Full Report

ibc Interdisciplinary Bio Central

Open Access, Open Review Journal

Inferring Relative Activity between Pathway and Downstream Genes to Classify Melanoma Cancer Progression

Inkyung Jung[1], Jungsul Lee[1], Chulhee Choi[1] and Dongsup Kim[1,*]

[1]Department of Bio and Brain Engineering, KAIST, Daejeon 305-701, Republic of Korea

Subject areas:
Bioinformatics/Computational biology/Molecular modeling

Author contribution: I.J., J.L., C.C. and D.K. designed the methods and experimental setup; I.J. and J.L. carried out the implementation of the various methods; I.J. wrote the manuscript under J.L., C.C. and D.K.; All authors have read and approved the final manuscript.

***Correspondence** and requests for materials should be addressed to D.K. (kds@kaist.ac.kr).

Reviewer: Daehee Hwang, POSTECH, Republic of Korea; Sun Shim Choi, Kangwon National University, Republic of Korea

Editor: Keun Woo Lee, Gyeongsang National University, Republic of Korea

Received February 15, 2011;
Accepted February 18, 2011;
Published February 28, 2011

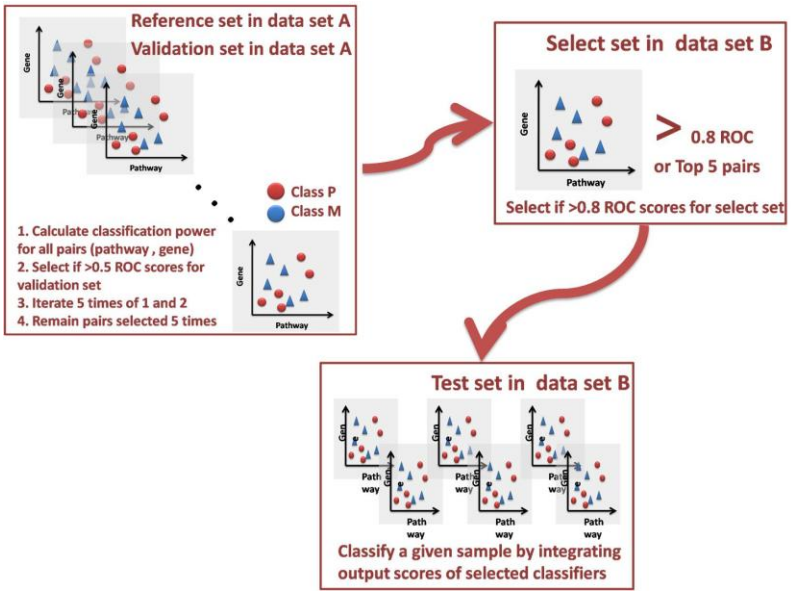
Citation: Jung, I., et al. Inferring Relative Activity between Pathway and Downstream Genes to Classify Melanoma Cancer Progression. IBC 2011, 3:5, 1-5.
doi: 10.4051/ibc.2011.3.1.0005

Funding: This work is supported by CHUNG Moon Soul Center for BioInformation and BioElectronics (CMSC), and by Korea Science and Engineering Foundation.

Competing interest: All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Introduction: Many signal transduction pathways mediate cell's behavior by regulating expression level of involved genes. Abnormal behavior indicates loss of regulatory potential of pathways, and this can be attributed to loss of expression regulation of downstream genes. Therefore, function of pathways should be assessed by activity of a pathway itself and relative activity between a pathway and downstream genes, simultaneously.
Results and Discussion: In this study, we suggested a new method to assess pathway's function by introducing concept of 'responsiveness'. The responsiveness was defined as a relative activity between a pathway itself and its downstream genes. The expression level of a downstream gene as a function of an upstream pathway activation characterizes disease status. In this aspect, by using the responsiveness we predicted potential progress in cancer development. We applied our method to predict primary and metastatic status of melanoma cancer. The result shows that the responsiveness-based approach achieves better performance than using gene or pathway information alone. The mean of ROC scores in the responsiveness-based approach was 0.90 for GSE7553 data set, increased more than 40% compared to a gene-based method. Moreover, identifying the abnormal regulatory patterns between pathway and its downstream genes provided more biologically interpretable information compared to gene or pathway based approaches.



Keywords: melanoma cancer progression, biomarker, responsiveness, classifier, drug target

Introduction

Phenotypic features of cells depend on signal transduction that regulates expression levels of specific phenotype related genes. Malicious phenotype might be attributed to abnormal regulation of signal transduction that induces abnormal gene expression patterns. Identification of over- or under- activated genes or pathways is a first step to understand the mechanism for disease associated phenotypes. Many researchers have tried to develop computational methods to detect abnormality of gene expression through analysis of genome-wide gene expression data[1-4]. These methods have focused on identification of differentially expressed genes to classify samples into phenotypically distinct groups. Although those methods have achieved great performance by using several powerful maker genes, they are limited in that they do not consider cellular heterogeneity within tissues and genetic heterogeneity across patients. For this reason, most of recognized maker genes have not been used in practical clinical use.

To solve this problem, protein-protein interaction or pathway information based approaches have been suggested. A network-based classification approach by using protein-protein interaction information provided powerful discriminate power in prediction of breast cancer metastasis[5]. The network-based approach was extended to a pathway-based approach which utilizes condition specific pathway information[6]. These approaches show high performance compared to conventional gene-based classifiers and give more robust results. Nevertheless, it is rather difficult to judge how they contribute to broaden our understanding of mechanism of complex disease because they do not focus on the role of pathway or group of genes in context of cellular network. To understand disease mechanism, dynamics of the responsiveness or interactions should be measured. Without consideration of interacting partners it is hard to understand the role and reason of abnormal patterns in given genes or pathways. In this point of view, a module based approach might be more relevant toward precise disease classification[7]. Breast cancer outcomes were well discriminated using dynamic modularity in a protein interaction network.

Here, we propose a novel classification method by assessing pathway's function with expression profiles, considering both pathway itself and genes that are regulated by the pathway. We identified several ill-operated pathway-downstream gene pairs as classifiers in primary and metastatic melanoma cancer samples. We named this concept as 'responsiveness' and concluded that the responsiveness is one of important information for understanding cancer progression.

Results and Discussion

Concept of 'responsiveness'

The metastatic progression in cancer development might be originated from abnormal regulation between pathway and its downstream genes. We defined abnormal regulation in terms of the responsiveness between upstream pathway activity and downstream gene expression. Figure 1 shows an example of abnormal regulation in metastatic melanoma cancer. Expression of interleukin 5 receptor (IL5RA, (A)) nor interleukin-2 pathway (IL-2, (B)) did not show significant difference between primary and metastatic samples. However, expression level of IL5RA gene tends to down regulated as IL-2 pathway is activated in primary samples, but not in metastatic samples (C). This result is consistent with previous report where IL5RA gene is revealed as being down regulated by IL-2 pathway[8]. We expected that the loss of expression regulation of IL5RA by IL-2 pathway is one of manifestations accompanied by the progression from primary to metastatic melanoma (D). This example suggests that the responsiveness is a different standpoint to understand disease mechanism, and considering both pathway activity and expression level of downstream genes allows us to recognize abnormal regulation.

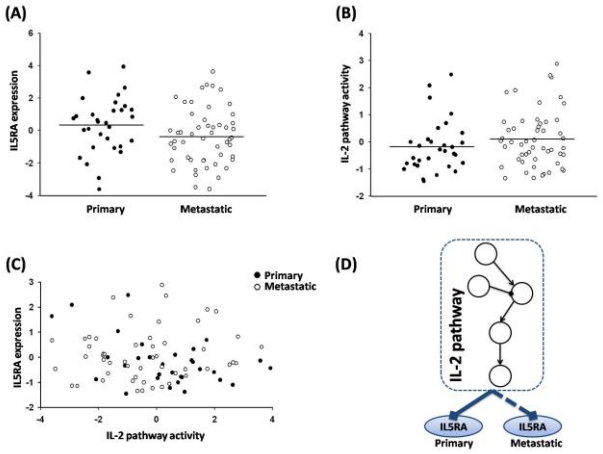


Figure 1. Abnormal responsiveness of IL5RA to IL-2 pathway in melanoma cancer. Expression of a gene (IL5RA, (A)) nor a pathway regulating the gene expression (IL-2, (B)) does not show difference in primary and metastatic samples. IL5RA expression has negative relationship with IL-2 signaling activity in primary melanoma samples, but not in metastatic melanoma cancer (C). Regulatory potential of IL-2 on its downstream gene IL5RA disappears in metastatic melanoma cancer (D).

Overview of classification with the responsiveness

The abnormality defined by the responsiveness might be one of important proprieties to characterize disease status. Based on this hypothesis, we applied the concept of the responsiveness to classify primary and metastatic samples in melanoma cancer.

The overall procedure is shown in Figure 2. We generated two independent data sets, A and B, one for building classifiers and another for measuring performance. To search candidate classifiers

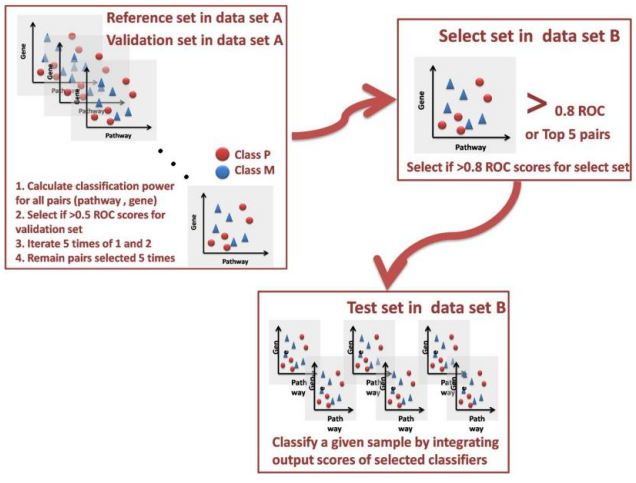


Figure 2. Overall procedure of the new method using the responsiveness to classify cancer samples into primary (Class P) and metastatic (Class M). In first step we searched candidate classifiers from data set A and then tested for data set B. To search candidate classifiers we divided data set A into 5 sub sets, one set as a validation set and others as a reference set. The classification ability was measured for all pair of a gene and a pathway using validation set. The output score of each pair was calculated according to a reference set. When we tested selected classifiers to data set B output scores were measured by integrating output scores of all classifiers.

from data set A, we divided data set A into 5 subsets where one set was assigned as a validation set and the others as a reference set. The classification ability was measured for all combinations of pathways and their downstream genes using the validation set. If the ROC score was over 0.5 for the given pair of a pathway and a gene, we selected the pair as a putative classifier. We iterated these procedures 5 times as changing a validation set and a reference set, and then selected candidate classifiers which were indentified every time during 5 iterations. For the test, the half of samples in data set B was used as a select set and another half as a test set. By using select set we filtered candidate classifiers with cutoff value of 0.8 ROC score. If there were not enough number of selected classifiers, top 5 classifiers were used. Finally, we calculated output scores by integrating all selected classifiers. The output scores were calculated as following.

$$S(t_i) = \sum_j \frac{Dist(t_i^j, M^j) + 1}{Dist(t_i^j, P^j) + 1}$$

where, t_i is a test sample i, and M and P a set of metastatic and primary samples, respectively, in a reference set. Also j indicates a classifier j. The $Dist(t_{ji},M_j)$ and $Dist(t_{ji},P_j)$ were calculated by taking shortest Euclidian distance between t_{ji} and all data points in M_j and P_j, respectively. If a given sample close to primary samples rather than metastatic samples in the reference set the scoring scheme gives a relatively high score.

Performance assessment with various methods

We predicted melanoma cancer progression by using responsiveness markers. We conducted cross-validation using two independent data sets. The performance was measured by calculating mean ROC scores for 200 trials. For the gene- or pathway- based approach, we used the same genes and pathways which were used in the new method. Figure 3 shows performance variation with various methods including responsiveness-, gene-, and pathway-based method. The mean ROC score of responsiveness-based approach achieves better performance than the other two methods. In Figure 3, GSE7553 (or GSE8401) indicates that classifiers were generated using GSE8401 (or GSE7553) and tested using GSE7553 (or GSE8401). The mean of ROC scores in the new method was 0.90 (for GSE7553) and 0.83 (for GSE8401). In GSE7553 the performance was increased more than 40% compared to a gene-based method. However, the performance was not significantly increased in GSE8401 since several genes were enough to classify samples. We conclude that the responsiveness-based approach is effective for predicting cancer progression.

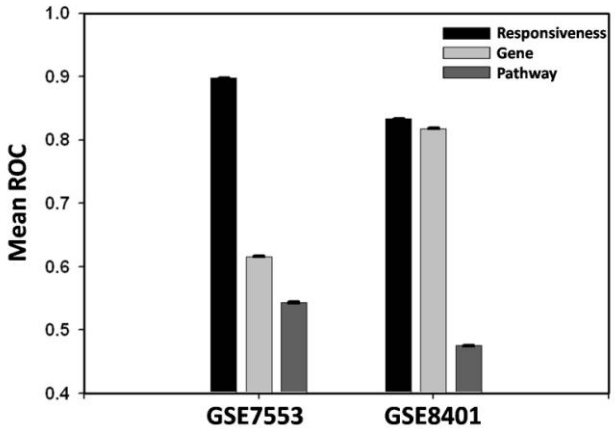


Figure 3. Various performance assessments. In Figure 'responsiveness' indicates the performance of classification by using the new method. The responsiveness shows better performance compared to other gene- or pathway-based methods in classification of melanoma cancer progression.

During 200 trials, various pairs of a gene and a pathway were selected as classifiers. Table 1 shows frequently selected classifiers in each data set. Among top 15 classifiers, 8 classifiers were overlapped in both data sets (bolded line). This result implies that the responsiveness-based approach is a robust measurement for predicting cancer progression, and these commonly recognized makers are effective for classifying the status of unknown samples.

The biological relevance of selective markers

From Table 1, we noted that several classifiers were commonly recognized in both data sets, implying that those classifiers provide putative therapeutic targets. EGFR1, TNFalpha, IL-1, and IL-4 were frequently selected as classifiers (Table 1). For all commonly recognized classifiers, gene expression levels were increased as corresponding pathways were activated in primary samples, but these relationships were not well maintained in metastatic samples. The pattern of primary samples matched with normal status, suggesting that primary status maintains proper relationship as a normal status but abnormal patterns reduce stability, leading to metastatic status. From literature search we noted that SERPINB3 (serine proteinase inhibitor member 3) was reported as up-regulated in benign hyperplasia in melanoma cancer[9]. Furthermore, in terms of strategies for targeting melanoma, a single targeted agent was designed to control EGFR1, and EGFR1 kinase inhibitors were used for antiangiogenic[10]. In metastatic state the relationship between SERPINB3 and EGFR1 was broken. To alter this abnormal state into proper state, two therapeutic approaches are possible; activation of SERPINB3 or inhibition of EGFR1

Table 1. Selected markers from both data sets

GSE7553 → GSE8401			GSE8401 → GSE7553		
Gene (HPRD ID)	Pathway	Frequency	Gene (HPRD ID)	Pathway	Frequency
02745 (up, SERPINB3)	**EGFR1**	**182**	**02745**	**IL-4**	**179**
02645 (up, S100A7)	**EGFR1**	**180**	**02745**	**EGFR1**	**143**
03185 (up, SFN)	**EGFR1**	**179**	**06371**	**EGFR1**	**142**
06371 (up, FGFBP1)	**EGFR1**	**168**	**02745**	**IL-1**	**133**
03185	IL-2	152	**02645**	**EGFR1**	**105**
01018	EGFR1	141	**03185**	**EGFR1**	**102**
02745 (up, SERPINB3)	**IL-1**	**132**	**03866**	**TNFalpha**	**101**
01012	EGFR1	131	01010	TGFbeta	100
02746	IL-1	116	05495	IL-4	82
09002	a6b4Integrin	111	**01017**	**TNFalpha**	**80**
02745 (up, SERPINB3)	**IL-4**	**105**	05495	TGFbeta	74
01017	TGFbeta	91	01420	IL-1	70
03866 (up, ANXA8)	**TNFalpha**	**85**	01015	IL-1	69
01017 (up, KRT14)	**TNFalpha**	**66**	00579	ID	67
00121	TGFbeta	64	08970	IL-2	64

The overlapped makers in both data sets are presented as bolded line.

pathway. EGFR1 kinase inhibitors might be a plausible approach to recover an abnormal relationship between SERPINB3 and EGFR1 into normal status.

The responsiveness-based classifiers are more relevant to biological context compared to gene- or pathway-based approaches since the function of gene or pathway should be assessed by considering their interacting partners. For this reason, interpretation of the classifiers generated by the new method can provide clues to identify new therapeutic targets or reveal unknown mechanism.

Conclusion and Prospects

The present study suggests that considering both pathway activity and expression level of its downstream gene is effective in predicting cancer progression. Identifying the abnormal regulatory patterns between pathway and its downstream gene helped to design effective classifiers. We found several powerful classifiers and one of them coincided with a previously reported experimental result. Based on these results, we argue that abnormal regulation between EGFR1, TNFalpha, IL-1, and IL-4 pathway and their downstream gene has important information for understanding of melanoma cancer progression. In addition, these newly found genes can be putative drug targets. All our results indicate that identifying abnormal regulation might be an effective way to understand cancer progression mechanism and give new information which cannot be found in conventional gene- or pathway-based approaches.

Materials and Methods

Microarray dataset and processing

We designed a new approach to discriminate cancer progression between primary and metastatic status in melanoma cancer. The data were downloaded from GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/projects/geo/) database. Two independent data sets, GSE7553[11] and GSE8401[12], were used; one for identification of candidate markers and the other for evaluation. For the data set we conducted inter-chip normalization by adjusting the median intensities to the same value. The GSE7553 and GSE8401 contain 42 and 31 primary samples, respectively, and 40 and 52 metastatic samples, respectively, for human. We converted all Affymetrix probe IDs into HPRD ID[13], then selected overlapped genes in both datasets. As a result, 8,654 genes were used to generate classifiers. For the duplicated genes the intensity was defined by using the average intensity of the probes. Intensity of each gene was z-transformed by subtracting average intensity and dividing by standard deviation of the each sample.

Pathway and responsive genes

In this study we focused on 20 pathways and its downstream genes which were manually curated as cancer related pathways in NetPath (http://www.netpath.org). Activity of the pathway was defined by calculating Welch's T-test t-score. The high t-score indicates that activation of genes included in a specific pathway shows significant different distribution compared to that of all genes. For the gene activity we used normalized z-transformed scores described in the above section.

Performance assessment

To test the performance, we conducted cross-validation by using independent data set. First, candidate classifiers were identified by using GSE7553 (or GSE8401) and filtered those classifiers by using half of samples in GSE8401 (or GSE7553), then tested another half of samples in GSE8401 (or GSE7553) by using filtered classifiers. This procedure was repeated 200 times, and the performance was measured by calculating the mean ROC scores[14]. Our scoring scheme was designed to give relatively high scores to primary samples compared to metastatic samples. We let primary samples as a positive example and metastatic samples as a negative example. ROC score is defined as the areas under the ROC curves, the plot of true positives as a function of the number of false positives

Acknowledgements

This work is supported by CHUNG Moon Soul Center for BioInformation and BioElectronics (CMSC), and by Korea Science and Engineering Foundation. Inkyung Jung thanks the members of Bioinformatics and Computational Biology Laboratory and Yoosun Lee for helpful discussions.

References

1. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503-511.
2. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537.
3. Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 99, 6567-6572.
4. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-5121.
5. Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D., and Ideker, T. (2007). Network-based classification of breast cancer metastasis. *Mol Syst Biol* 3, 140.
6. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T., and Lee, D. (2008). Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* 4, e1000217.
7. Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., and Wrana, J.L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol* 27, 199-204.
8. Fung, M.M., Chu, Y.L., Fink, J.L., Wallace, A., and McGuire, K.L. (2005). IL-2- and STAT5-regulated cytokine gene expression in cells expressing the Tax protein of HTLV-1. *Oncogene* 24, 4624-4633.
9. Haider, A.S., Peters, S.B., Kaporis, H., Cardinale, I., Fei, J., Ott, J., Blumenberg, M., Bowcock, A.M., Krueger, J.G., and Carucci, J.A. (2006). Genomic analysis defines a cancer-specific gene expression signature for human squamous cell carcinoma and distinguishes malignant hyperproliferation from benign hyperplasia. *J Invest Dermatol* 126, 869-881.
10. Sosman, J.A., and Puzanov, I. (2006). Molecular targets in melanoma from angiogenesis to apoptosis. *Clin Cancer Res* 12, 2376s-2383s.
11. Riker, A.I., Enkemann, S.A., Fodstad, O., Liu, S., Ren, S., Morris, C., Xi, Y., Howell, P., Metge, B., Samant, R.S., et al. (2008). The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC Med Genomics* 1, 13.
12. Xu, L., Shen, S.S., Hoshida, Y., Subramanian, A., Ross, K., Brunet, J.P., Wagner, S.N., Ramaswamy, S., Mesirov, J.P., and Hynes, R.O. (2008). Gene expression changes in an animal melanoma model correlate with aggressiveness of human melanoma metastases. *Mol Cancer Res* 6, 760-769.
13. Prasad, T.S., Kandasamy, K., and Pandey, A. (2009). Human Protein Reference Database and Human Proteinpedia as

discovery tools for systems biology. *Methods Mol Biol* 577, 67-79.

14. Gribskov, M., and Robinson, N.L. (1996). Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. *Comput Chem* 20, 25-33.