

Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates

Fan Zhang^{*,§}, Min Wu^{†,¶}, Xue-Juan Li^{*,||}, Xiao-Li Li^{†,***}, Chee Keong Kwoh^{*,††}
and Jie Zheng^{*,‡,‡‡}

^{*}*School of Computer Engineering, Nanyang Technological University
Singapore 639798*

[†]*Data Analytic Department, Institute for Infocomm Research
A*STAR, Singapore 138632*

[‡]*Genome Institute of Singapore, A*STAR, Singapore 138672*

[§]*fzhang005@e.ntu.edu.sg*

[¶]*wumin@i2r.a-star.edu.sg*

^{||}*lixj@ntu.edu.sg*

^{***}*xlli@i2r.a-star.edu.sg*

^{††}*asckkwh@ntu.edu.sg*

^{‡‡}*zhengjie@ntu.edu.sg*

Received 1 September 2014

Revised 17 November 2014

Accepted 24 December 2014

Published 10 February 2015

A major goal of personalized anti-cancer therapy is to increase the drug effects while reducing the side effects as much as possible. A novel therapeutic strategy called synthetic lethality (SL) provides a great opportunity to achieve this goal. SL arises if mutations of both genes lead to cell death while mutation of either single gene does not. Hence, the SL partner of a gene mutated only in cancer cells could be a promising drug target, and the identification of SL pairs of genes is of great significance in pharmaceutical industry. In this paper, we propose a hybridized method to predict SL pairs of genes. We combine a data-driven model with knowledge of signalling pathways to simulate the influence of single gene knock-down and double genes knock-down to cell death. A pair of genes is considered as an SL candidate when double knock-down increases the probability of cell death significantly, but single knock-down does not. The single gene knock-down is confirmed according to the human essential genes database. Our validation against literatures shows that the predicted SL candidates agree well with wet-lab experiments. A few novel reliable SL candidates are also predicted by our model.

Keywords: Synthetic lethality; signaling pathways; data-driven.

1. Introduction

A major challenge to establish treatments for human cancer is how to kill cancer cells specifically, but spare normal cells. Among the developed anti-cancer

^{‡‡}Corresponding author.

therapies, the lack of selectivity of the drugs is likely to lead to the elimination of both the tumour and the healthy cells, thus causing various side effects. To improve the classic cytotoxic therapies, a novel anti-cancer strategy named synthetic lethality (SL)¹ has been introduced and has shown great potential to increase the selectivity of the drugs. Two genes have SL relationship when the combination of mutations is able to kill a cell, whereas the mutation by one gene cannot. Since a lot more genetic mutations exist in tumour cells than in normal cells, the identification of the SL pair of genes could play an important role in pharmaceutical industry when one of the genes is a cancer-specific mutated gene. The drug that targets the partner gene is thus able to give rise to SL and kill the tumour cells selectively.

However, the majority of the established SL identification techniques, such as RNA interference (RNAi),² are faced with multiple issues like the design of an effective small interfering RNA (siRNA) sequence and the stability across different platforms or cancer subtypes. Moreover, while the underlying mechanisms that lead to SL are crucial for developing reliable anti-cancer therapies, they can hardly be reflected by the screening-based methods. Wu *et al.*³ proposed a meta-analysis data-mining method that was able to predict SL based on genomic and proteomic features, Güell *et al.*⁴ also proposed a computational method to screen synthetic lethal reaction pairs from the perspective of metabolism, however, their results mainly focused on yeast. Another computational method was developed by Heiskanen and Aittokallio,⁵ which infers human SL from yeast SL based on the conserved features between the two species. However, key features that are specific to human were likely to be lost.

In this paper, we propose a computational method to predict SL pairs of genes by combining a data-driven model^{6,7} with knowledge of the underlying mechanisms (e.g. signaling pathways information). The data-driven model is used to construct a function that describes the relationship between the activities of signaling proteins and the probability of cell death. Then the single gene knock-down and the double gene knock-down are simulated by taking into account the pathway information as well (e.g. what downstream proteins should be affected by the knock-down events). A predicted probability of cell death is calculated after gene knock-down. A pair of genes is considered as an SL candidate when the predicted probability of cell death after double genes knock-down is highly increased compared with real data, while the predictions after the knock-down of either single gene are not. We use human essential genes to validate the simulation of single gene knock-down. The Syn-Lethality database,⁸ which consists of 113 SL pairs manually collected from literatures of wet-lab experiments, is employed to evaluate the SL predictions of our model. The two SL pairs of genes that coexist in the Syn-Lethality database and in the employed data set are successfully identified. Explanations of the mechanisms of the novel predicted SL pairs of genes are also presented.

2. Methods

2.1. Data

We downloaded the time-series signaling data for proteins from the work of Lee *et al.*⁹ This signaling data set contains the phosphorylation levels of 32 signaling proteins at 5 time points, as well as the cell fates (e.g., apoptosis, proliferation) measured using flow cytometry at the corresponding time points. In addition, the signaling data contains six treatments (i.e. six groups), which were designed on breast cancer cells (e.g. cell line BT20). We chose the control group (treated with DMSO) data since the other five groups had been treated by different drugs which may introduce unpredictable biases to our model. For each signalling protein, the values of its phosphorylation levels from all the biological experiments were normalized into an open interval between 0 and 1. Suppose max and min represent the maximum and minimum measurements of a protein, respectively, then for every biological measurement (denoted as mea) of this protein, the normalized value is $(\text{mea} - \text{min} * 0.99) / (\text{max} * 1.01 - \text{min} * 0.99)$. Finally, we approximately estimated the probability of cell death as the proportion of dead cells under each treatment.

2.2. Pathways

Pathway information was extracted from GeneGO database.¹⁰ Classic signaling pathways about the regulation of cell fates (e.g. apoptosis and proliferation) were extracted and combined into a network as shown in Fig. 1. Here, we constructed a generic pathway where only the proteins with phosphorylation measurements available in the data set and their direct interaction neighbors were retained. In this generic pathway, there are 59 signaling proteins and three cell fates (i.e. apoptosis, proliferation, cell cycle). And among the 59 proteins, 28 had measurements in the data set we downloaded, i.e. SMAC, 4EBP1, Beclin1, H2AX, ERK, S6, S6K, Cabl, Casp6, Casp9, CDC25, Chk1, p27, PUMA, Wee1, HSP27, AKT, JNK, p38, BIM, BID, DAPK1, CYCLIND1, STAT3, p53, Casp8, EGFR and HER2 (green nodes in Fig. 1).

2.3. Data-driven model

We first constructed a mathematical model in Eq. (1)⁷ to relate the cell signals to the cell fates. At each time point, P is the probability of cell death, x_i (where $i = 1, \dots, 28$) represents the activity (e.g. phosphorylation level) of the i th signaling protein, α_i indicates the influence of the corresponding protein to the cell death (as such, it is not propagation through the network, but rather direct regression to the target node, i.e. apoptosis) and ε is a small constant representing random error.

$$P = e^{\alpha_0} \cdot \prod_{i=1}^n x_i^{\alpha_i} + \varepsilon. \quad (1)$$

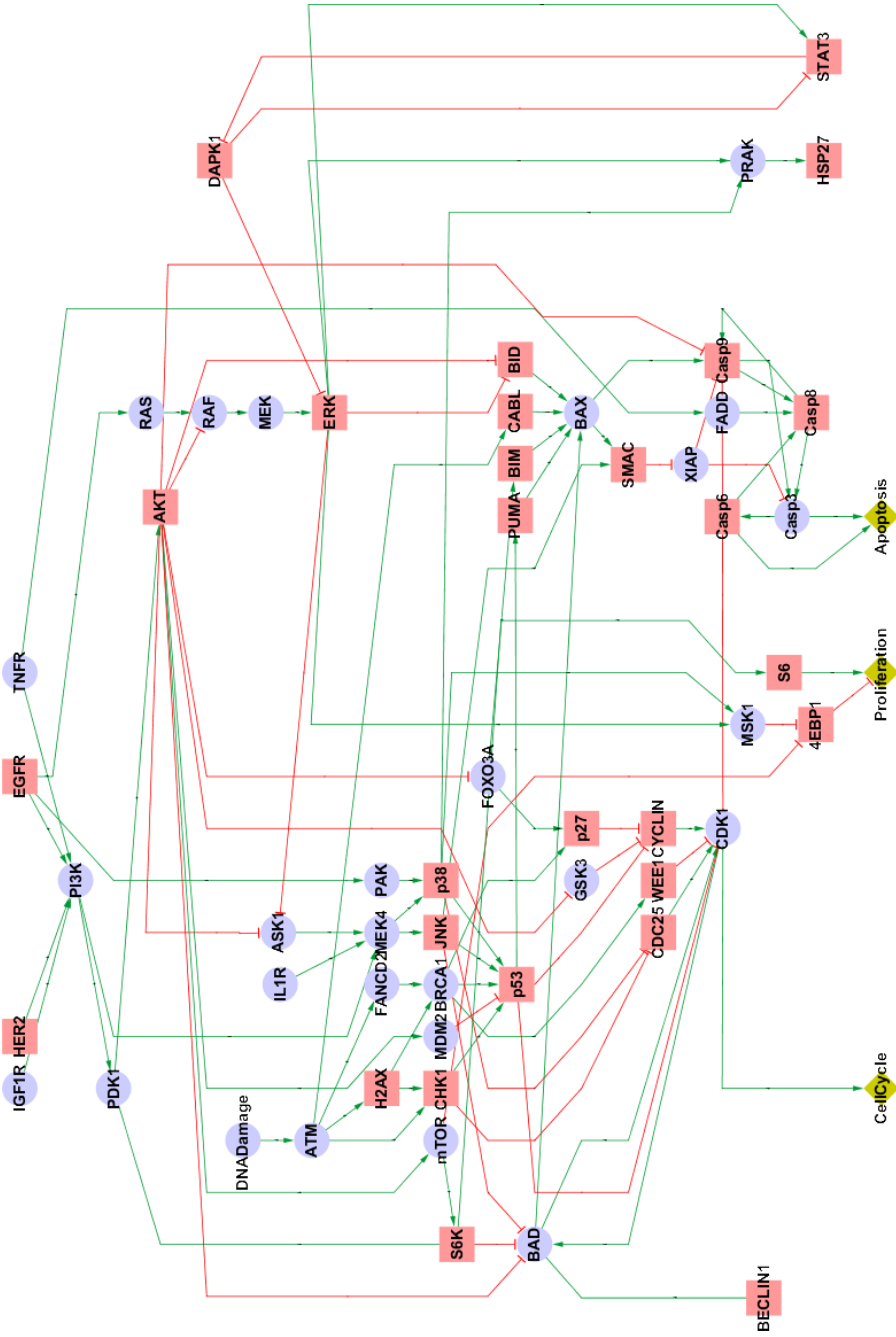


Fig. 1. Pathway information extracted from GeneGO MetaCore database. Signaling proteins with and without experimental measurements are represented as pink rectangular and blue circle, respectively. Cell fates are represented as yellow diamond. Activation and inhibition interactions are denoted as green and red edges, respectively.

Next, we built the functions in Eq. (2)⁷ to describe the relationship between a signalling protein and all its measured upstream nodes in Fig. 1. For each signalling protein (e.g. the i th protein), we selected it as the dependent variable in Eq. (2) and all its measured upstream nodes as independent variables, with λ_{ij} representing the contribution of the j th measured upstream node to the node i and all the λ parameters forming a matrix M . To identify the upstream nodes of the node i , we considered not only its direct parent nodes, but also the measured proteins that had at least one path leading to the node i . For example, from Fig. 1, we can see that SMAC has two direct parent nodes (JNK and BAX, but only JNK has measurements), and since there is a path BIM-BAX-SMAC and BIM has measurements, both JNK and BIM will be considered as the measured upstream nodes of SMAC.

$$x_i = e^{\lambda_0} \cdot \prod_j x_j^{\lambda_{ij}} + \varepsilon_i. \quad (2)$$

Then the parameters in Eqs. (1) and (2) (e.g. α_i and λ_{ij}) were learnt using partial least squares regression (PLSR) method as in our previous work⁷ based on the data set (Algorithms 1 and 2).

2.4. Gene knock-down simulation

Suppose that the signalling protein u is a measured upstream node of protein v in our generic pathway, and u is knocked down (e.g. mutated or blocked). The signals received by v are thus changed under the assumption that all the other upstream nodes of v remain unchanged. For example, v will be downregulated or upregulated after the removal of u when u activates or inhibits v , respectively. Assuming that u is knocked down, we next estimate the changes of its measured downstream nodes and the corresponding probability of cell death, based on our data-driven models in Eqs. (1) and (2).

Algorithm 1. Learning parameters of Eq. (1).

INPUT: Time-series signalling protein data $X_{(t \times n)}$ and cell death data $P_{t \times 1}$, where t is the number of measurements and $n = 28$.

OUTPUT: The vector $A : (\alpha_0, \alpha_1, \dots, \alpha_n)$ with $n + 1$ elements.

//Take the logarithm of Equation (1).

$\ln(P - \varepsilon) = \alpha_0 + \sum_{i=1}^n \alpha_i \cdot \ln(x_i)$;

Independent variables \leftarrow signalling proteins;

Dependent variables \leftarrow cell death;

Use partial least squares regression (PLSR) method (e.g., the *plsregress* function in MATLAB) to do the regression and learn the parameters $(\alpha_0, \alpha_1, \dots, \alpha_n)$;

Return The vector A .

Algorithm 2. Learning parameters of Eq. (2).

INPUT: Time-series signalling protein data $X_{(t \times n)}$; adjacency matrix $AD_{(N \times N)}$, where $N = 59$ is the number of nodes in the pathways and $N_{ij} = 1$ if there is an edge leading from node N_j to N_i (i.e., N_j is a direct parent node of N_i).

OUTPUT: The matrix $M_{(n \times n)}$ where each element m_{ij} is the parameter λ_{ij} .

//Extract measured upstream nodes.

for each signalling protein i **do**

 //Extract direct parent nodes of protein i .

$rowAD(i) \leftarrow i$ -th row of matrix AD ;

for each parent node j ($j \neq i$) in $rowAD(i)$ **do**

if the j -th parent node has measurements and $m_{ij} = 0$ **do**

$m_{ij} \leftarrow 1$;

elseif the j -th parent node has no measurements and $m_{ij} = 0$ **do**

 //Extract direct parent nodes of protein j .

$rowAD(j) \leftarrow j$ -th row of matrix AD ;

$rowAD(i) \leftarrow \text{union}(rowAD(i), rowAD(j))$;

end for when the i -th row of M is conserved or reach the maximum steps of loops;

end for

//Calculate λ_{ij} .

for each signalling protein i **do**

 //Extract measured upstream nodes of protein i .

$rowM(i) \leftarrow i$ -th row of matrix M ;

 //Take the logarithm of Equation (2)

$\ln(x_i - \varepsilon_i) = \lambda_{i0} + \sum_j \lambda_{ij} \cdot \ln(x_j)$;

 Independent variables \leftarrow proteins with nonzero values in $rowM(i)$;

 Dependent variables \leftarrow protein i ;

 Use PLSR to learn the parameters λ_{ij} as in Algorithm 1;

$m_{ij} \leftarrow \lambda_{ij}$;

end for

Return The matrix M .

For the knock-down of a single gene (gene and protein are used interchangeably in this paper), we removed not only its contribution to the cell death, but also its influence on all its downstream nodes. For example, the j th node in the pathway was knocked down, and the effect of its knock-down on cell death was estimated as follows. First, we removed the influence of the j th protein to all its measured downstream nodes by setting the j th column in Matrix M to be 0 ($\lambda_{ij} = 0$, for all i) and recalculated the activity levels of its downstream nodes. Second, we set $\alpha_j = 0$ to remove the contribution of the j th protein to cell death. Then we estimated the probability of cell death by substituting the updated protein data and vector $(\alpha_0, \alpha_1, \dots, \alpha_n)$ into Eq. (1). Fourth, we employed the Kullback–Leibler divergence (the relative entropy, given in Eq. (3)) to estimate how much the cell death predictions differ from biological experimental measurements, where t is the number of predictions (or measurements). In information theory, the Kullback–Leibler divergence is a measure of the difference between two probability distributions. Algorithm 3 gives the pseudocode of the process. We defined the divergence as the “single knock-down score” of the j th protein. If the single knock-down score was equal to zero, there was no increment of cell death due to knock-down event. Therefore, the bigger the single knock-down score is, the greater the cell death is promoted.

$$D_{\text{KL}}(\text{Predicted}||\text{Measured}) = \sum_t \text{Predicted}(t) \ln \left(\frac{\text{Predicted}(t)}{\text{Measured}(t)} \right). \quad (3)$$

Similarly, we can infer the effect on cell death when we knocked down a pair of genes. For each protein, we extracted the set of its downstream nodes. Given a pair of proteins to be knocked down, we first computed the union set of their downstream nodes and recalculated the activity levels for the nodes in this union set by modifying the parameters in Eq. (2) to be 0. Then, we estimated the probability of cell death by feeding the updated activity levels and vector $(\alpha_0, \alpha_1, \dots, \alpha_n)$ to Eq. (1), and similarly used Eq. (3) to define the “double knock-down score” to measure the effect of the knock-down of a pair of genes.

2.5. Prediction of SL

For a protein pair (u, v) , let $d(u, v)$ denotes its double knock-down score. $s(u)$ and $s(v)$ are the single knock-down scores for u and v , respectively. Our assumption was that a protein pair (u, v) with large $d(u, v)$ and low $s(u)$ and $s(v)$ tended to be a SL pair. Hence, we defined the SL score for the protein pair (u, v) as in Eq. (4), to quantify their mutual dependence on knock-down events. For example, if u and v are independent, then knocking down u does not give any influence on knock-down of v and vice versa, so their SL score is zero. The log-ratio term $(\ln(\frac{d(u,v)}{s(u) \times s(v)}))$ helps to identify the pairs whose double knock-down score increases significantly compared with single knock-down score. However, large log-ratio is insufficient to clarify the discovery of an SL candidate since an SL candidate should have a large double knock-down score in the first place. Therefore, by timing the double knock-down

Algorithm 3. Single gene knock-down simulation.

INPUT: Time-series signalling protein data $X_{(t \times n)}$ and cell death data $P_{t \times 1}$; the vector A and the matrix M .

OUTPUT: Single knock-down score for each gene.

for each signalling protein j **do**

$X' \leftarrow X$;

$M' \leftarrow M$;

//Extract measured downstream nodes of protein j .

$colM(j) \leftarrow$ the j -th column of matrix M ;

for each downstream node i in $colM(j)$ **do**

$m'_{ij} \leftarrow 0$;

//Extract measured upstream nodes of protein i .

$rowM'(i) \leftarrow$ the i -th row of the matrix M' ;

Recalculate the activity levels of protein i by substituting the matrix X and $rowM'(i)$ into Equation (2);

Update the i -th column of X' using the recalculate activity levels of protein i ;

end for

$A' \leftarrow A$;

$\alpha'_j \leftarrow 0$;

Recalculate the probability of cell death by substituting the matrix X' and the vector A' into Equation (1);

Use Equation (3) to calculate the Kullback Leibler divergence between the predicted and measured cell death as the knock-down score of protein j ;

end for

Return List of single knock-down scores for all genes.

$score(d(u, v))$, the pairs with large log-ratio but small double knock-down score (such that the SL score is small) are filtered out to reduce the false positive. The larger the SL score is, the more likely the pair is considered as SL candidate in this paper.

$$SL(u, v) = d(u, v) \ln \left(\frac{d(u, v)}{s(u) \times s(v)} \right). \quad (4)$$

3. Results

3.1. Single knock-down and essential genes prediction

To evaluate the performance of single gene knock-down simulation, 28 measured signaling proteins were ranked according to their single knock-down scores. Out of these 28 proteins, 15 proteins had single knock-down scores larger than 0. Table 1 shows the top-five proteins (i.e. AKT, p53, CHK1, S6K and CYCLIND1), whose removal brings the highest impact on the probability of cell death.

AKT, which is also known as Protein kinase B (PKB), plays an important role in many regulation processes such as proliferation, apoptosis, cell cycle and metabolism. From Fig. 1, we can see that the activation of mTOR by AKT contributes to cell proliferation. AKT regulates also cell apoptosis via several pathways such as AKT-Caps9, and controls cell cycle by phosphorylating its substrates including GSK3.¹¹ The tumor suppressor p53 is able to induce cell cycle arrest, initiate DNA repair process when DNA is damaged and activate cell apoptosis if DNA is irreparable. Therefore, it is crucial in multiple cellular mechanisms.¹²⁻¹⁴ CHK1, which is also referred to as checkpoint kinase 1, responds to checkpoint-mediated cell cycle arrest and DNA damage.¹⁵ S6K (Ribosomal protein S6 kinase beta-1), which responds to mTOR and growth factors, is responsible for regulating protein synthesis, cell growth and cell proliferation.¹⁶ CYCLIND1, acting as the regulator of cyclin-dependent kinase (CDK), belongs to a family whose members have a significant periodicity in their abundance over cell cycle.¹⁷ Moreover, all these five genes were identified as human essential genes in Online GEne Essentiality database (OGEE).¹⁸

3.2. Double knock-down and SL prediction

We performed double knock-down simulation over all the 378 protein pairs (i.e. $\binom{28}{2}$ as we had 28 measured signaling proteins). 252 of them had double knock-down scores larger than 0, indicating that the double knock-down event tended to increase the probability of cell death.

SL scores were calculated based on Eq. (4). Table 2 gives the top 20 pairs of SL candidates after ranking based on the SL scores. A recent work has investigated the

Table 1. Rank of the single knock-down simulation. For example, AKT was in the top of the list since knock-down of AKT gave rise to the biggest growth of the probability of cell death according to Eq. (3).

Protein	Gene	Single knock-down score	Rank
AKT	ENSG00000142208	0.037	1
p53	ENSG00000141510	0.031	2
CHK1	ENSG00000149554	0.021	3
S6K	ENSG00000108443	0.016	4
CYCLIND1	ENSG00000110092	0.015	5

Table 2. Rank of the SL candidates. For example, the pair of AKT and BID was in the top of the list since it had the highest SL score based on Eq. (4).

Protein pairs		SL score	Rank
AKT	BID	2.053	1
AKT	Cabl	2.029	2
AKT	BIM	2.021	3
AKT	CYCLIN	1.983	4
AKT	Casp9	1.978	5
p53	BID	1.976	6
p53	CHK1	1.971	7
AKT	Wee1	1.969	8
p53	BIM	1.954	9
p53	Cabl	1.933	10
p53	Wee1	1.914	11
p53	Casp8	1.907	12
AKT	Casp6	1.902	13
AKT	Casp8	1.899	14
p53	S6K	1.897	15
p53	Casp9	1.889	16
p53	CYCLIN	1.874	17
AKT	S6K	1.860	18
p53	Casp6	1.841	19
CHK1	CYCLIN	1.807	20

SL interactions in the signalling pathways related to DNA damage and checkpoint control.¹⁹ Two pairs, (*p53*, *CHK1*), as well as (*p53*, *Wee1*) were reported as SL pairs in their work, and these two pairs are ranked the seventh and the eleventh by our method. As reported by Kaelin² and Le Meur *et al.*,²⁰ three types of mechanisms were proposed, from the perspectives of signalling pathways, that could be the candidate explanations of SL interactions in human cancers. First, two proteins that coexist in a linear essential pathway are likely to have SL interaction, such as (*AKT*, *BID*) in the cascade AKT-BID-BAX-Casp9 and (*AKT*, *BIM*) in the pathway AKT-FOXO3A-BIM-BAX-Casp9 (Fig. 1). The mutation of the first protein (e.g. AKT) would decrease the signals transmitted in the pathway whereas the mutation of both destroys the pathway. Second, two proteins from two parallel essential pathways could be SL partners when they are backups of each other. For example, (*p53*, *Cabl*) is predicted since p53 and Cabl are in two parallel essential pathways, i.e. ATM-CHK1-p53-PUMA-BAX and ATM-Cabl-BAX, respectively. Another example is (*p53*, *Wee1*), where p53 is in the pathway BRCA1-p53-CDK1 and Wee1 is in the pathway BRCA1-Wee1-CDK1. Third, two components have the same essential function or they both contribute to the construction of an essential protein complexes. Moreover, according to the Syn-Lethality database,⁸ only two SL pairs of proteins reported in the literatures have measurements in the data set⁹ i.e. (*p53*, *CHK1*) and (*p53*, *Wee1*), and both of them have been successfully predicted by our method.

4. Discussion and Conclusion

Anti-cancer therapy based on SL strategy shows great potential for its capability of treating the cancer and the normal cells differently. Finding SL pairs of genes provides great opportunities for drug target identification if one of the genes in the pair has been mutated in the cancer cells. However, due to the high cost of the screening technology, highly reliable results generated from wet-lab experiments are growing slowly. Therefore, we proposed a computational method which hybridizes a data-driven model with knowledge of signalling pathways to predict potential SL pairs of genes. We first related the activities of signalling proteins to the probability of cell death using a mathematical function. Second, we identified the relationship between each protein and all its measured upstream nodes based on both the biological experimental data and the signalling pathways. Next we simulated the single knock-down of each measured protein and the double knock-down of all possible pairs of proteins, to estimate the significance of knock-down events to cell death. We then defined the SL score to do virtual screening for the candidates of SL pairs of genes. The single knock-down simulations were confirmed according to the human essential genes. And the double knock-down simulations gave both wet-lab confirmed results and novel predictions which are suggested to be reliable by evidence from literatures.

In spite of promising performance of our proposed method, limitations have been noticed which point to our future work. First, due to the lack of information, underlying mechanisms are insufficiently considered. For example, the interaction types (e.g. phosphorylation and transcription regulation) in Fig. 1 are unclear. Also the interaction between two proteins could be activation in one cell type while inhibition under other circumstance due to different genetic contexts. Second, some well-established network models such as Boolean network and ODE-based model should be involved to gain better performance and mechanistic understanding. And we believe that by employing these models, our method will be improved.

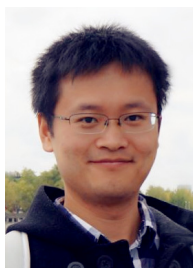
Acknowledgment

This project is supported by AcRF Tier 2 grant ARC39/13 (MOE2013-T2-1-079), Ministry of Education, Singapore.

References

1. Kaelin WG Jr, Synthetic lethality: A framework for the development of wiser cancer therapeutics, *Genome Med* 1(10):99.1–99.6, 2009.
2. Kaelin WG Jr, The concept of synthetic lethality in the context of anticancer therapy, *Nat Rev Cancer* 5(9):689–698, 2005.
3. Wu M, Li X, Zhang F, Li XL, Kwoh CK, Zheng J, Silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer, *Cancer Informatics* 13(S3):71–80, 2014.
4. Güell O, Sagués F, Ángeles Serrano M, Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis, *PLoS Comput Biol* 10(5):e1003637, 2014.

5. Heiskanen MA, Aittokallio T, Mining high-throughput screens for cancer drug targets—lessons from yeast chemical-genomic profiling and synthetic lethality, *WIREs Data Mining Knowl Discov* **2**(3):263–272, 2012.
6. Janes KA, Yaffe MB, Data-driven modelling of signal-transduction networks, *Nat Rev Mol Cell Biol* **7**(11):820–828, 2006.
7. Zhang F, Kwoh CK, Wu M, Zheng J, Data-driven prediction of cancer cell fates with a nonlinear model of signaling pathways, *ACM BCB*, 436–444, 2014.
8. Li XJ, Mishra SK, Wu M, Zhang F, Zheng J, Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies, *BioMed Res Int* **2014**:196034, 2014.
9. Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, MacBeath G, Yaffe MB, Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks, *Cell* **149**(4):780–794, 2012.
10. Ekins S, Nikolsky Y, Bugrim A, Kirillov E, Nikolskaya T, Pathway mapping tools for analysis of high content data, *Methods Mol Biol* **356**:319–350, 2007.
11. Basu A, Sivaprasad U, Protein kinase cepsilon makes the life and death decision, *Cell Signal* **19**(9):1633–1642, 2007.
12. Carvajal LA, Hamard PJ, Tonnessen C, Manfredi JJ, E2f7, a novel target, is up-regulated by p53 and mediates DNA damage-dependent transcriptional repression. *Genes Dev* **26**(14):1533–1545, 2012.
13. Liu S, Mirza A, Wang L, Generation of p53 target database via integration of microarray and global p53 DNA-binding site analysis, *Methods Mol Biol* **281**:33–54, 2004.
14. Reyes-Zurita FJ, Pachn-Pea G, Lizrraga D, Rufino-Palomares EE, Cascante M, Lupiez JA, The natural triterpene maslinic acid induces apoptosis in ht29 colon cancer cells by a jnk-p53-dependent mechanism, *BMC Cancer* **11**(154), 2011.
15. McNeely S, Beckmann R, Bence Lin AK, Chek again: Revisiting the development of chk1 inhibitors for cancer therapy, *Pharmacol Ther* **142**(1):1–10, 2014.
16. Chiang GG, Abraham RT, Phosphorylation of mammalian target of rapamycin (mTOR) at ser-2448 is mediated by p70s6 kinase, *J Biol Chem* **280**(27):25485–25490, 2005.
17. Greider C, Chattopadhyay A, Parkhurst C, Yang E, Bcl-x(l) and bcl2 delay myc-induced cell cycle entry through elevation of p27 and inhibition of g1 cyclin-dependent kinases, *Oncogene* **21**(51):7765–7775, 2002.
18. Chen WH, Minguez P, Lercher MJ, Bork P, Ogee: An online gene essentiality database, *Nucl Acids Res Database issue* **40**:D901–D906, 2012.
19. Morandell S, Yaffe MB, Exploiting synthetic lethal interactions between DNA damage signaling, checkpoint control, and p53 for targeted cancer therapy, *Prog Mol Biol Transl Sci* **110**:289–314, 2012.
20. Le Meur N, Gentleman R, Modeling synthetic lethality, *Genome Biol* **9**(9):R135, 2008.



Fan Zhang is currently a Ph.D. candidate in Medical Informatics Research Laboratory of School of Computer Engineering in Nanyang Technological University, Singapore. He received his Bachelor degree from Harbin Institute of Technology (HIT), China in 2009 and Master degree from Zhejiang University (ZJU), China in 2011, respectively. His current research interests include bioinformatics, systems biology and data mining.



Min Wu is currently a Research Scientist in the Data Analytics Department at Institute for Infocomm Research (I2R) under the Agency for Science, Technology and Research (A*STAR), Singapore. He received the B.Eng. from University of Science and Technology of China (USTC), China in 2006 and his Ph.D. degree from Nanyang Technological University, Singapore in 2011, respectively. His current research interests include machine learning, data mining and bioinformatics.



Xue-Juan Li is a Research Assistant at the Bioinformatics Research Center (BIRC) in Nanyang Technological University of Singapore. She has been engaged in the bioinformatics approach applied to understanding the mechanism of disease after she received Master degree from Shandong University of Technology of China in 2008.



Xiao-Li Li is currently a deputy department head and a senior scientist at the Institute for Infocomm Research, A*STAR, Singapore. He also holds Adjunct Associate Professor positions at National University Singapore and Nanyang Technological University. His research interests include data mining, machine learning and bioinformatics. He has been serving as the PC members/workshop chairs/session chairs in the leading conferences, such as *KDD*, *ICDM*, *SDM*, *PKDD/ECML*, *PAKDD*, *WWW*, *AAAI* and *CIKM*. He has published more than 100 research papers in the impact international journals and conferences. In 2005, he received Best Paper Award in the *16th International Conference on Genome Informatics (GIW 2005)*. In 2008, he received the Best Poster Award in the *12th Annual International Conference Research in Computational Molecular Biology (RECOMB 2008)*. In 2011, he received Best Paper Runner-Up Award of the *16th International Conference on Database Systems for Advanced Applications (DASFAA 2011)*. He also won two Best Performance Awards in international benchmarking competitions, i.e. bioinformatics competition DREAM 2007 USA, EU activity recognition challenge 2011.



Chee Keong Kwoh is currently in the School of Computer Engineering since 1993. He is the Assistant Chair of Graduate Studies; School Appraisal Committee for Services and the Programme Director, M.Sc. (Bioinformatics). He received his Bachelor degree in Electrical Engineering (1st Class) and Master in Industrial System Engineering from the National University of Singapore in 1987 and 1991, respectively. He received his Ph.D. degrees from the Imperial College of Science, Technology and Medicine, University of London in 1995. His research interests include data analytics and mining, soft computing and graph-based inference; applications areas include bioinformatics, health informatics and biomedical engineering. He has done significant research work in his research areas and published many quality international conferences and journal papers. He is the Editorial Board Member and Associate Editor of *The International Journal of Data Mining and Bioinformatics*, *IEEE Access*, *The Scientific World JOURNAL*, *Network Modeling and Analysis in Health Informatics and Bioinformatics (NetMAHIB)*, *Theoretical Biology Insights*, and *Bioinformation*. He has been Guest Editor for many journals such as *JMMB*, *International Journal on Biomedical and Pharmaceutical Engineering* and others. He has been often invited as a organizing member or referee and reviewer for a number of premier conferences and journals, including *GIW*, *IEEE BIBM*, *RECOMB*, *PRIB*, *BIBM*, *ICDM* and *iCBBE* just to name a few. Dr. Kwoh is a member of The Institution of Engineers Singapore, Association for Medical and Bioinformatics, Imperial College Alumni Association of Singapore (ICAAS). He has provided many service to professional bodies and was conferred the Public Service Medal by the President of Singapore in 2008.



Jie Zheng is a tenure-track Assistant Professor at School of Computer Engineering, Nanyang Technological University (NTU), Singapore, and an adjunct faculty member (Senior Research Scientist) with Genome Institute of Singapore (GIS), A*STAR. He received Ph.D. in 2006 from the University of California, Riverside, USA and B.Eng. (first class honor) in 2000 from Zhejiang University in China, both in Computer Science. Before joining NTU in February 2011, he did postdoc training at the National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), USA. His research interests are bioinformatics, computational systems biology and genomics, aiming to develop novel algorithms and in silico models to help answer biomedical questions (such as the mechanisms of cancer). Although trained in Computer Science, Dr. Zheng maintains active and long-standing collaborations with life scientists.