BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 5 2011, pages 649–654
doi:10.1093/bioinformatics/btq714

Gene expression

Advance Access publication December 24, 2010

A sub-pathway-based approach for identifying drug response principal network

Xiujie Chen[1,*,†], Jiankai Xu[1,2,†], Bangqing Huang[1,†], Jin Li[1], Xin Wu[3], Ling Ma[1], Xiaodong Jia[1], Xiusen Bian[1], Fujian Tan[1], Lei Liu[1], Sheng Chen[1] and Xia Li[1,*]

[1]College of Bioinformatics Science and Technology, [2]Department of Pharmacology and [3]The Third Affiliated Hospital, Harbin Medical University, Harbin, China

Associate Editor: John Quackenbush

ABSTRACT

Motivation: The high redundancy of and high degree of cross-talk between biological pathways hint that a sub-pathway may respond more effectively or sensitively than the whole pathway. However, few current pathway enrichment analysis methods account for the sub-pathways or structures of the tested pathways. We present a sub-pathway-based enrichment approach for identifying a drug response principal network, which takes into consideration the quantitative structures of the pathways.

Result: We validated this new approach on a microarray experiment that captures the transcriptional profile of dexamethasone (DEX)-treated human prostate cancer PC3 cells. Compared with GeneTrail and DAVID, our approach is more sensitive to the DEX response pathways. Specifically, not only pathways but also the principal components of sub-pathways and networks related to prostate cancer and DEX response could be identified and verified by literature retrieval.

Contact: chenxiujie@ems.hrbmu.edu.cn; lixia@hrbmu.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 13, 2010; revised on December 14, 2010; accepted on December 17, 2010

1 INTRODUCTION

It is increasingly important to understand the effects of treatment drugs on cellular pathways and networks. Microarrays provide a novel way to investigate the expression of thousands of genes simultaneously, with a view to dissecting the molecular pathways (Huang da *et al.*, 2009). With the availability of biological pathway databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) or BioCarta (www.biocarta.com), several resources have been developed to visualize and analyze microarray data in the context of known biological networks/pathways. A number of statistical tests combined with known biological databases for detecting significant pathways are available for data interpretation, including GenMAPP (Dahlquist *et al.*, 2002), ChipInfo (Zhong *et al.*, 2003), DAVID (Sherman *et al.*, 2007), GoMiner (Zeeberg *et al.*, 2005), GSEA-P (Subramanian *et al.*, 2007), Onto-Tools (Khatri *et al.*, 2007) and GeneTrail (Backes *et al.*, 2007).

These functional analysis tools are valuable but are unsuitable for the detection of drug response pathways, due to a number of limitations. First, most traditional strategies use only pre-selected gene lists (often the differentially expressed genes) to compute the statistic that tests the enrichment of each pathway. Common statistical methods, such as chi-square, Fisher's exact test, binomial probability and hypergeometric distribution, are used to compare number of genes in the list that coincide with a given pathway by random chance. Neglecting less relevant genes, and merely counting the number of pathway members contained in the reduced list, leads to loss of information such as gene order in the pathways. Second, while the prevalent gene-set enrichment analysis (GSEA) (Subramanian *et al.*, 2005) does account for gene order, the effect of the gene-set size, the influence of other gene sets and normalization procedures may lead to inaccurate assessment of statistical significance in some instances (Damian and Gorfine, 2004). Third, the correlation structure of pathways is not considered. A pathway extracted from biological databases is the collection of specific sub-pathways or modules that perform certain functions. Different sub-pathways may perform the same function in the same pathway and different pathways may use the same sub-pathways in similar roles. Testing the whole pathway is too universal to determine which individual sub-pathways respond to a particular biological condition. Fourth, drug-induced cell microarray experiments usually contain few samples (typically, a treated sample and a control sample). Thus, the pathway significance test is limited to the traditional strategies mentioned above.

To overcome these limitations, we propose a framework of sub-pathway-based strategy to identify the drug response pathways/network. The framework consists of the generation of a large number of relative sub-pathways (sub-graphs of the pathways obtained from the KEGG public database), mapping of the unfiltered expression data onto them and statistically scoring for identification of the principal component of sub-pathways (PCSs) that is most perturbed by two stage designs. PCSs are then combined into a larger drug response network, on which topological and biological analyses are performed. In an analysis of the dexamethasone (DEX) dataset, we found that this approach can enhance the biological interpretation of drug effect.

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

2 METHODS

2.1 Data description, preprocessing and normalization

Our DEX response dataset contained seven cell intensity files (CEL) downloaded from CMAP (Lamb, 2007) (Instance ID 2079), one derived from PC3 cell lines treated with 9.2 μM DEX and the other six derived from dimethyl sulfoxide (DMSO)-treated PC3 cell lines. The raw data were scaled using the R RMA package. For each sample, the expression values of all probes for a given gene were reduced to a single value by taking the maximum expression value. For the DMSO-treated samples, the expression value for a given gene was the average of the six samples, effectively reducing these samples to a single DMSO-treated sample.

2.2 Obtainment of sub-pathways by parsing the KEGG pathway

Each pathway map stored in KEGG can be downloaded in its own XML format (named KGML), which contains information about nodes in the map, such as their relationships, their coordinates, and what genes or compounds exist in the nodes. The pathway maps describe the combination of ordered linear sequences of protein direct/indirect interactions or metabolic reactions from biologically meaningful start-points (membrane receptors or their ligands) to end-points (transcriptional factors or their immediate targets). In such a map, the start-points have no parent nodes and the end-points have no descendant nodes. The sub-pathway is defined by an individual path from a start-point to an end-point in a pathway map. For a given KEGG pathway, the sub-pathways were obtained by searching all possible paths between start-points and end-points in the adjacency matrix generated by node relationships extracted from KGML files.

2.3 Identification of significant drug–response principal sub-pathways

To identify significant principal sub-pathways responding to a drug, two stage tests are employed (Lin, 2006; Nguyen *et al*., 2009; Skol *et al*., 2007; Wang *et al*., 2006).The first is used to define a subset of candidate sub-pathways, from which the second can capture the principal sub-pathways.

In the first-stage test, the sub-pathway state is represented by a vector $p_j=[g_1,g_2,\ldots,g_n]$, where g_i is the expression value of gene i in sub-pathway j. Here, we employ Euclidean distance $d_j=\|p'_j-p_j\|_2$ to describe the state change of each sub-pathway, where p'_j and p_j represent the states of sub-pathway j after and before drug treatment, respectively. A larger distance corresponds to greater sensitivity of the sub-pathway to the drug. When there are several genes in a sub-pathway node, the maximum expression-fold-change gene is chosen for the node. To identify significant drug response sub-pathways, we take the statistic for all sub-pathways and draw 100 000 random gene sets of the same size from all the genes to represent the background distribution. The P-value is calculated as the fraction of re-sampled gene set statistic that exceeds the observed value (P-value cutoff 0.05).

The second-stage test focuses on the key genes (PCS) that drive the drug response sub-pathways. Our goal is to develop an approach that would identify PCSs and their networks most influenced by drugs. First, we selected the top n genes that gave cumulative contribution at least up to 80% to significant sub-pathway state change and designated these PCS. The contribution was calculated by

$$I_j=\frac{\sum_{k=1}^{k_0}(g_i'^{(k)}-g_i^{(k)})^2}{\sum_{i=1}^{n}(g'_i-g_i)^2},$$

where g'_i and g_i are the expression values of gene i in the presence or absence of drugs, respectively, k is the rank of gene i contribution and n is the size of sub-pathway j. Next, the significance of PCSs were assessed by the same method used to identify significant drug response sub-pathways (cf. the first stage test in the above paragraph), and the P-value was adjusted by false discovery rate (FDR, cutoff 0.05) (Benjamini and Hochberg, 1995). Thus, the significant PCS is the sub-pathway since this PCS dominates the state change of the sub-pathways. Effectively, the sub-pathways containing the significant PCSs represent significant responses to a drug. To prioritize these PCSs with respect to DEX response in PC3 cell lines, the PCS were scored by

$$S_c=\frac{\sum_{q=1}^{u}f_{cq}}{\sum_{c=1}^{w}\sum_{q=1}^{u}f_{cq}},$$

where f_{cq} is the frequency of gene q in PCS c (composed of u genes) appearing in w PCSs.

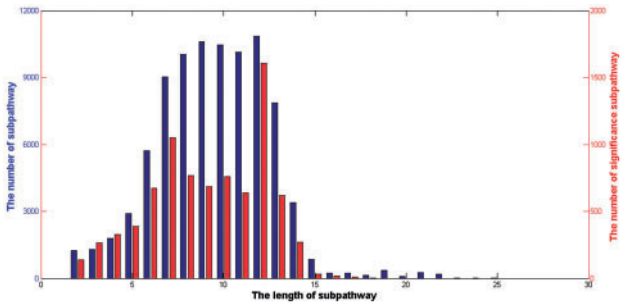

Fig. 1. The length distribution of (significant) sub-pathways. The blue bar is for the whole sub-pathways parsed from 169 pathways and the red bar is for the significant sub-pathways.

2.4 Construction and analyses of principal network to drug

From the identified PCSs and their structures (key gene relationships) maintained from their original sub-pathways, we constructed a principal drug response network and analyzed its topological properties using NetworkAnalyzer (Assenov *et al*., 2008).

3 RESULTS

3.1 General information of sub-pathway

The 201 pathway XML files were downloaded from KEGG, and 169 pathways involving metabolism, genetic information processing, environmental information processing, cellular processes and human diseases were used for further analyses after excluding the pathways lacking node relationships. In total, 87 732 sub-pathways were generated from the 169 KEGG pathway maps by the depth-first search algorithm. These contained 4179 genes and were highly redundant since each gene participated in approximately 20 sub-pathways. The maximum number of sub-pathways (16 154) was identified for the glycosphingolipid biosynthesis-lacto and neolacto series pathways, while several pathway maps, such as the tetrachloroethene degradation pathway, were parsed into only one sub-pathway (Supplementary Table 1).

3.2 Significant response sub-pathway to DEX in first-stage test

A total of 8252 sub-pathways containing 825 genes were significantly identified ($P<0.05$), indicating that the response genes were highly shared, or cross-talked, between the sub-pathways or pathways. Significant sub-pathway lengths, derived from 92 KEGG pathways (Supplementary Table 2), ranged from 2 to 18, with an average of 9 (Fig. 1).

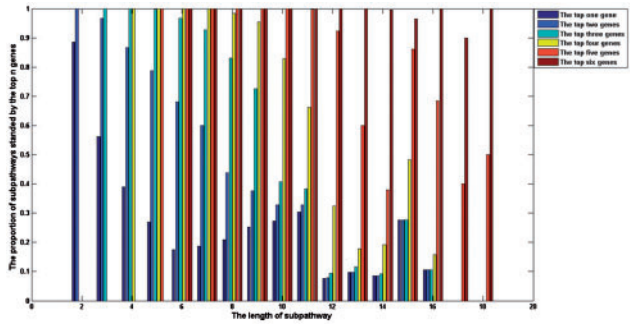

Fig. 2. The rate distribution of cumulative contribution of the top n genes up to 80% for the different length of significant sub-pathways. At most, six genes could represent the majority of sub-pathways, with only three sub-pathways requiring more key genes (i.e. only three sub-pathways could not be utilized under our specified conditions).

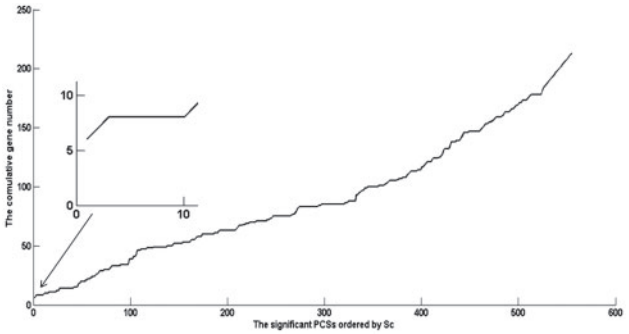

Fig. 3. The comulative gene number in the significant PCSs. The first 10 PCSs ordered by Sc are steadily composed of eight genes (Fig. 4).

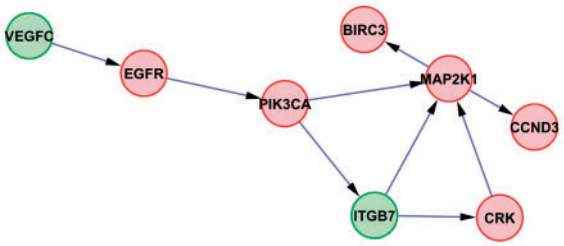

Fig. 4. The sub-pathway merging the first 10 PCSs by Sc. The pink circle is for increased expression level and the green circle is for decreased expression level. The decreased VEGFC and ITGB7 lead to the block of angiogenesis and the suppression of lymphocyte migration, respectively; the increased expression of other genes promotes PC3 cell proliferation and inhibits the apoptosis.

The proportions of significant sub-pathways varied for different pathways. For example, in the Circadian rhythm pathway, all sub-pathways were significant, whereas the proportion of significant sub-pathways in the Arachidonic acid metabolism pathway was only 0.063%.

The state changes in these significant sub-pathways were determined by the genes within them. However, these genes affected their sub-pathway states to different extents. The expression level changes of these genes dominated the whole sub-pathway, thus their cumulative contributions were scored by the statistic (Supplementary Figure 1). More than 90% of sub-pathways containing less than nine genes and more than 85% of sub-pathways containing 9–18 genes, were controlled by 2–3 genes and by 4–6 genes, respectively (Supplementary Figure 1). This implies that only a few genes (PCSs) regulate the drug response sub-pathways. Therefore, focusing on PCS genes might provide an economical way of elucidating the molecular mechanisms of drug response.

3.3 PCSs to DEX

As described above, the entire sub-pathway seems to be governed by only a few genes. It is therefore necessary to know which genes dominate the sub-pathway response state and the maximum number of these genes required. Therefore, we ranked the genes by their contributions to the state change of sub-pathway from high to low, and defined PCS standing for the sub-pathway by the cumulative contribution of the top n genes up to 80%. We then calculated the proportions of different length PCSs to different length sub-pathways (Fig. 2). We obtained 555 PCSs (FDR < 0.05) covering 213 genes, and found not only that one or several genes were highly shared in more than two PCSs, but also that the PCSs were highly shared by sub-pathways; indeed, one PCS could represent an average of 16 sub-pathways. These PCSs represented the principal states of DEX response pathways/sub-pathways. In addition, as shown in Figure 2, at most six genes can represent one sub-pathway. As they were highly shared, it could be inferred that shorter sub-pathways are linked in longer sub-pathways, consistent with the idea that signal cascades choose the most economical path *in vivo*.

Conventionally, the next step is to select the most important PCSs in the drug response sub-pathway. To this end, we applied the Sc statistic to evaluate the priorities of PCSs to DEX in PC3 cell lines (Supplementary Table 3). However, it was difficult to determine the optimal PCS since the PCSs were too highly cross-talked. For example, the top 10 PCSs contained eight genes, with four genes in common and with no gene additions from the top third to the top tenth PCS (Fig. 3). This indicates that the first 10 PCSs are strongly related. This trend was apparent throughout the PCS list; the appearance of a new gene in a PCS signified that no new genes would appear in the next several PCSs (Fig. 3; Supplementary Table 3). Therefore, we used the gene direct/indirect connections from their original sub-pathways to merge the first 10 PCSs into a single sub-pathway (Fig. 4). The new sub-pathway reveals the basic molecular mechanisms and effects of DEX in PC3 cell lines.

3.4 The principal response network to DEX

To reveal the comprehensive molecular mechanisms of DEX in PC3 cell lines, we merged all PCSs and constructed a DEX response principal network, since, as described above, PCSs for the significant sub-pathways are highly cross-talked (Fig. 5). The correlation coefficient of node degree distribution between the principal network and the network merging of all significant sub-pathways was 0.69 ($P < 0.01$), implying that the principal network had inherited the main structure of the latter. The top 10 genes with degree and betweenness centrality are shown in Supplementary Table 4. These signify important genes through which DEX exerts its complex effects in PC3 cell lines. In addition, there exist several isolated

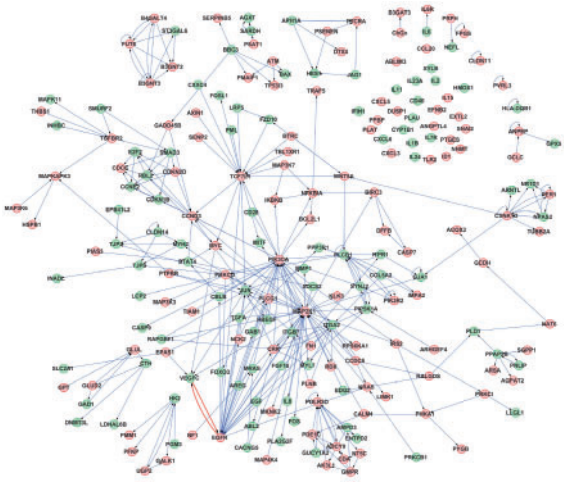

Fig. 5. DEX response principal network. The pink circle is for increased expression level and the green circle is for decreased expression level. The double-link between VEGFC and EGFR yields additional information about the interaction between the effects of DEX and cancer cell self-protection.

genes in the network which respond positively to DEX and which contribute more than 80% to their own sub-pathways. These were either the direct target of glucocorticoid receptor (e.g. ANGPTl4) (Koliwad *et al*., 2009) or the genes influenced by DEX (e.g. ILB1, CXCL5; Goleva *et al*., 2008; Waterman *et al*., 2006) (Supplementary Table 5). It should be noted that it was doubly-linked between VEGFC and EGFR, contrary to what was observed in the sub-pathway merging the top 10 PCSs (Fig. 4). DEX was found to stabilize EGFR and decrease VEGFC to suppress angiogenesis, whereas PC3 cells increased EGFR as a survival response. The double-link yields additional information about the interaction between the effects of DEX and cancer cell self-protection.

3.5 The comparison with the method which links the differentially expressed gene pairs in the KEGG pathways

To identify drug response sub-pathways, it was a direct and simple method (M2) to link the differentially expressed gene pairs in the KEGG pathways. Therefore, we compared M2 with our method (M1) and found that few differentially expressed gene pairs in the KEGG pathways could be connected (Supplementary Table 6). When the fold change cutoff was 2, 28 genes were differentially expressed and 13 genes were found in the KEGG pathways and no path (or connection) was found between any differentially expressed gene pair in one KEGG pathway. When the fold change cutoff lowered at 1.5, 62 genes were differentially expressed and 28 genes were decentralized in 54 KEGG pathways and only one path of four genes (Senate-Notch-CSL-Hes1/5, Senate and Hes1/5 differentially expressed) was found in Notch Signaling pathway. These results show that most genes of linear sub-pathways or pathways do not change their expression very large based on current knowledge of pathways. Therefore, we considered the sub-pathway as a whole responding to drug and developed M1. Further, we lowered the fold change cutoff from 1.4 to 1.2 and the result of M2 varied very large (Supplementary Table 6). It is not easy to choose a fold change cutoff here. Cross-talks are very universal between pathways or between sub-pathways. For example, in Jak-stat signaling pathway, SHP2, GRB and SOS are common genes in the seven sub-pathways which start from STAM, EPO, IFN, IL2/3, IL6, UPD or SHP1. When two or more sub-pathways have two common genes differentially expressed, M2 does not test which sub-pathway is significant response. M1 considers this problem and tests the sub-pathway significance in the first stage design. In addition, M2 has to keep the genes connecting differentially expressed genes to stand for the response sub-pathways, even if their expression changes are so small that their contributions to the state change of sub-pathways could be ignored. M1 selects the significant top n genes that give cumulative contribution at least up to 80% to significant sub-pathway state change and keeps their direct/indirect relationships to stand for the sub-pathways.

4 DISCUSSION

Current approaches typically study whole pathways, whether by singular enrichment analysis or by gene set enrichment analysis (Huang da *et al*., 2009). However, these methods may be unsuitable for the identification of drug response genes. A pathway comprises sets of expressed genes that lead to a complex phenotype (e.g. prostate cancer) or which perform a function (e.g. apoptosis). We generated 87 732 sub-pathways from 169 KEGG pathways. Within part of a pathway, some individual sub-pathways may give rise to disease or respond significantly to a drug. Single sub-pathways performing relatively simple functions coupled with specific paths may be more suitable for assessing drug response, than complex pathways which are currently poorly understood. To illustrate our point, we detected DEX response pathways in PC3 cell lines using GeneTrail. None of the pathways obtained from GeneTrail were related to biological pathways such as prostate cancer, cancer pathways in general, WNT signaling pathway, MAPK signaling pathway, etc. DAVID, with 28 genes (fold change >2) or 62 genes (fold change >1.5) detected null or only one pathway (cytokine–cytokine receptor interaction). In contrast, our methods identified 52 significant pathways (Supplementary Table 7), of which a subset was unequivocally associated with DEX response and prostate cancer (as determined by the number of significant sub-pathways containing significant PCSs in the pathways, evaluated by the hypergeometric test). Many pathways detected were triggered directly by DEX, and there was evidence of high cross-talk among the pathways. When a sub-pathway responds, its pathway must respond also, rendering further tests unnecessary. Combined, our results suggest that the sub-pathway based strategy is more sensitive to drug response than whole-pathway approaches.

A total of 825 genes were associated with 8252 significant sub-pathways, with an average of one gene participating in 10 sub-pathways, revealing high cross-talk between the genes. We hypothesize that the important genes change their expression often in the sub-pathways and have a high degree of standing in the sub-pathway network, since they are stimulated by multiple signals. To test this idea, we extracted 555 PCSs according to their contributions to the sub-pathway state changes, and found that the standing degrees of those genes were on average higher (t-test, $P = 5.59E$-08) than those of non-PCS genes in the sub-pathway-based network. Afterwards, we designed the statistic to evaluate the priorities of these PCSs to DEX in PC3 cell lines, which effectively ranked PCSs and hinted at strong links between the top 10 PCSs and DEX-treated

prostate cancer cells (Fig. 4). In addition, the cross-talking among sub-pathways could supplement existing pathway knowledge. For example, inspired by Rhee *et al*. (1995), we deduced that CCND3 influences proliferation of CCND1-expressing cells in the prostate cancer pathway (Fig. 4).

From supplementary Figure 2, we found that, at first stage, only 2–3 genes contributed >90% (median) to the significant sub-pathways containing less than 9 genes, and that 4–6 genes contributed >85% (median) to the sub-pathways containing 9–18 genes. As already mentioned, this implies that the response sub-pathways are mainly caused or represented by a few genes only (the PCSs). We defined six as the default number of genes in a PCS. PCSs with size less than six might not represent all significant sub-pathways in the first stage, since their contribution might not reach 80%. Figure 2 illustrates that, at most, six genes could represent the majority of sub-pathways, with only three sub-pathways requiring more key genes (i.e. only three sub-pathways could not be utilized under our specified conditions). Thus, fixing the contribution rate cutoff at 80% could minimize the number of lost significant sub-pathways in the first stage, while maintaining the maximum contribution rate.

The genes in PCSs are highly shared (213 genes in 555 PCSs, Supplementary Table 3), hence it is necessary to construct a drug response principal network by merging all of them for a systematic explanation of DEX's behaviors in PC3 cells. For 213 genes (26% of 825 genes), the principal response network represents at least the 80% response state of PC3 cells to DEX, as each PCS contributes more than 80% to its sub-pathway. In the principal response network, DEX's biological effects are reflected in the node functions. For example, increase of PIK3CA and decrease of either FOXO3 (Shang *et al*., 2009; Skurk *et al*., 2004) or CASP9 (Hakem *et al*., 1998) hint that DEX may promote proliferation and block apoptosis in PC3 cells. It is worth noting that the up and down nodes may disaccord as EGF and EGFR, since EGFR is triggered by a multitude of other signaling molecules. Although gene-by-gene analysis of the network reveals the main response molecular mechanisms of DEX in PC3 cells, we suggest that the function of a sub-pathway is better explained through its original pathway map. DEX, a synthetic glucocorticoid, has been shown to exhibit single-agent activity in prostate cancer (Venkitaraman *et al*., 2008). Therefore, we investigated the PCSs in pathways in cancer (KEGG ID: hsa05200), and found that DEX suppresses prostate cancer mainly by inhibiting sustained angiogenesis, while promoting apoptosis evasion and cell proliferation (Supplementary Figure 2). Similarly, we found that PCSs in the prostate cancer pathway (KEGG ID: hsa05215) play roles in apoptosis inhibition, cell proliferation and cell cycle progression (Supplementary Figure 3). Encouragingly, this finding is reflected in the top 10 PCSs ranked by the Sc statistic (Fig. 4).

DEX is the most commonly used type of steroid to treat the prostate cancer that has spread. (Alimirah *et al*., 2006). The response principal network exhibits several postulated mechanisms by which DEX exerts its effects in cancer. DEX has been shown to attenuate the transcriptional activity of NF-kB and AP1 (FOS and JUN) (Nishimura *et al*., 2001), resulting in the inhibition of IL6 and the NF–kB–IL6 pathways in prostate cancer (Akakura *et al*., 2003; De Bosscher *et al*., 2003). DEX has also been shown to suppress angiogenesis in a prostate cancer xenograft model, reducing IL8 and VEGF expression *in vitro* and *in vivo*, respectively (Yano *et al*., 2006). In addition, our results indicate that suppressed PC3 cells take a strong line of self-defense by activating proliferation sub-pathways, by inhibiting apoptosis sub-pathways, and even by promoting the circadian rhythm pathway to activate the WNT pathway (CSNK1D-WNT5A-TCF7L1-JUN-MAP2K1), which increases cytoactivity (Fu and Lee, 2003; Schwarz-Romond *et al*., 2002). These results suggest that DEX should be combined with other anti-tumor drugs (especially the EGFR inhibitors) in the effective treatment of prostate cancer. Consistent with this, combined inhibition of VEGF and EGFR is effective in EGFR inhibitor-resistant cell lines, and increases benefit compared with either of these anti-EGFR agents alone or when combined with chemotherapy (Tortora *et al*., 2008).

Reproducibility analysis of the principal DEX response network was performed using another batch instance (ID 5797), held under the same conditions as batch instance 2079. A total of 9045 sub-pathways (composed of 808 genes) and 632 PCSs (composed of 217 genes) were significantly identified, of which 4350 sub-pathways (approximately half of sub-pathways) and 71 key genes (approximately one-third of PCS genes) were identical to those found in batch instance 2079. The low reproducibility of key genes was expected, since it reflects the apparently low reproducibility of true discoveries in microarray studies (Zhang *et al*., 2008). The higher reproducibility of significant sub-pathways indicates that they perform specific functions, again consistent with the results of microarray studies, which are more reproducible at functional level (Gong *et al*., 2010). In addition, the functional similarity of the Gene Ontology (GO) terms enriched in the two PCS sets, evaluated by G-SESAME (Du *et al*., 2009) was 0.88, significantly greater ($P < 0.01$) than that of random gene set pairs of the same size. Moreover, a dataset of 90 samples, DEX-treated myeloma cells (GEO ID: GSE8546) was analyzed. The DEX response principal network (composed of 274 genes) in Myeloma, were significantly identified. Most genes (approximately 1/2) are associated with DEX, Myeloma or cancer (Supplementary Table 8). The functional similarity of GO terms enriched in the principal network from GSE8546, compared with those from instance 2079 and 5797, evaluated by G-SESAME (Du *et al*., 2009), were 0.837 and 0.864, significantly greater ($P < 0.01$) than that of random gene-set pairs of the same size. These results show that our sub-pathway-based approach can potentially detect stable drug response principal network at the functional level.

ACKNOWLEDGEMENTS

We thank Associate Editor John Quackenbush and the three reviewers for their suggestions and comments.

Funding: The Scientific and Technological Project of Heilongjiang Province (GC07C337 and ZD200804-01); Provincial Education Department Project of Heilongjiang (11541121); National Natural Science Foundation of China (30871394, 30972538 and 30772238); the National High Tech Development Project of China, the 863 Program (2007AA02Z329); the key national S&T Program-Major New Drug Development (2009ZX09103-380); The Innovation Manpower Fund of Harbin Science and Technology Bureau (2010RFXXS053); Master Innovation Funds of Harbin Medical University (HCX2009008 and HCXS2010007).

Conflict of Interest: none declared.

REFERENCES

Akakura,K. *et al*. (2003) Possible mechanism of dexamethasone therapy for prostate cancer: suppression of circulating level of interleukin-6. *Prostate*, **56**, 106–109.

Assenov,Y. *et al*. (2008) Computing topological parameters of biological networks. *Bioinformatics*, **24**, 282–284.

Alimirah,F. *et al*. (2006) DU-145 and PC-3 human prostate cancer cell lines express androgen receptor: implications for the androgen receptor functions and regulation. *FEBS Lett*., **580**, 2294–2300.

Backes,C. *et al*. (2007) GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res.*, **35**, W186–W192.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

Dahlquist,K.D. *et al*. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **31**, 19–20.

Damian,D. and Gorfine,M. (2004) Statistical concerns about the GSEA procedure. *Nat. Genet.*, **36**, 663; author reply 663.

De Bosscher,K. *et al*. (2003) The interplay between the glucocorticoid receptor and nuclear factor-kappaB or activator protein-1: molecular mechanisms for gene repression. *Endocr. Rev.*, **24**, 488–522.

Du,Z. *et al*. (2009) G-SESAME: web tools for GO-term-based gene similarity analysis and knowledge discovery. *Nucleic Acids Res.*, **37**, W345–W349.

Fu,L. and Lee,C.C. (2003) The circadian clock: pacemaker and tumour suppressor. *Nat. Rev. Cancer*, **3**, 350–361.

Goleva,E. *et al*. (2008) Corticosteroid-resistant asthma is associated with classical antimicrobial activation of airway macrophages. *J. Allergy Clin. Immunol.*, **122**, 550–559.e3.

Gong,X. *et al*. (2010) Extracting consistent knowledge from highly inconsistent cancer gene data sources. *BMC Bioinformatics*, **11**, 76.

Hakem,R. *et al*. (1998) Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell*, **94**, 339–352.

Huang da,W. *et al*. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.

Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Khatri,P. *et al*. (2007) Onto-Tools: new additions and improvements in 2006. *Nucleic Acids Res.*, **35**, W206–W211.

Koliwad,S.K. *et al*. (2009) Angiopoietin-like 4 (ANGPTL4, fasting-induced adipose factor) is a direct glucocorticoid receptor target and participates in glucocorticoid-regulated triglyceride metabolism. *J. Biol. Chem.*, **284**, 25593–25601.

Lamb,J. (2007) The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer*, **7**, 54–60.

Lin,D.Y. (2006) Evaluating statistical significance in two-stage genomewide association studies. *Am. J. Hum. Genet.*, **78**, 505–509.

Nguyen,T.T. *et al*. (2009) Optimal robust two-stage designs for genome-wide association studies. *Ann. Hum. Genet.*, **73**, 638–651.

Nishimura,K. *et al*. (2001) Potential mechanism for the effects of dexamethasone on growth of androgen-independent prostate cancer. *J. Natl Cancer Inst.*, **93**, 1739–1746.

Rhee,K. *et al*. (1995) c-Myc and cyclin D3 (CcnD3) genes are independent targets for glucocorticoid inhibition of lymphoid cell proliferation. *Cancer Res.*, **55**, 4188–4195.

Schwarz-Romond,T. *et al*. (2002) The ankyrin repeat protein diversin recruits Casein kinase I epsilon to the beta-catenin degradation complex and acts in both canonical Wnt and Wnt/JNK signaling. *Genes Dev.*, **16**, 2073–2084.

Shang,Y.C. *et al*. (2009) FoxO3a governs early microglial proliferation and employs mitochondrial depolarization with caspase 3, 8, and 9 cleavage during oxidant induced apoptosis. *Curr. Neurovasc. Res.*, **6**, 223–238.

Sherman,B.T. *et al*. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. *BMC Bioinformatics*, **8**, 426.

Skol,A.D. *et al*. (2007) Optimal designs for two-stage genome-wide association studies. *Genet. Epidemiol.*, **31**, 776–788.

Skurk,C. *et al*. (2004) The Akt-regulated forkhead transcription factor FOXO3a controls endothelial cell viability through modulation of the caspase-8 inhibitor FLIP. *J. Biol. Chem.*, **279**, 1513–1525.

Subramanian,A. *et al*. (2007) GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics*, **23**, 3251–3253.

Subramanian,A. *et al*. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.

Tortora,G. *et al*. (2008) Combined targeting of EGFR-dependent and VEGF-dependent pathways: rationale, preclinical studies and clinical applications. *Nat. Clin. Pract. Oncol.*, **5**, 521–530.

Venkitaraman,R. *et al*. (2008) Efficacy of low-dose dexamethasone in castration-refractory prostate cancer. *BJU Int.*, **101**, 440–443.

Wang,H. *et al*. (2006) Optimal two-stage genotyping designs for genome-wide association scans. *Genet. Epidemiol.*, **30**, 356–368.

Waterman,W.R. *et al*. (2006) Glucocorticoid inhibits the human pro-interleukin lbeta gene (ILIB) by decreasing DNA binding of transactivators to the signal-responsive enhancer. *Mol. Immunol.*, **43**, 773–782.

Yano,A. *et al*. (2006) Glucocorticoids suppress tumor angiogenesis and in vivo growth of prostate cancer cells. *Clin. Cancer Res.*, **12**, 3003–3009.

Zeeberg,B.R. *et al*. (2005) High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics*, **6**, 168.

Zhang,M. *et al*. (2008) Apparently low reproducibility of true differential expression discoveries in microarray studies. *Bioinformatics*, **24**, 2057–2063.

Zhong,S. *et al*. (2003) ChipInfo: software for extracting gene annotation and gene ontology information for microarray analysis. *Nucleic Acids Res.*, **31**, 3483–3486.