Deciphering Drug Action and Escape Pathways: An Example on Nasopharyngeal Carcinoma

Difeng Dong[1], Chun-Ying Cui[2], Benjamin Mow[3], and Limsoon Wong[1]

[1]National University of Singapore, Singapore
[2]Capital Medical University, China
[3]West Clinic Excellence Cancer Center, Singapore
{dong.difeng,wongls}@comp.nus.edu.sg
ccy@ccmu.edu.cn
bmow@westexcellence.com

Abstract. We design a drug pathway identification system, which we called Drug Pathway Decipherer (Decipherer), to generate hypotheses of treatment responsive pathway. Decipherer takes in both pre- and post-treatment gene expression data, and evaluates known biologcial pathways against the data. We applied Decipherer to two gene expression datasets of human nasopharyngeal carcinoma (NPC) treated with CYC202. Results show that the identified RAS-ERK pathway and PI3K-NFκB-IAP pathway are closely associated with treatment outcome. Decipherer is implemented in Java, and it is available together with supplementary material at `http://www.comp.nus.edu.sg/~wongls/projects/drug-pathway`.

Key words: gene expression, drug pathway, NPC, CYC202

1 Introduction

Biological pathways have been incorporated into gene expression analysis to understand drug treatment response in disease population [17]. Some works focus on enrichment analysis of gene groups extracted from pathway [25, 2, 18]. Zeeberg *et. al.* [25] and Doniger *et. al.* [2] use a hypergeometric test to determine statistically over-represented pathways in a list of differentially expressed genes in treatment. Subramanian *et. al.* [18] propose the gene set enrichment analysis (GSEA), which uses a weighted Kolmogorov-Smirnov statistics to compare two sets of distributions and uses resampling to estimate false discovery rates (FDR). Other research groups identify responsive genetic networks under drug treatment [26, 6, 4]. Zien *et. al.* [26] exhaustively enumerate all possible gene combinations on a metabolic pathway, and select the most co-expressed gene group as the responsive pathway. Ideker *et. al.* [6] follow their work by extending metabolic pathway to a protein-protein interaction network, and use an annealed random method to generate candidate gene subnetworks for statistical evaluation. Guo *et. al.* [4] follow Ideker *et. al.* [6], but their evaluation is based on the co-expression between interacted genes rather than the significance of expression change of genes in the identified subnetworks. A more recent

work is called PathwayExpress [3], which is a web-based application to evaluate gene expression data in framework of KEGG pathway database [7]. However, most existing works fall short on several issues [17]: these works provide little information on the interplay between selected genes; the collection of pathways that can be used, evaluated and ranked against the observed expression data is limited; and the generated hypotheses are still too general to guide further research and treatment. In this paper, we present a drug pathway identification system, which we called Drug Pathway Decipherer (Decipherer), to generate hypotheses for drug responsive pathways. We applied Decipherer to two NPC gene expression datasets. CYC202 (Cyclacel Ltd, Dundee, United Kingdom; Seliciclib; R-roscovitine), a CDK inhibitor, is studied for its anti-tumor effect in human NPC cells *in vitro* and *in vivo*. 3 NPC cell lines and 13 NPC patients were treated with CYC202, and the expression of selected genes were measured during the process of the treatment. Both cell lines and patients in the study responded to the drug treatment differently. Our target is to understand the drug action of CYC202 in these NPC samples as well as to identify escape pathways for the drug-resistant NPC individuals.

2 System and Methods

2.1 Overview

Decipherer is a framework for statistical evaluation of known biological pathways against gene expression data. It consists of 4 partitions distributed on two biological levels. Figure 1 shows the diagram of its workflow.

2.2 Data source

Both NPC gene expression datasets contain 380 genes selected for apoptosis, cell proliferation, and cell cycle regulation. For the *in vitro* dataset, 3 cell lines, CNE1, CNE2 and HK1 were measured for their gene expression before treatment of CYC202, and 2hs, 4hs, 6hs, 12hs and 24hs after treatment, respectively. As a result, CNE1 responded poorly to the treatment; CNE2 responded in a limited way; HK1 fully responded. For the *in vivo* dataset, 12 NPC samples and 1 non-tumor sample were taken from NPC patients. Gene expression were measured before and after treatment. 7 patients were reported to have molecular response to the treatment. (See supplementary material for more details of patients and cell lines.)

Although the panel of genes studied in the NPC case only covers a limited number of pathways, we tended to include more signaling pathways to cope with other study cases. Thus, we collected 108 signaling pathways from KEGG pathway database (September 14, 2008) [7] and 49 signaling pathways from the Ingenuity Pathway database (July 12, 2008) [27]. (See supplementary material for more details of the collected pathway information.)

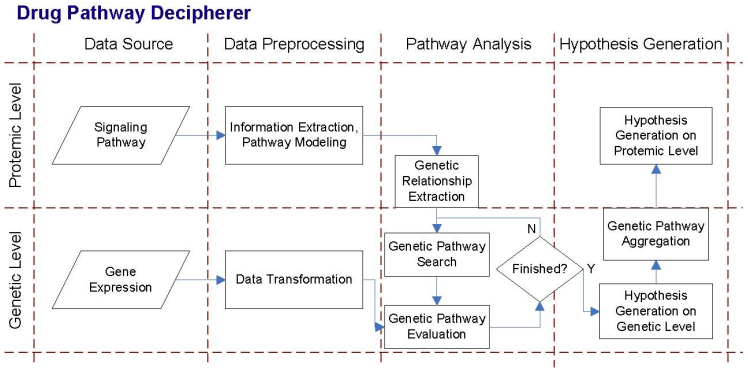

Fig. 1. The workflow of Decipherer. (a) Data source: structured signaling pathways and drug treatment gene expression data are taken as the input. (b) Data preprocessing: signaling pathways are modeled as graphs, and relative expression indicating gene expression change under drug treatment are computed. (c) Pathway analysis: pairwise genetic relationships are extracted from the modeled signaling pathways and evaluated against the relative expression data. (d) Hypothesis generation: Co-expressed genetic relationships are selected to be connected into complete genetic pathways, and statistical tools are performed to generate drug pathway hypotheses. Signaling pathway status are estimated based on the hypothesized genetic pathways.

2.3 Preprocessing data source

To capture gene expression change in response to drug treatment, the original gene expression data are transformed into relative expression (RE) values. Given pre- and post-treatment gene expression value e and e', if $e' \geqslant e$, then RE is $e'/e-1$; otherwise, RE is $1-e/e'$. Intuitively, rather than log-ratio transformation, RE describes expression change in multiples in a linear scale, which allows pairwise drug effect on gene expression data to be measured by a linear correlation metric.

Signaling pathways are modeled by directed graphs. Formally, A signaling pathway γ is a directed graph (P, I), with P the vertex set, representing the collection of proteins in pathway, and I the edge set, representing the collection of interactions between proteins. An interaction is a triplet $i = \langle p_1, p_2, s \rangle$, with $p_1, p_2 \in P$ and $s \in S$, where $S = \{\$activation, \$inhibition\}$ is the set of terms used to denote interaction type[1].

2.4 Extracting genetic relationships

Since protein activity can not be directly observed with gene expression data, we need to associate an interaction with one or more genetic relationships. For-

[1] Decipherer only considers the simplest interaction types. According to KEGG pathway database, we reduce other interaction types as: expression → activation, repression → inhibition, binding (association) → activation, dissociation → inhibition, ubiquitination → inhibition.

mally, a genetic relationship (or simply a relationship) is a triplet $q = \langle g_1, g_2, s \rangle$, with $g_1, g_2 \in G$ and and $s \in S$, where G is the whole gene collection and S is defined in Section 2.3. To extract relationships, proteins in an interaction are mapped into their encoding genes, and the interaction type is reserved in extracted relationships. As one protein can be encoded by more than one gene, multiple relationships may be extracted from one interaction. For example, $\langle$Ras,Raf,\$*activation*$\rangle$ is an interaction. Since Ras can be encoded with either HRAS or KRAS, and Raf is encoded only with RAF1, the extracted relationships are $\langle$HRAS,RAF1,\$*activation*$\rangle$ and $\langle$KRAS,RAF1,\$*activation*$\rangle$.

2.5 Scoring a genetic pathway

A genetic pathway is a chain of consecutive relationships, with each of the first gene of a relationship (except for the first relationship), equals to the second gene of its previous relationship. For example, relationships $\langle$HRAS,RAF1,\$*activation*$\rangle$, $\langle$RAF1,MAP2K1,\$*activation*$\rangle$, and $\langle$MAP2K1,MAPK1,\$*activation*$\rangle$ construct a genetic pathway. Particularly, the first gene of the first relationship, HRAS in the example, is called source gene; the second gene of the last relationship, MAPK1 in the example, is called sink gene. Theoretically, a drug can target any of the known molecules in a biological system, and drug action can be terminated in any way, for example with a phenotype, a mutaion, or a negative feedback, which means we should not impose any requirement on source and sink gene. However, this makes the number of genetic pathway candidates increases exponentially as the length of pathway increases. Thus, we require a genetic pathway starts with a membrane gene and ends with either a transcription factor, a feedback, or a phynotype regulator (encoding genes).

To score a genetic pathway ϑ, we first introduce the score function of relationship. Given a relationship $q = \langle g_1, g_2, s \rangle$, if gene expression are measured at multiple time points (as our *in vitro* dataset), the correlation of q is:

$$Corr(q) = Corr(\overrightarrow{r_{g_1}}, \overrightarrow{r_{g_2}}), \tag{1}$$

where $Corr(\overrightarrow{r_{g_1}}, \overrightarrow{r_{g_2}})$ is Pearson's correlation coefficient between RE vectors of g_1 and g_2. If gene expression are only measured at two time points (as our *in vivo* dataset), then the correlation is estimated simply by comparing the post-treatment RE of the two genes:

$$Corr(q) = sign(r_{g_1}^{post} \times r_{g_2}^{post}) \times \frac{\min_{i=1,2} |r_{g_i}^{post}|}{\max_{j=1,2} |r_{g_j}^{post}|}. \tag{2}$$

$Corr(q)$ is then transformed into a z-score, $z(q)$, against sample background. $z(q)$ are then summed up over all k relationships in ϑ into an aggregated z-score, $z(\vartheta)$, for the entire genetic pathway [6]:

$$z(\vartheta) = \frac{1}{\sqrt{k}} \sum_{q \in \vartheta} (-1)^{\alpha} z(q), \tag{3}$$

where $\alpha = 0$ if $q.type = \$activation$; $\alpha = 1$ if $q.type = \$inhibition$, which means that if the type is $\$inhibition$, the genes are expected to be affected by drug treatment in opposite directions.

Genes in ϑ are permutated for 10000 times to estimate the p-value of $z(\vartheta)$, denoted by $score(\vartheta)$. Intuitively, pathway score represents the consistency between a genetic pathway and expression change of genes in it.

2.6 Generating hypotheses

Genetic pathways satisfying statistical requirement of p-value and FDR [5] are selected as genetic hypotheses. Genetic hypotheses of same signaling pathway are then integrated to generate hypotheses of signaling pathway status. For a genetic pathway ϑ, each gene g in ϑ has an impact to the sink gene, denoted by $impact_\vartheta(g)$. If g is an inhibitor of the sink gene, then $impact_\vartheta(g) = -1$; otherwise, $impact_\vartheta(g) = 1$. In biological system, some genes play dual functions, namely to be both activator and inhibitor to same downstream gene. However, this does not conflict with the definition of impact, since in a single genetic pathway, the role of each gene is fixed.

Thus, for a signaling pathway γ, let $\vartheta \sim \gamma$ denote the hypothesized genetic pathway ϑ in γ, and G_ϑ denote the gene set in ϑ. The hypothesis of signaling pathway status Z_i^γ at time point i is a weighted average of RE of genes in the hypothesized genetic pathways of γ, which is in formula:

$$Z_i^\gamma = \frac{\sum_{\vartheta \sim \gamma} \sum_{g \in G_\vartheta} \left(\frac{1}{|G_\vartheta|} \times impact_\vartheta(g) \times r_g^i \right)}{| \vartheta \sim \gamma |}. \tag{4}$$

3 Results and Discussion

3.1 Signaling pathway database

The current signaling pathway database of Decipherer consists of three parts of data source: 49 pathways manually constructed from the Ingenuity Pathway database, 24 pathways manually constructed from KEGG pathway database, and another 84 pathways automatically constructed by invoking KEGG pathway database API. There are 748 distinct genes in the database and 181949 genetic pathways are compiled as candidates for hypothesis.

3.2 Application on the *in vitro* NPC dataset

We applied Decipherer to the *in vitro* NPC datasets with p$\leq$0.05 and FDR$\leq$0.25. Results are shown in Table 1. RAS-ERK cell proliferation pathway and PI3K-NFκB-IAP anti-apoptosis pathway are observed in all 3 cell lines, which suggests CYC202 may directly access to control cell growth and cell death. Figure 2 further compares the regulation of these two pathways among the three cell lines

Table 1. Hypothesized drug pathways by Decipherer and by PathwayExpress in three NPC cell lines.

DECIPHERER				PathwayExpress	
Source	Pathway	p-value	FDR	Source	p-value
CNE1					
ErbB signaling pathway	BTC→ErbB4→...→ERK→Elk	5.00E-4	0.035	Leukocyte transendothelial migration	3.65E-13
Apoptosis	NGF→PI3K→...→NFkB→IAP	1.00E-3	0.035	MAPK signaling pathway	1.01E-12
Regulation of actin cytoskeleton	GF→RTK→...→PI3K→PI4P5K	4.9E-3	0.11	Toll-like receptor signaling pathway	2.26E-13
MAPK signaling pathway	NF1⊣Ras→...→ERK→RSK2	6.00E-3	0.11	Regulation of actin cytoskeleton	3.40E-12
Focal adhesion	GF→RTK→...→Rac→Actin	9.70E-3	0.12	ErbB signaling pathway	1.10E-14
CNE2					
MAPK signaling pathway	NF1⊣Ras→...→ERK→Elk1	8.00E-4	0.050	Regulation of actin cytoskeleton	3.40E-12
ErbB signaling pathway	EGF→ErbB1→...→JNK→Elk	2.00E-3	0.063	Toll-like receptor signaling pathway	2.26E-13
Focal adhesion	GF→RTK→...→Rac→Actin	5.50E-3	0.12	Pathogenic Escherichia coli infection	8.69E-8
Apoptosis	TNFA→TNFR1→...→NFkB→IAP	8.60E-3	0.14	Type II diabetes mellitus	7.89E-13
p53 signaling pathway	ATM→ATR→...→CASP8→CASP3	1.49E-2	0.19	ErbB signaling pathway	1.10E-14
HK1					
GnRH signaling pathway	GnRH→GnRHR→...→MKK3/6→p38	9.00E-4	0.046	Small cell lung cancer	1.10E-14
Apoptosis	IL-1→IL-1R→...→NFkB→IAP	1.30E-3	0.046	ErbB signaling pathway	1.10E-14
MAPK signaling pathway	NGF→TrkA/B→...→ERK→Tau	2.60E-3	0.062	Gap junction	6.8E-14
Fc epsilon RI signaling pathway	Lyn→Syk→PKC	6.90E-3	0.098	Melanogenesis	1.94E-13
Regulation of actin cytoskeleton	GF→RTK→...→Rac→PAK	6.90E-3	0.098	Toll-like receptor signaling pathway	2.26E-13

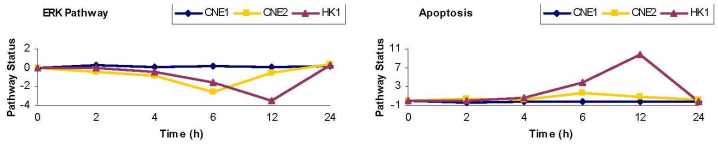

Fig. 2. Comparable status of ERK pathway and Apoptosis pathway of the 3 NPC cell lines along the treatment of CYC202.

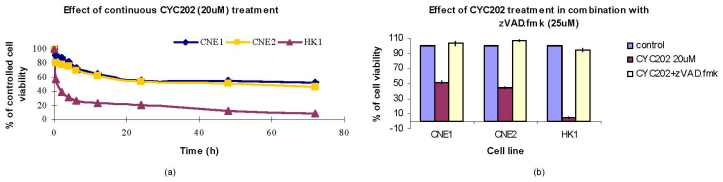

Fig. 3. Results of the associated medical assays to measure the cell viability and apoptosis level under the treatment of CYC202 for NPC cell lines: (a) The results of trypan blue test for measuring the cell viability along the drug treatment. (b) The extent of caspase-dependent apoptosis. zVAD.fmk is a caspase activity inhibitor.

Table 2. Hypothesized post-treatment signaling pathway status in patients.

Patient	Response	ERK	Apoptosis
Pt5	P(ositive)	-2.25	1.34
Pt8	P	-	0.82
Pt9	P	-0.97	-
Pt14	P	-	-0.86
Pt16	P	-0.20	1.42
Pt17	P	-1.02	1.01
Pt19	P	-	0.91
Pt18	No Tumor	**-0.15**	**0.13**
Pt1	N(egative)	0.21	-1.00
Pt7	N	-0.10	0.11
Pt10	N	1.02	-1.57
Pt15	N	-	-1.01
Pt20	N	1.30	-1.68

under drug treatment. From the figure, ERK pathway is significantly suppressed in the responder, HK1, but less down regulated or almost unchanged in the half-responder, CNE2, and the resister, CNE1. The apoptosis pathway is on the opposite. ERK pathway regulates cell survival, proliferation and differentiation. Suppression of this pathway represses cell viability. Therefore, the observation of ERK pathway in Figure 2 is consistent with known treatment outcome. Apoptosis pathway, on the other hand, regulates cell death. Since this pathway is induced in HK1, HK1 should be more sensitive to the treatment. To biologically prove the hypothese of these two pathways, trypan blue test and tunel assay were used to measure the level of cell proliferation and apoptosis in these three cell lines. Results support our conclusion (Figure 3).

We compare our results with that of PathwayExpress [3] (shown in Table 1). PathwayExpress is a web-based application which evaluates gene expression data in framework of KEGG pathway database. The main issue of PathwayExpress is that it does not differentiate regulations within a signaling pathway. In Table 1, we list the top five pathways identified by each method. Results of PathwayExpress show a very high statistical significance level. However, we suspect the correctness of the p-value measurement, since 61 out of 83 signaling pathways are more significant than 10^{-4} when we applied PathwayExpress to CNE1. (see supplementary material for whole results)

3.3 Application on the *in vivo* NPC dataset

We applied Decipherer to the *in vivo* NPC datasets with $p \leq 0.05$ and $FDR \leq 0.25$. Table 2 gives a summary of the hypothesized post-treatment signaling pathway status (see supplementary material for genetic pathway hypotheses). Pt18 is a non-tumor sample. Other patients are classified into two groups according to their treatment response. In Table 2, since Pt18 is a normal sample and pathway status of Pt18 does not change much after treatment, we consider the pathway

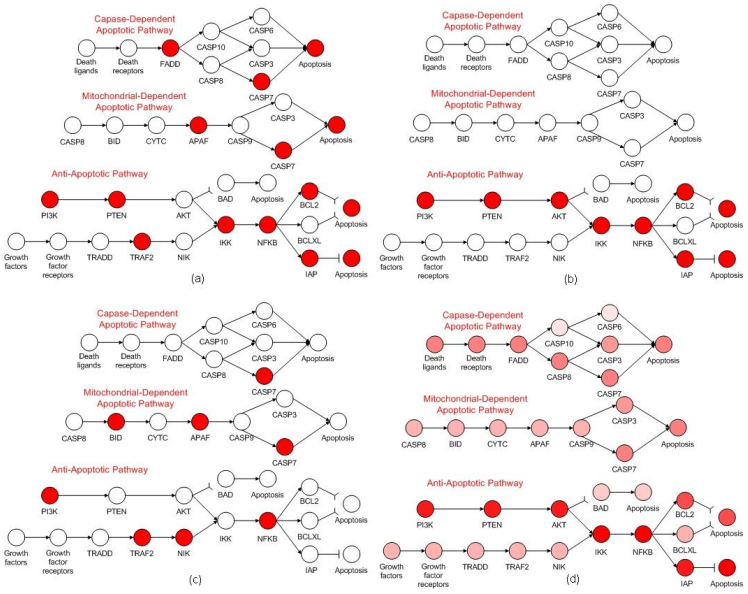

Fig. 4. Contrast results of the genes identified by the leading edge analysis of GSEA and the pathways identified by Decipherer: (a) GSEA performed on the *in vitro* dataset. The identified genes are highlighted. (b) Decipherer performed on the *in vitro* dataset. Since the identifications of CNE1 and HK1 on Apoptosis pathway are the same, the pathway is highlighted in the figure. (c) GSEA performed on the *in vivo* dataset. (d) Decipherer performed on the *in vivo* dataset. Color density represents the frequency of a pathway identified in patients.

status of Pt18 to be a benchmark to evaluate drug response of other patients. An interesting observation is that post-treatment status of ERK pathway and apoptosis pathway in two responding groups can be almost perfectly separated by that of Pt18 (except for Pt14 in apoptosis pathway). This observation suggests suppression of ERK pathway and induction of apoptosis pathway have correlation with effective CYC202 treatment *in vivo*, and this argument also agrees with the conclusion of the *in vitro* experiment.

The leading edge analysis of GSEA selects a subset of genes that mostly differentiate two phenotype groups. Decipherer generates hypotheses of genetic pathway, which can be regarded as selecting a subset of genes from the whole gene set defined by signaling pathway. Thus, we compare Decipherer with the leading edge analysis of GSEA. The same RE values were taken as input to GSEA. All parameters were remained with default values. Gene sets were extracted from Decipherer pathway database.

As shown in Figure 4, we report the results of comparison on apoptosis pathway since it is one of our main concerned pathways in this study, yet it is identified by GSEA in both *in vivo* and *in vitro* datasets with statistical signifi-

cance ($p \leq 0.05$ and FDR≤ 0.25). For cell lines, both GSEA and Decipherer identified the pattern of PI3K-NFκB-IAP anti-apoptosis pathway (Figure 4a and 4b). However, GSEA missed AKT, and selected some other irrelevant genes. This is because the purpose of the leading edge analysis is to select genes expressed differently between the two response groups, but relationships between selected genes are ignored. When applied to the *in vivo* dataset, GSEA did not identify any strong genetic pathway pattern (Figure 4c), but for Decipherer, multiple pathways with different significance were identified (Figure 4d). The most significant identification is still PI3K-NFκB-IAP pathway, which indicates the main genetic response of patients in apoptosis pathway is similar to that of cell lines. Another discovery is the death receptor regulated pro-apoptosis pathway. This pathway is previously undiscovered in the *in vitro* experiments, which means there exists alternative apoptosis regulation pathway for CYC202 in NPC patients rather than in cell lines. Thus, we show that, compared with the leading edge analysis of GSEA, Decipherer generates more biologically meaningful results, which can be used as a guide for further drug research and disease treatment.

3.4 Biological reasoning and discussion

Epstein-Barr Virus (EBV) infection dysregulates NFκB, MAPK, JAK-STAT and PI3K-AKT pathways [21], and thus plays critical role in NPC pathogenesis [12]. Up regulation of NFKB2 and BIRC5 (IAP) induced by EBV increases resistance of NPC cells to apoptosis, which has been confirmed by RNA interference [15]. On the other hand, CYC202 inhibits CDK2, 7 and 9 through competitive inhibition of ATP binding [10]. CDK7 and CDK9 phosphorylate the carboxyl terminal domain of RNA polymerase II, which initiates the gene transcription. NFκB regulated genes and IAP family members are greatly affected because of their short protein halflife [9]. The suppression of genes in ERK pathway and anti-apoptosis pathway, for example MAPK1, MAPK3, MCL1, BCL2, BIRC4 and BIRC5, are frequently observed in CYC202 treatment [11, 22, 1, 14, 16, 8]. In this study, Decipherer identified different regulation of RAS-ERK cell proliferation pathway and PI3K-NFκB-IAP anti-apoptosis pathway between two outcome groups both *in vitro* and *in vivo*. With the support of literature, we conclude that these two pathways are the main drug pathways of CYC202 in human NPC cells. On the other hand, due to the diversity of individual genetic environment of patients, the hypothesized escape pathways are heterogeneous. The dysregulation of NFκB pathway and MAPK pathway are both commonly observed in CYC202 resisters. Based on these observations, we made an individual treatment proposal for these CYC202 resisters (See supplementary material for treatment proposal).

4 Conclusion and Future Work

During the past decade, mRNA microarray techniques have been greatly developed and have found many significant applications in biomedical research. In

this paper, we present a novel statistical gene expression evaluation framework to discover drug responsive pathways in treatment gene expression data. We decide to report the method, since we have applied Decipherer to the NPC study case and have found some meaningful results, which have been considered to be applied to improve the CYC202 based NPC treatment in clinic. Thus, we think of Decipherer to be a potential valuable direction to follow. However, we realize that Decipherer has some caveats. We list following caveats as a conclusion of this paper:

- Signaling pathway is a protein level description, while gene expression is an mRNA level measurement. In Decipherer, we use correlation of expression of adjacent genes to socre a genetic pathway, which seems to lack enough biological support. Our explanation is that we try to impose a data-driven induction that if we can find a pathway satisfying the criteria, then we have evidence to support the pathway to be interesting. Readers may suspect the correctness of our criteria. We believe that the criteria are still very initial and need to be improved.
- We compare pathways from KEGG database, the Pathway Ingenuity database, and WikiPathways database [28]. A surprising observation is that pathways with same pathway name may have maximal approaching 50% disagreement with each other (personal communication with Soh). This observation has questioned our framework since Decipherer is only capable of evaluating known pathways. We believe that some pathway inferences should be included into the future version.
- Metabolic pathways are another important data source for drug action hypothesizing. Patients may response differently to treatment because of their different metabolic rates. For example, two of us recently found, in a colon cancer study, that Fluorouracil, a pyrimidine analog, affected two different metabolic pathways: an effective pathway and a drug degradation pathway. The expression of genes on these two pathways were observed to have strong correlation with treatment outcome [20]. Thus, we believe that metabolic pathways should be considered as well.

Currently, a new version of Decipherer is in development. We have made improvements to overcome some of the caveats listed above:

- A new set of rules has been designed for Decipherer to generate hypotheses. Firstly, a hypothesized pathway should not have broken logic in any responder. For example, if $\langle$HRAS,RAF1,\$*activation*$\rangle$ is a relationship for evaluation, we need to exclude the case that HRAS is greatly induced while RAF1 is significantly repressed. Secondly, a hypothesized pathway should be significantly perturbated by treatment in all responders. Thirdly, the regulation of a hypothsized pathway should be consistent in all responders. Finally, a resister should at least violate one of the three criteria above.
- The challange of taking metabolic pathway into the current design of Decipherer is that metabolic pathways are commonly circular systems, and thus

it is difficult to define source and sink genes in a pathway. This issue comes similar with the problem we have mentioned in Section 2.5 that arbitrarily taking source and sink genes in signaling pathways makes the number of hypothesis candidates increase exponentially. To solve the problem, we employ a dynamic procedure to only keep and extend valid candidates for each length, rather than generate all candidates at one time.

Acknowledgement

The NPC patient data are kindly provided by Dr. Boon Cher Goh, National University Hospital Singapore. This work is supported in part by a NUS research scholarship (Dong) and a MOE AcRF Tier 1 grant (Wong).

References

1. A. Alvi, B. Austen, V. Weston, et. al. A novel CDK inhibitor, CYC202 (R-roscovitine), overcomes the defect in p53-independent apoptosis in B-CLL by down-regulation of genes involved in transcription regulation and survival. *Blood*, 105:4484-4491, 2005.
2. S. W. Doniger, N. Salomonis, K. D. Dahlquist, et. al. MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology*, 4(1):R7, 2003.
3. S. Draghici, P. Khatri, A. Tarca, et. al. A systems biology approach for pathway level analysis. *Genome Research*, 17(10):1537-1545, 2007.
4. Z. Guo, Y. Li, X. Gong, *et. al.* Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network, *Bioinformatics*, 23:2121-2128, 2007.
5. H. Herrington. Controlling the false discovery rate in multiple hypothesis testing, `http://www.unt.edu/benchmarks/archives/2002/april02/rss.htm`.
6. T. Ideker, O. Ozier, B. Schwikowski, and A. F. Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18(Suppl. 1):S233-S240, 2002.
7. M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya. The KEGG database at GenomeNet. *Nucleic Acids Research*, 30(1):42-46, 2002.
8. K. Lacrima, A. Valentini, C. Lambertini, et. al. *In vitro* activity of cyclin-dependent kinase inhibitor CYC202 (Seliciclib, R-roscovitine) in mantle cell lymphomas. *Annals of Oncology*, 16:1169-1176, 2005.
9. L. Lam, O. Pickeral, A. Peng, et. al. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, 2(10):research004, 2001.
10. S. McClue, D. Blake, R. Clarke, et. al. *In vitro* and *in vivo* antitumor properties of the cyclin dependent kinase inhibitor CYC202 (R-ROSCOVITINE). *International Journal of Cancer*, 102:463-468, 2002.
11. L. Meijer, A. Borgne, O. Mulner, et. al. Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinase cdc2, cdk2 and cdk5. *European Journal of Biochemistry*, 243:527-536, 1997.

12. R. Pathmanathan, U. Prasad, R. Sadler, et. al. Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. *The New England Journal of Medicine*, 333:693-698, 1995.
13. C. Pui and W. Evans. Acute lymphoblastic leukemia. *New England Journal of Medicine*, 339:605-615, 1998.
14. N. Raje, S. Kumar, T. Hideshima, et. al. Seliciclib (CYC202 or R-roscovitine), a small-molecule cyclin-dependent kinase inhibitor, mediates activity via down-regulation of MCL1 in multiple myeloma. *Blood*, 106:1042-1047, 2005.
15. W. Shi, C. Bastianutto, A. Li, et. al. Multiple dysregulated pathways in nasopharyngeal carcinoma revealed by gene expression profiling. *International Journal of Cancer*, 119:2467-2475, 2006.
16. P. Smith and E. Yue. *Inhibitors of Cyclin-dependent Kinases as Anti-tumor Agents*, Taylor and Francis Group, 2006.
17. D. Soh, D. Dong, Y. Guo and L. Wong. Enabling more sophisticated gene expression analysis for understanding diseases and optimizing treatments, *ACM SIGKDD Explorations*, 9:3-14, 2007.
18. A. Subramanian, P. Tamayo, V. Mootha, *et. al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proceedings of the National Academy of Science of the United States of America*, 102:15545-15550, 2005.
19. K. Sultanem, H. K. Shu, P. Xia, et. al. Three-dimensional intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: the University of California, San Francisco experience. *International journal of radiation oncology, biology, physics*, 48:711-722, 2000.
20. W. Tan, D. Dong, M. Loh, et. al. Pathway determinants of 5-Fluorouracil activity. Poster, presented on 20th EORTC-NCI-AACR symposium on "Molecular targets and Cancer Therapeutics", 2008.
21. S. Tsao, G. Tramoutanis, C. Dawson, et. al. The significance of LMP1 expression in nasopharyngeal carcinoma. *Cancer Biology*, 12:473-487, 2002.
22. S. Whittaker, M. Walton, M. Garrett and P. Workman. The cyclin-dependent kinase inhibitor CYC202 (R-Roscovitine) inhibits retinoblastoma protein phosphorylation, causes loss of cyclin D1, and activates the mitogen-activated protein kinase pathway. *Cancer Research*, 64:262–272, 2004.
23. E. Yeoh, M. Ross, S. Shurtleff, et. al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, 1:133-143, 2002.
24. M. Yu and J. Yuan. Epidemiology of nasopharyngeal carcinoma. *Seminars in Cancer Biology*, 12:421-429, 2002.
25. B. R. Zeeberg, W. Feng, G. Wang, et. al. GoMiner: A resource for biological interpretation of genomic and proteomic data. *Genome Biology*, 4(4):R28, 2003.
26. A. Zien, R. Kuffner, R. Zimmer and T. Lengauer. Analysis of gene expression data with pathway scores, in *Proceedings of International Conference on Intelligent Systems for Molecular Biology*, 8:407-417, 2000.
27. Pathway Ingenuity database, `http://www.ingenuity.com/`.
28. WikiPathways, `http://www.wikipathways.org/index.php/WikiPathways`.