.7 μM, 0.74 μM, and 0.27 μM, respectively). As the three digitalis glycosides share similar 2D structures, they may be able to target *CD82* leading to SL in *KRAS*-mutant colon cancer cells. However, no literature reports an association between *CD82* and digitoxin or digoxin. Furthermore, digitonin is an active compound in *RAS*-mutant colon cancer cell lines [13]. Lower *CD82* expression and high miR-633 are biomarkers of a worse prognosis for melanoma combined with CRC cancers [37]. The miR-633 molecule is another possible *CD82* target. According to bioinformatics analyses and *in vitro* studies, miR-633 could target and regulate *CD82* expression [38]. Therefore, the inhibition of miR-633 results in decreased cell viability and migration, indicating that it may be a potential target for CD82.

*CD82*, a tumor suppressor gene, is downregulated during tumor progression in human cancers and can be activated by p53 [39]. *CD82* functions with *BCL2L12* via the *AKT*/*STAT5* signaling pathway in acute myelogenous leukemia cells [40]. *CD82* may interact with the *KRAS* gene through p53 or may be directly associated with growth factors [41]. Except for bioassay data in PubChem [12], no literature to date has directly demonstrated the *CD82-KRAS* SL pair in colon cancer. To confirm our SL prediction model and test the sensitivity of cancer cells to drug inhibition, we established isogenic mutation model cell lines, including DLD-1 *KRAS* wild-type (W/−) as a positive control and DLD-1 *KRAS* mutation (W/M) siRNA for gene knockdown in stabilized cancer cell lines. We administered a drug to inhibit receptor tyrosine kinase signaling in *KRAS*-mutated cells and knocked down essential genes to identify the proteins involved in downstream signaling pathways using western blotting. Finally, based on cell viability, we demonstrated *CD82-KRAS* SL in DLD-1 colon cancer cells.

This study aims to identify SL pairs through different angles to prioritize the SL candidates. To demonstrate the co-dependency of *CD82* and *KRAS* mutation in colorectal cancer cells, we used isogenic DLD-1 cancer cell lines, including DLD-1 W/M and DLD-1 W/−, in our study. Fig. 5A compares baseline *CD82* expression in scrambled siRNA-treated DLD-1 W/M and DLD-1 W/− cancer cells. When *KRAS* mutations are knocked out in DLD-1 W/− isogenic cancer cells, *CD82* expression and p-MEK decrease compared to *KRAS*-mutant colorectal cancer cells DLD-1 W/M. Using DLD-1 isogenic cell lines, we demonstrated the co-dependency between *CD82* and *KRAS* mutations. However, our study has several limitations. Firstly, we did not use more cancer cell lines as positive or negative controls to validate our findings. Through RNAi screening of different cell lines, Barbie et al. demonstrate that *TBK1* is required for oncogenic *KRAS*-driven cancers [42]. We used the *RAS*-mutant HTS database with a genome-wide RNA interference screen in DLD-1 W/M to validate the SL gene pairs. Secondly, the performance of our text-mining method may be improved using other methods or data resources. Thirdly, some SL pairs are presented comprehensively, such as the *BCL2L1-KRAS*-mutant SL pair in colorectal cancer [2]. However, we briefly introduced text-mining into the literature rather than comprehensively; therefore, our study may not have identified additional SL pairs. Lastly, in most cases, automatically generated trigger terms were shown to yield higher recalls but lower precisions.

## 5. Conclusions

This study developed an SL prediction system based on a text-mining method. We identified and validated *CD82-KRAS* SL using cell-based experiments. Our findings highlight *CD82* as a novel target for *KRAS* mutations in colon cancer and demonstrated digitonin as a potential therapeutic agent via the bioassay database for *KRAS*-mutant colon cancer cell lines. Accordingly, further studies using digitonin are warranted.

## Declarations

**Consent for publication:** All authors agree to publication.

**Availability of data and material:** The data generated in this study are available in Supplementary Table S1-S7. Publicly available data used in the study are listed in the key resources table. The code used for the analysis in this paper is available online at https://github.com/imwilly3737/SL_prediction_by_literature. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**Funding:** This work was supported in part by the Ministry of Science and Technology (MOST), Taiwan, under Research Grant of MOST 111-2634-F-006-002, MOST 111-2634-F-006-007, and MOST 110-2634-F-006-020, Ministry of Health and Welfare (MOHW111-TDU-B-221-014005), and National Cheng Kung University Hospital (NCKUH-11102061).

**Authors' contributions:** Conception and study design: HT Yang, MY Chien, JH Chiang, and PC Lin; Development of methodology: HT Yang and MY Chien; Acquisition of data: HT Yang, MY Chien, and PC Lin; Statistical and computational analysis: HT Yang, MY Chien, and PC Lin; Writing, review, and/or revision of the manuscript: HT Yang, MY Chien, and PC Lin; Study supervision: HT Yang and PC Lin.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The authors gratefully acknowledge the significant contribution of Sina Abdollahi for his helpful suggestions and technical support.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.09.025.

## References

[1] Hong DS, Fakih MG, Strickler JH, et al. KRASG12C inhibition with sotorasib in advanced solid tumors. N Engl J Med 2020;383(13):1207–17. https://doi.org/10.1056/NEJMoa1917239.
[2] Aguirre AJ, Hahn WC. Synthetic lethal vulnerabilities in KRAS-mutant cancers. Cold Spring Harb Perspect Med 2018;8(8):a031518. https://doi.org/10.1101/cshperspect.a031518.
[3] Dobzhabsky T. Genetics of natural populations; recombination and variability in populations of Drosophila pseudoobscura. Genetics 1946;31(3):269–90. https://doi.org/10.1093/genetics/31.3.269.
[4] Jonsson P, Bandlamudi C, Cheng ML, et al. Tumour lineage shapes BRCA-mediated phenotypes. Nature 2019;571(7766):576–9. https://doi.org/10.1038/s41586-019-1382-1.
[5] Helleday T. The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. Mol Oncol 2011;5(4):387–93. https://doi.org/10.1016/j.molonc.2011.07.001.
[6] Fong PC, Boss DS, Yap TA, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med 2009;361(2):123–34. https://doi.org/10.1056/NEJMoa0900212.
[7] Li K, Luo H, Huang L, Luo H, Zhu X. Microsatellite instability: a review of what the oncologist should know. Cancer Cell Int 2020;13(20):16. https://doi.org/10.1186/s12935-019-1091-8.
[8] Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a cancer dependency map. Cell 2017;170(3):564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.
[9] Chan EM, Shibue T, McFarland JM, et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. Nature 2019;568(7753):551–6. https://doi.org/10.1038/s41586-019-1102-x.
[10] Nijman SM. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS Lett 2011;585(1):1–6. https://doi.org/10.1016/j.febslet.2010.11.024.
[11] Guo J, Liu H, Zheng J. SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. Nucleic Acids Res 2016; 44 (D1):D1011-D1017. doi:10.1093/nar/gkv1108.
[12] Charrin S, Manié S, Thiele C, et al. A physical and functional link between cholesterol and tetraspanins. Eur J Immunol 2003;33(9):2479–89. https://doi.org/10.1002/eji.200323884.
[13] Wang Y, Xiao J, Suzek TO, et al. PubChem's BioAssay database. Nucleic Acids Res 2012; 40(Database issue):D400-D412. doi:10.1093/nar/gkr1132.
[14] Yang HT, Ju JH, Wong YT, Shmulevich I, Chiang JH. Literature-based discovery of new candidates for drug repurposing. Brief Bioinform 2017;18(3):488–97. https://doi.org/10.1093/bib/bbw030. PMID: 27113728.
[15] Cer D, de Marneffe MC, Jurafsky D, Manning C. Parsing to Stanford dependencies: Trade- offs between speed and accuracy. In LREC 2010.
[16] Koh JL, Brown KR, Sayad A, Kasimer D, Ketela T, Moffat J. COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. Nucleic Acids Res 2012; 40(Database issue):D957-D963. doi:10.1093/nar/gkr959.
[17] Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015; 43 (Database issue):D805-D811. doi:10.1093/nar/gku1075.
[18] Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483 (7391):603–7.
[19] Kaewphan S, Van Landeghem S, Ohta T, Van de Peer Y, Ginter F, Pyysalo S. Cell line name recognition in support of the identification of synthetic lethality in cancer from text. Bioinformatics 2016;32(2):276–82. https://doi.org/10.1093/bioinformatics/btv570.
[20] Blank JL, Liu XJ, Cosmopoulos K, et al. Novel DNA damage checkpoints mediating cell death induced by the NEDD8-activating enzyme inhibitor MLN4924. Cancer Res 2013;73(1):225–34. https://doi.org/10.1158/0008-5472.CAN-12-1729.
[21] Luo J, Emanuele MJ, Li D, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 2009;137(5):835–48. https://doi.org/10.1016/j.cell.2009.05.006.
[22] Haagensen EJ, Thomas HD, Mudd C, et al. Pre-clinical use of isogenic cell lines and tumours in vitro and in vivo for predictive biomarker discovery; impact of KRAS and PI3KCA mutation status on MEK inhibitor activity is model dependent. Eur J Cancer 2016;56:69–76. https://doi.org/10.1016/j.ejca.2015.12.012.
[23] Felth J, Rickardson L, Rosén J, et al. Cytotoxic effects of cardiac glycosides in colon cancer cells, alone and in combination with standard chemotherapeutic drugs. J Nat Prod 2009;72(11):1969–74. https://doi.org/10.1021/np900210m.
[24] Floren M, Restrepo Cruz S, Termini CM, Marjon KD, Lidke KA, Gillette JM. Tetraspanin CD82 drives acute myeloid leukemia chemoresistance by modulating protein kinase C alpha and β1 integrin activation. Oncogene 2020;39(19):3910–25. https://doi.org/10.1038/s41388-020-1261-0.
[25] Costanzo M, Baryshnikova A, Bellay J, et al. The genetic landscape of a cell. Science 2010;327(5964):425–31. https://doi.org/10.1126/science.1180823.
[26] Bassik MC, Kampmann M, Lebbink RJ, et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell 2013;152(4):909–22. https://doi.org/10.1016/j.cell.2013.01.030.
[27] Laufer C, Fischer B, Billmann M, Huber W, Boutros M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. Nat Methods 2013;10(5):427–31. https://doi.org/10.1038/nmeth.2436.
[28] Sinha S, Thomas D, Chan S, et al. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. Nat Commun 2017;8:15580. https://doi.org/10.1038/ncomms15580.
[29] Das S, Deng X, Camphausen K, Shankavaram U. DiscoverSL: an R package for multi-omic data driven prediction of synthetic lethality in cancers. Bioinformatics 2019;35(4):701–2. https://doi.org/10.1093/bioinformatics/bty673.
[30] Jerby-Arnon L, Pfetzer N, Waldman YY, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell 2014;158 (5):1199–209. https://doi.org/10.1016/j.cell.2014.07.027.
[31] Ryan CJ, Lord CJ, Ashworth A. DAISY: picking synthetic lethals from cancer genomes. Cancer Cell 2014;26(3):306–8.
[32] Sanjiv K, Hagenkort A, Calderón-Montaño JM, et al. Cancer-specific synthetic lethality between ATR and CHK1 kinase activities. Cell Rep 2016;14 (2):298–309. https://doi.org/10.1016/j.celrep.2015.12.032.
[33] Skoulidis F, Li BT, Dy GK, et al. Sotorasib for lung cancers with KRAS p. G12C mutation. N Engl J Med 2021;384(25):2371–81. https://doi.org/10.1056/NEJMoa2103695.
[34] Fakih MG, Kopetz S, Kuboki Y, et al. Sotorasib for previously treated colorectal cancers with KRASG12C mutation (CodeBreaK100): a prespecified analysis of a single-arm, phase 2 trial. Lancet Oncol 2022;23(1):115–24. https://doi.org/10.1016/S1470-2045(21)00605-7.
[35] Plaut D. Cardiac glycosides: digitalis, digoxin, and digitoxin. Am J Med Technol 1978;44(9):896–904.
[36] Pitha J, Szente L. Digitonin derivatives of low toxicity: potential solubilizers for lipophilic compounds. J Pharm Sci 1984;73(2):240–3. https://doi.org/10.1002/jps.2600730224.
[37] Du X, Wang B, Liu L, Li Y, Wang ZX, Zhang GJ, Yang XF. Lower expression of KAI1 as a biomarker of poor survival prognosis of melanoma combined with colorectal cancer metastasis. J Int Med Res 2022; 50(8):3000605221116758. doi: 10.1177/03000605221116758.
[38] Wang Z, Liu Y. MicroRNA-633 enhances melanoma cell proliferation and migration by suppressing KAI1. Oncol Lett 2021 Feb;21(2):88. https://doi.org/10.3892/ol.2020.12349.
[39] Guo C, Liu QG, Zhang L, Song T, Yang X. Expression and clinical significance of p53, JunB and KAI1/CD82 in human hepatocellular carcinoma. Hepatobiliary Pancreat Dis Int 2009;8(4):389–96.
[40] Nishioka C, Ikezoe T, Takeuchi A, Nobumoto A, Tsuda M, Yokoyama A. The novel function of CD82 and its impact on BCL2L12 via AKT/STAT5 signal pathway in acute myelogenous leukemia cells. Leukemia 2015;29 (12):2296–306. https://doi.org/10.1038/leu.2015.219.
[41] Barzegar F, Zaefizadeh M, Yarir R, Salehzadeh A. Balancing The Expression of K-ras and CD82 Genes By Magnetic Nano-oleuropein as a New Mechanism for Inhibition of AGS Gastric Cancer Cells. Journal of epigenetics 2020; 2(1):17-26. Available: https://www.sid.ir/en/journal/ViewPaper.aspx?id=817192..
[42] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462(7269):108–12. https://doi.org/10.1038/nature08460.