Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 909525, 8 pages
http://dx.doi.org/10.1155/2013/909525

Research Article

Molecular Signature of Cancer at Gene Level or Pathway Level? Case Studies of Colorectal Cancer and Prostate Cancer Microarray Data

Jiajia Chen,[1,2] Ying Wang,[1,3] Bairong Shen,[1] and Daqing Zhang[1]

[1] *Center for Systems Biology, Soochow University, Jiangsu, Suzhou 215006, China*
[2] *Department of Chemistry and Biological Engineering, Suzhou University of Science and Technology, Jiangsu, Suzhou 215011, China*
[3] *Laboratory of Gene and Viral Therapy, Eastern Hepatobiliary Surgical Hospital, Second Military Medical University, Shanghai 200438, China*

Correspondence should be addressed to Daqing Zhang; szdaq@126.com

Received 2 November 2012; Accepted 23 December 2012

Academic Editor: Tianhai Tian

Copyright © 2013 Jiajia Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With recent advances in microarray technology, there has been a flourish in genome-scale identification of molecular signatures for cancer. However, the differentially expressed genes obtained by different laboratories are highly divergent. The present discrepancy at gene level indicates a need for a novel strategy to obtain more robust signatures for cancer. In this paper we hypothesize that (1) the expression signatures of different cancer microarray datasets are more similar at pathway level than at gene level; (2) the comparability of the cancer molecular mechanisms of different individuals is related to their genetic similarities. In support of the hypotheses, we summarized theoretical and experimental evidences, and conducted case studies on colorectal and prostate cancer microarray datasets. Based on the above assumption, we propose that reliable cancer signatures should be investigated in the context of biological pathways, within a cohort of genetically homogeneous population. It is hoped that the hypotheses can guide future research in cancer mechanism and signature discovery.

1. Introduction

Microarray technology has evolved rapidly in the past several years as a powerful tool for large-scale gene expression profiling [1]. By monitoring changes in gene expression patterns, microarray technology is widely utilized in search of molecular signatures for many medical conditions including cancer. However, evidence is mounting that differentially expressed gene (DEG) lists detected from different studies for the same disease are often inconsistent [2, 3]. One might attribute the inconsistency to the variation in microarray platforms, experimental samples, normalization and analysis methods, and inherent biological uncertainty. Yet this discordance remains even in technical replicate tests using identical samples as in the case of Ein-Dor et al. [4]. Therefore, signature identification at the level of differential genes has been challenged about its robustness and reliability. In light of the inconsistency between DEG lists obtained from different datasets, we propose herein two hypotheses: (1) the expression signatures of different cancer microarray datasets are more similar at pathway level than at gene level; (2) the comparability of the cancer molecular mechanisms of different individuals is related to their genetic similarities. The hypotheses are subsequently verified by case studies of colorectal cancer and prostate cancer microarray datasets, respectively. Hopefully, the hypotheses would explain the inconsistency of the DEG lists derived from multiple experiments and provide novel methods for discovering robust and specific biomarkers of cancer.

2. Materials and Methods

2.1. Data Collection. We collected 5 gene expression profiling datasets on colorectal cancer and 10 datasets on prostate cancer from public gene expression data repositories, for example, Gene Expression Omnibus (GEO), Oncomine

Table 1: Colorectal cancer gene expression datasets used in the meta-analysis.

Dataset	Platform	Total genes	Total samples	Experimental design		Statistical method
				Normal	Tumor	
Hong	HGU133	54675	22	10	12	t-test
Sabates-Bellver	HGU133	54675	64	32	32	Mann-Whitney test
Galamb1	HGU133	54675	30	11	19	SAM
Galamb2	HGU133	54675	38	8	30	PAM
Graudens	cDNA	23232	30	12	18	z-statistics

SAM: significance analysis of microarrays; PAM: prediction analysis of microarrays.

[5] and Supplementary Materials from published literatures. The detailed information of the datasets was summarized in Table 1 for colorectal cancer and Supplementary Table 1 (see Supplementary Material available online at http://dx.doi.org/10.1155/2013/909525) for prostate cancer. These data were collected from two types of platforms, that is, cDNA two-channel arrays and Affymetrix microarray platforms including Human 6800 Affy gene chips, HG-U95A and HG-U133 series. Each dataset was named after the first author of the original literature. Only profiles of normal and cancer tissues were extracted for further analysis.

2.2. Preprocessing of Raw Data. The images of the cDNA array were processed using GenePix Pro 5.0.1.24 software. Background correction was performed by subtracting the median background intensities from the median foreground intensities of all spots in both channels. The raw datasets measured with Affymetrix chips were analysed via MAS5.0 algorithm in R platform. To eliminate the systematic error from heterogeneous datasets before the identification of signatures, we performed Locally Weighted Scatter Plot Smoothing (LOWESS) method for within-chip normalization of cDNA array's dataset and Median Absolute Deviation (MAD) method for between-chip normalization of all datasets. In addition, data was filtered to eliminate bad spots, and the filter criterion was defined as 60% absence across all of the samples. All of the data of preprocessing procedures were performed in R programming environment.

2.3. Determination of the Differentially Expressed Outlier Genes. Cancer Outlier Profile Analysis (COPA) method was performed for detecting genes that were differentially expressed between cancer and normal samples. We used COPA package by MacDonald and Ghosh [6] in R platform. According to the COPA package guidelines, the data was centered and scaled on a rowwise basis using median average difference. The rows of microarray expression data matrix were genes, and the columns were samples. The COPA function calculates a "COPA" score from a set of microarrays. As a preliminary step the function used a percentile for prefiltering the data. The number of outlier samples for each gene was calculated, and all genes with a number of outlier samples less than the percentile (default 95th) were removed from further consideration. A threshold cutoff for "outlier" status was set as 1.7 and applied to all genes.

2.4. Functional Enrichment of Outlier Genes. The significant outlier genes were subsequently mapped to functional databases, for example, GSEA [7], KEGG [8], and GeneGO (GeneGO, Inc.) for the pathway enrichment analysis. GSEA analysis and KEGG pathway analysis were performed using Gene Set Enrichment Analysis (GSEA) tool [7] and Onto-Express [9, 10], respectively. GSEA tool used a collection of gene sets from molecular signatures database (MSigDB), which was divided into five major collections. In our work, we used C2 curated gene sets. Enriched GeneGO pathways were detected by MetaCore (GeneGO, Inc) [11] software. P-value was used to evaluate the statistical significance of each candidate pathway. In MetaCore, the statistics significance (P-value) was calculated by using hypergeometric distribution. False Discovery Rate (FDR) adjustment was applied for multiple test correction.

2.5. Pairwise Overlapping Comparison at Gene/Pathway Level. The overlapping percentage between two datasets is calculated as follows:

$$\text{Overlapping percentage} = \frac{m}{n_1 + n_2 - m} \times 100\%, \quad (1)$$

where n_1 is the number of all the data in dataset 1, n_2 is the number of all the data in dataset 2, and m is the number of overlapping data between two datasets.

3. Results

3.1. Outlier Detection Using Novel Statistic Method. Table 1 listed the statistical methods for identifying differentially expressed genes by the original articles. Most of the prevailing analytical methods like t-test, SAM, and z-statistic considered the average value of gene intensities in the cancer samples. These statistical methods, however, would fail to find "outlier genes" which are only involved in subsets of the cancer samples. Despite their scarcity, outlier genes are nontrivial and may present a hallmark of potential oncogenes. These conventional methods are not suitable for detecting such subset-specific oncogene expression profiles as proposed by Tomlins et al. [12] and Lian [13]. Through applications to public cancer microarray datasets in our previous study [14], we have demonstrated that some newly developed statistics showed superior performance than traditional t-statistics in outlier detection. We herein applied Cancer Outlier Profile Analysis (COPA), a novel significant genes analysis method

Table 2: The number of pathway/gene sets enriched by differentially expressed gene for five colorectal cancer datasets.

Dataset	Number of enriched pathways in GeneGO	Number of enriched gene sets in GSEA
Hong	71	154
Sabates-Bellver	50	303
Galamb1	78	91
Galamb2	36	128
Graudens	149	172

proposed by Tomlins et al. [12], to meta-analyze multiple cancer datasets.

3.2. Signatures Are More Similar at Pathway Level across Multiple Colorectal Cancer Datasets. In order to verify our first hypothesis, we performed meta-analysis of 5 colorectal cancer gene expression profiling datasets from independent laboratories [15–19].

After COPA analysis, we identified 3258 genes differentially expressed between normal colorectal and colorectal tumor samples. The searches in the Entrez PubMed database showed that only 450 out of 3258 (13.8%) identified genes by COPA method were associated with colorectal cancer.

The number of overexpressed genes was obviously discrepant across all groups because of the different samples, arrays, and platforms. To decrease the discrepancy, we tried to understand the cancer molecular mechanism at systems biological level. We then mapped the DEGs identified by COPA using Gene Set Enrichment Analysis (GSEA) and MetaCore software for pathway enrichment analysis, respectively. Totally we found 262 enriched pathways in GeneGO's database with a P value threshold of 0.05; the detailed list of the pathways are provided in Supplementary Table 2. In addition, we performed the gene sets enrichment analysis in GSEA by using C2 curated file, which includes 1892 gene sets/pathway annotation. 111 outlier gene sets with NOM P-value <0.05 and FDR < 0.05 were also found and listed in Supplementary Table 3. The numbers of significant GeneGO pathways or GSEA gene sets enriched by the differentially expressed gene for 5 colorectal cancer datasets were listed in Table 2.

We performed pairwise comparison between 5 datasets in terms of DEGs, GSEA's enriched gene sets, and GeneGO's enriched pathways, respectively. For 5 different datasets, 10 pairs of datasets are available for comparison. Figure 1 showed the pairwise overlapping percentage at different observation levels. A significantly higher overlap at pathway level than at gene level is observed with 70% of the dataset pairs by GeneGO and 60% of the dataset pairs by GSEA. This observation supports our first hypothesis that the overlapping percentage at the pathway level is higher than that at the gene level.

Moreover, we found 4 GeneGO pathways that were shared by 4 datasets. These pathways were considered to be most overlapped and listed in Table 3. Among them, ECM remodeling, chemokines, and adhesion pathways, belonging

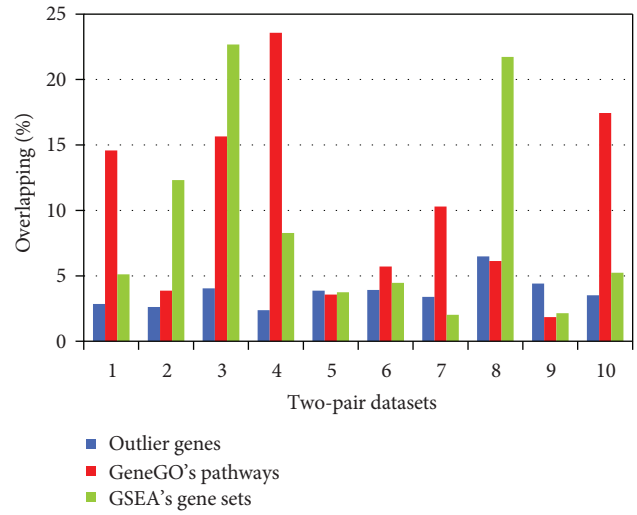


Figure 1: Pairwise overlapping percentage of 5 datasets among differentially expressed genes, enriched gene sets in GSEA, and enriched pathways in GeneGO database. The x-axis represented all the two-pair combination of 5 datasets. The y-axis represented the overlapping percentage.

to cell adhesion category, were previously reported to play a role in colorectal cancer. The other two pathways, integrin outside-in signalling pathway and L-selenoamino acids incorporation in proteins during translation pathway, have not been reported as colorectal cancer associated pathways. The network objects in both of the pathways, however, have been widely reported in colorectal cancer. Integrins are heterodimeric adhesion receptors, and most of them recognize ECM proteins. A major function of integrin signaling is to link ECM proteins to intracellular actin filaments via interactions of integrins with actin-binding proteins. Therefore, the correlation between integrin signaling and ECM pathway may play an active role in colorectal cancer. We infer that these two pathways might be putative novel colorectal cancer related pathways which could provide crucial guidance for biological scientists. Their roles in colorectal cancer need further experimental validation in the future.

We performed paired t-test to decide whether the different overlapping percentages observed between different levels are significant. The P-values for the difference between outlier genes and GeneGO's enriched pathways were 0.01354 by paired t-test and 0.02441 by Wilcoxon test. The P-values for the difference between outlier genes and GSEA gene sets were 0.028 by paired t-test and 0.08 by Wilcoxon test, respectively. The P-values indicate that the overlapping percentages at gene set or pathway level are significantly higher than that at individual gene level. We thus came to the conclusion that the expression signatures of independent datasets at higher functional level are significantly more consistent than that at gene level.

3.3. The Prostate Cancer Outlier Gene Enriched Pathways Show a Regional Distribution Feature. In support of the second hypothesis, we performed a regional analysis of 10

Table 3: The top 4 most overlapped GeneGO's pathways shared by 4 datasets.

GeneGO ontology	Pathway name	Pubmed citation count
Translation	(L)-selenoamino acids incorporation in proteins during translation	0
Cytoskeleton remodeling	Integrin outside-in signaling	0
Cell adhesion	ECM remodeling	64
Cell adhesion	Chemokines and adhesion	1117

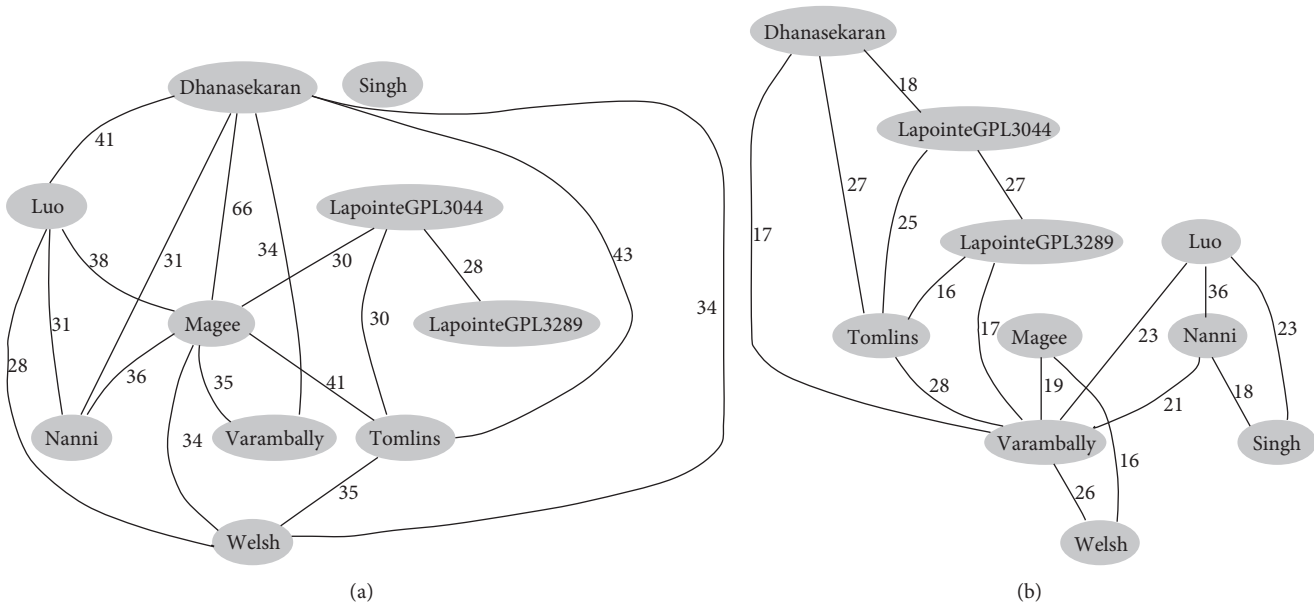


Figure 2: A simple network that associates datasets according to their similarity distances. The distances were calculated based on the overlapping percentage of the enriched pathways identified by (a) GeneGO and (b) KEGG. The lines between two datasets mean that their overlapping is more than two-thirds of the all. Each circle represented a dataset, and the overlapping percentage was shown on the lines.

publicly available prostate cancer gene-expression datasets from different locations [20–28].

We first conducted KEGG and GeneGO pathway enrichment analysis on these datasets, followed by a pairwise comparison of pathway overlapping percentage among them. Only the significantly enriched pathways with previous evidence of prostate cancer association were adopted for the comparison. Text mining was performed to make sure that there was at least one published paper describing the function of these pathways in prostate cancer.

Based on pathway overlapping analysis, we calculated the distance matrices between these datasets and generated a network to display their association. Five common distances, that is, Euclidean distance, Pearson correlational distance, Manhattan distance, Kendall's tau correlational distance, and Hamming distance were used to measure the similarity of these datasets. Based on these distances, a network graph was generated to display the association of these datasets. Figures 2(a) and 2(b) illustrate the association of the datasets based on GeneGO pathways and KEGG pathways, respectively.

Figure 2 revealed an essential regional distribution feature of significant pathways across multiple datasets. It is obvious from the graph that the distance between two Lapointed [29] datasets is the closest among all the datasets. Datasets by Dhanasekaran et al. [20], Tomlins et al. [25], and Magee et al. [23] feature a high pathway overlap which could be reflected by distances, indicating their similarities. The datasets from Singh et al. [26], Luo et al. [22], Welsh et al. [24], and Nanni et al. [27] diverge less from each other than those from the other six datasets.

We then investigated the regional sources of the tissue specimens for each dataset, as listed in Table 4. Samples of Dhanasekaran et al. [20] and Tomlins et al. [25] were obtained from the same place; those of Magee et al. [23] were close to them. Samples of Singh et al. [26], Welsh et al. [24] and Luo et al. [22, 30] came from adjacent states in America. Although the samples of Lapointe et al. [21] were not given a specific location, the author informed us their two experiment datasets were taken from patients of the same population. Apparently, there is obvious concordance between dataset similarity and sample source distribution.

Considering the influence by different microarray platforms, we compared the total unique genes of each dataset in order to testify that the significant pathway distribution feature is caused by different data sources rather than different experimental platforms. As implied in Figure 3, the similarities of the experimental platforms, here the overlapping proportion of the nonredundant probes used in different platforms, are not correlated to the regional distribution. Therefore, the regional distribution of cancer signature at pathway level is independent of the experimental platforms.

Table 4: Tissue specimen sources of each prostate cancer expression dataset.

Datasets	Tissue specimens sources	Locations
Dhanasekaran	University of Michigan Specialized Program of Research Excellence in Prostate Cancer (SPORE) tumor bank	America, Michigan (MI)
Lapointe	Stanford University; Karolinska Institute; Johns Hopkins University	America, California (CA); Sweden, just outside Stockholm; America, Maryland (MD);
Tomlins	University of Michigan	America, Michigan (MI)
Luo	Johns Hopkins Hospital	America, Maryland (MD)
Magee	Washington University School of Medicine; University of Washington Medical Center	America, Missouri (MO); America, Washington (WA);
Welsh	University of Virginia (UVA)	America, Virginia (VA)
Varambally	University of Michigan Prostate Cancer Specialized Program of Research Excellence (SPORE) Tissue Core	America, Michigan (MI)
Singh	Brigham and Women's Hospital	America, Massachusetts (MA)
Nanni	Regina Elena Cancer Institute	Italy, Rome

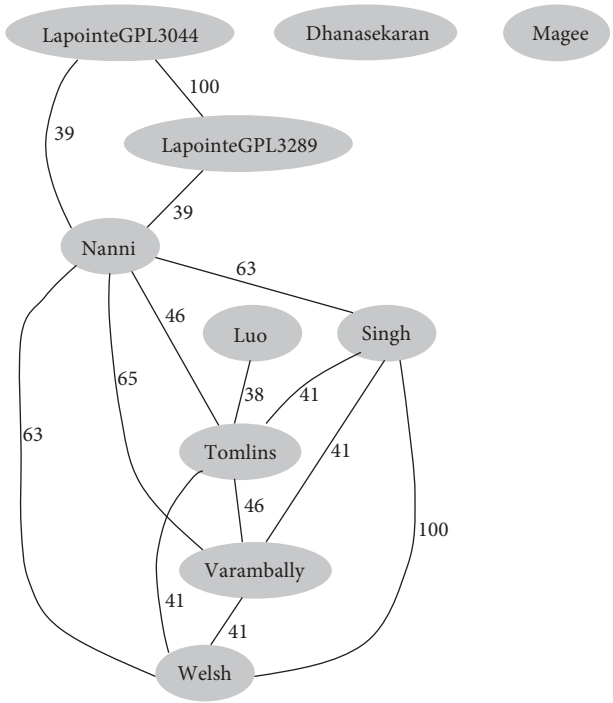


Figure 3: A simple network that associates datasets according to the similarity in microarray platforms. The distances represent the overlapping proportion of the probes used in different platforms.

4. Discussion

4.1. Comparison of DEGs between Different Experiments Revealed Little Overlap. The application of DNA microarrays for the investigation of cancer has led to numerous microarray studies that examined the same clinical conditions. Nevertheless, experiments from different groups have given dissimilar results when DEG lists are directly compared. The disparity was demonstrated in this study, where a meta-analysis of 5 colorectal cancer microarray expression datasets from 4 independent laboratories was performed. We calculated the pairwise overlapping proportion of DEGs between any two datasets, only to find that the overlap between the two lists was disappointingly small (~5%).

Such inconsistency has been observed in gene expression profiling of various types of cancer. For example, in two prominent studies that aimed to predict survival of breast cancer patients [31, 32], both groups claimed to have generated gene lists with predictive power, but only 17 genes appeared on both lists. In another attempt to predict the 5-year metastasis of breast cancer, van't Veer et al. [31] and Wang et al. [33] reported a list of gene sets with good prediction performance, respectively. But the predictive success of their studies was frustrated by the fact that the sets of metastasis-related genes identified by these two independent studies had only 3 overlapping genes. More recently our colleagues [3] meta-analyzed 10 independent microarray datasets associated with prostate cancer, but the resulting set of DEGs had only ~20% overlap between each datasets.

The most straightforward explanation of this lack of agreement is the variation in microarray platforms, experimental samples, normalization, and analysis methods. The open question is, however, whether the inconsistency can be attributed only to these trivial reasons?

To address the issue, Ein-Dor et al. [4] sought to remove all the technical differences mentioned above by analyzing a single breast cancer dataset [31] with a single method. By randomly generating training datasets, they demonstrated that the same analysis could have obtained many equally predictive gene lists and that two such lists share, typically, only a small number of genes. This finding indicates that low consistency occurs even in technical replicate tests using identical samples. The reason for this inconsistency or instability would be that (1) the number of DEGs is large whereas the number of samples is limited; (2) the resulting set of DEGs fluctuates according to the subset of patients used for gene selection.

4.2. Identifying Robust Molecular Signatures at Functional Modules Level or Pathway Level. In this study we evaluated the consistency of signatures across 5 colorectal cancer datasets produced by different platforms. Although the DEG lists selected had only ~5% overlaps, their enriched pathways were still consistent. Consistency analysis at different levels provides solid evidence that cancer signatures at pathway level diminish the discrepancies observed in direct comparisons of DEGs and are more consistent across multiple datasets than at gene level.

As the understanding of tumor biology deepens, it is well recognized that carcinogenesis is characterized with coordinated molecular changes. Functionally correlated genes often display coordinated expression to accomplish their roles; one would therefore expect that the inconsistent DEG lists across independent experiments are functionally more consistent. In other words, the discrepancies of DEGs would be less pronounced when they are mapped to functional groups or biological pathways.

Following this line, some previous studies have shifted their focus from individual genes to the biologically related groups of genes in the analysis of cancer microarray data. For example, in order to investigate the robustness of biological themes, Hosack et al. [34] applied the Expression Analysis Systematic Explorer (EASE) to determine the biological theme for DEG lists generated by various gene selection methods. Their research provided strong evidence that biological themes are stable to varying methods of gene selection. Zhu et al. [35] developed a novel tool for identifying cancer signatures at functional modules level. Its applications to two cancer types demonstrated that the functional modules enjoy explicit relevance to cancer biology. Recently, Yang et al. [36] proposed semantic similarity measure for DEG lists detected under varied statistical thresholds and from different studies. They reported that gene lists could be functionally consistent according to their semantic similarity. In addition, Gorlov et al. [37] conducted functional annotation analysis of the prostate cancer genes identified by two different methods. They observed a considerable overlap between biological functions identified by varied methods.

In recent years, pathway analysis has received a great deal of attention in the study of cancer microarray data [7, 34]. Pathway analysis typically correlates the identified DEGs with predefined pathway databases. It is reported that pathway analysis applied to differential gene lists detected under varied statistical methods yielded common results [38]. This discovery was validated in our previous study by Wang et al. [3], who evaluated the consistency of signature across 10 prostate cancer datasets produced by different platforms. Although the datasets share disappointingly few DEGs, their DEG-enriched pathways were still consistent.

4.3. Searching for Common Signatures among a Cohort of Genetic Homogeneous Population. As for the second hypothesis we assume that the individuals bearing similar genetic/environmental factors tend to share more common pathways. However, the information on the genetic/environmental characteristics of the patient samples is generally lacking. We believe it should be statistically reasonable to take the geographical location of the sample resources as the measurement of the similarities of their genetic/environmental factors. According to the similarity of outlier enriched pathways found by GeneGO and KEGG, we are able to classify 10 different prostate cancer related datasets into several groups. The datasets from same or adjacent geographical locations tend to reside within the same group. In other words, we observed an essential regional distribution feature of significant pathways across multiple datasets. In this sense molecular signatures from the geographically adjacent tissue specimens would be more consistent than those generated from geographically isolated samples. This observation is basically in accordance with our hypothesis that the comparability of the cancer molecular mechanisms of different individuals is related to their genetic similarities.

Cancer represents a heterogeneous disease, which reflects the interaction of a myriad of etiological and genetic contributions [39]. Therefore the gene expression profiles of cancer patients are diverse, depending on factors such as genetic information, environment effect, and personal behaviors. The role of genetic and environmental factors in modulating gene expression variation in humans has been extensively investigated. Most of the previous studies on cancer microarray profiling, however, ignored the interindividual variation in gene expression. It is likely that differences in expression that appear to be related with the disease may in fact represent random genetic variation. This situation will further introduce false discoveries and reduce the overall reproducibility of DEG detection. This concern was mentioned by Michiels et al. [40], who investigated the stability of seven published datasets to predict prognosis of cancer patients. It was observed that the predictive gene lists reported by the various groups were highly unstable and depended strongly on the subset of samples chosen for training.

It is assessed that, to achieve a typical overlap of 50% between two predictive lists of genes, the expression profiles of several thousands of patients would be needed [41]. Unfortunately, obtaining such a large number of samples is currently impractical due to limited tissue availability and financial constraints. A more practical approach would be to search for common signatures among a genetically homogeneous human population other than those among a mixed population. Although different individuals may have different regulatory mechanisms and discrepant cancer associated pathways, we assume that the individuals bearing similar genetic and environmental factors tend to share more common pathways.

Thus it would be reasonable to group patients into well-defined small subgroups on the basis of each person's unique genetic and environmental information. In this way, the individual difference of cancer mechanism is accounted when we analyze cancer expression data from different resources. This kind of investigation will help to find population-specific cancer pathways and facilitate personalized medicine.

5. Conclusions

Based on previous observations, we proposed herein two novel points of view for the cancer signatures identification. The pathway-based approach suggested in this paper would hopefully improve the comparability of different microarray datasets and, therefore, may lead to more valid and reliable biological interpretation of microarray results. Moreover, the generation of the population-specific cancer signatures would help to deliver effective therapy to patients most likely to benefit from such treatment and enable "personalized medicine." With increasing amount of cancer datasets available, the challenge in the future is to collect more cancer datasets from independent populations to prove our hypotheses.

Conflict of Interests

The authors declare they have no direct financial relation with the trademarks mentioned in this paper that might lead to a conflict of interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (91230117, 31170795) Grants, the Specialized Research Fund for the Doctoral Program of Higher Education of China (20113201110015), the International S&T Cooperation Program of Suzhou (SH201120), and the National High Technology Research and Development Program of China (863 program, Grant No. 2012AA02A601).

References

[1] J. Dopazo, E. Zanders, I. Dragoni, G. Amphlett, and F. Falciani, "Methods and approaches in the analysis of gene expression data," *Journal of Immunological Methods*, vol. 250, no. 1-2, pp. 93–112, 2001.

[2] M. Zhang, C. Yao, Z. Guo et al., "Apparently low reproducibility of true differential expression discoveries in microarray studies," *Bioinformatics*, vol. 24, no. 18, pp. 2057–2063, 2008.

[3] Y. Wang, J. Chen, Q. Li et al., "Identifying novel prostate cancer associated pathways based on integrative microarray data analysis," *Computational Biology and Chemistry*, vol. 35, no. 3, pp. 151–158, 2011.

[4] L. Ein-Dor, I. Kela, G. Getz, D. Givol, and E. Domany, "Outcome signature genes in breast cancer: is there a unique set?" *Bioinformatics*, vol. 21, no. 2, pp. 171–178, 2005.

[5] D. R. Rhodes, S. Kalyana-Sundaram, V. Mahavisno et al., "Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles," *Neoplasia*, vol. 9, no. 2, pp. 166–180, 2007.

[6] J. W. MacDonald and D. Ghosh, "COPA—cancer outlier profile analysis," *Bioinformatics*, vol. 22, no. 23, pp. 2950–2951, 2006.

[7] A. Subramanian, P. Tamayo, V. K. Mootha et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[8] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[9] S. Drăghici, P. Khatri, R. P. Martins, G. C. Ostermeier, and S. A. Krawetz, "Global functional profiling of gene expression," *Genomics*, vol. 81, no. 2, pp. 98–104, 2003.

[10] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture," *Nature Genetics*, vol. 22, no. 3, pp. 281–285, 1999.

[11] S. Ekins, A. Bugrim, L. Brovold et al., "Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms," *Xenobiotica*, vol. 36, no. 10-11, pp. 877–901, 2006.

[12] S. A. Tomlins, D. R. Rhodes, S. Perner et al., "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer," *Science*, vol. 310, no. 5748, pp. 644–648, 2005.

[13] H. Lian, "MOST: detecting cancer differential gene expression," *Biostatistics*, vol. 9, no. 3, pp. 411–418, 2008.

[14] Y. Tang, J. Chen, C. Luo, A. Kaipia, and B. Shen, "MicroRNA expression analysis reveals significant biological pathways in human prostate cancer," in *Proceedings of the 5th IEEE International Conference on Systems Biology (ISB '11)*, pp. 203–210, IEEE Computer Society, Zhuhai, China, September 2011.

[15] E. Graudens, V. Boulanger, C. Mollard et al., "Deciphering cellular states of innate tumor drug responses," *Genome Biology*, vol. 7, no. 3, article R19, 2006.

[16] O. Galamb, B. Györffy, F. Sipos et al., "Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature," *Disease Markers*, vol. 25, no. 1, pp. 1–16, 2008.

[17] O. Galamb, F. Sipos, N. Solymosi et al., "Diagnostic mRNA expression patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results," *Cancer Epidemiology Biomarkers and Prevention*, vol. 17, no. 10, pp. 2835–2845, 2008.

[18] Y. Hong, K. S. Ho, K. W. Eu, and P. Y. Cheah, "A susceptibility gene set for early onset colorectal cancer that integrates diverse signaling pathways: implication for tumorigenesis," *Clinical Cancer Research*, vol. 13, no. 4, pp. 1107–1114, 2007.

[19] J. Sabates-Bellver, L. G. Van der Flier, M. de Palo et al., "Transcriptome profile of human colorectal adenomas," *Molecular Cancer Research*, vol. 5, no. 12, pp. 1263–1275, 2007.

[20] S. M. Dhanasekaran, T. R. Barrette, D. Ghosh et al., "Delineation of prognostic biomarkers in prostate cancer," *Nature*, vol. 412, no. 6849, pp. 822–826, 2001.

[21] J. Lapointe, C. Li, J. P. Higgins et al., "Gene expression profiling identifies clinically relevant subtypes of prostate cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 3, pp. 811–816, 2004.

[22] J. Luo, D. J. Duggan, Y. Chen et al., "Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling," *Cancer Research*, vol. 61, no. 12, pp. 4683–4688, 2001.

[23] J. A. Magee, T. Araki, S. Patil et al., "Expression profiling reveals hepsin overexpression in prostate cancer," *Cancer Research*, vol. 61, no. 15, pp. 5692–5696, 2001.

[24] J. B. Welsh, L. M. Sapinoso, A. I. Su et al., "Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer," *Cancer Research*, vol. 61, no. 16, pp. 5974–5978, 2001.

[25] S. A. Tomlins, R. Mehra, D. R. Rhodes et al., "Integrative molecular concept modeling of prostate cancer progression," *Nature Genetics*, vol. 39, no. 1, pp. 41–51, 2007.

[26] D. Singh, P. G. Febbo, K. Ross et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.

[27] S. Nanni, C. Priolo, A. Grasselli et al., "Epithelial-restricted gene profile of primary cultures from human prostate tumors: a molecular approach to predict clinical behavior of prostate cancer," *Molecular Cancer Research*, vol. 4, no. 2, pp. 79–92, 2006.

[28] S. Varambally, J. Yu, B. Laxman et al., "Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression," *Cancer Cell*, vol. 8, no. 5, pp. 393–406, 2005.

[29] J. Lapointe, C. Li, J. P. Higgins et al., "Gene expression profiling identifies clinically relevant subtypes of prostate cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 3, pp. 811–816, 2004.

[30] J. Luo, D. J. Duggan, Y. Chen et al., "Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling," *Cancer Research*, vol. 61, no. 12, pp. 4683–4688, 2001.

[31] L. J. van't Veer, H. Dai, M. J. Van de Vijver et al., "Gene expression profiling predicts clinical outcome of breast cancer," *Nature*, vol. 415, no. 6871, pp. 530–536, 2002.

[32] T. Sørlie, C. M. Perou, R. Tibshirani et al., "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 19, pp. 10869–10874, 2001.

[33] Y. Wang, J. G. M. Klijn, Y. Zhang et al., "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *The Lancet*, vol. 365, no. 9460, pp. 671–679, 2005.

[34] D. A. Hosack, G. Dennis Jr., B. T. Sherman, H. C. Lane, and R. A. Lempicki, "Identifying biological themes within lists of genes with EASE," *Genome Biology*, vol. 4, no. 10, article R70, 2003.

[35] J. Zhu, J. Wang, Z. Guo et al., "GO-2D: identifying 2-dimensional cellular-localized functional modules in Gene Ontology," *BMC Genomics*, vol. 8, article 30, 2007.

[36] D. Yang, Y. Li, H. Xiao et al., "Gaining confidence in biological interpretation of the microarray data: the functional consistence of the significant GO categories," *Bioinformatics*, vol. 24, no. 2, pp. 265–271, 2008.

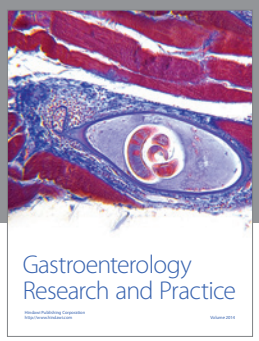
[37] I. P. Gorlov, G. E. Gallick, O. Y. Gorlova, C. Amos, and C. J. Logothetis, "GWAS meets microarray: are the results of genome-wide association studies and gene-expression profiling consistent? Prostate cancer as an example," *PLoS One*, vol. 4, no. 8, Article ID e6511, 2009.

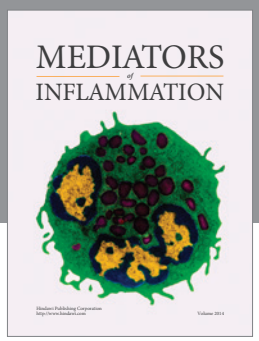
[38] T. Manoli, N. Gretz, H. J. Gröne, M. Kenzelmann, R. Eils, and B. Brors, "Group testing for pathway analysis improves comparability of different microarray datasets," *Bioinformatics*, vol. 22, no. 20, pp. 2500–2506, 2006.

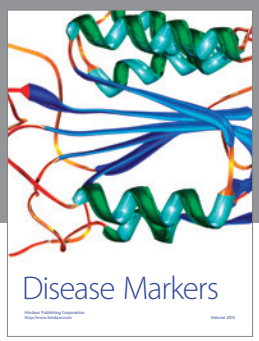
[39] J. Chen, Y. Wang, D. Guo, and B. Shen, "A systems biology perspective on rational design of peptide vaccine against virus infections," *Current Topics in Medicinal Chemistry*, vol. 12, no. 12, pp. 1310–1319, 2012.

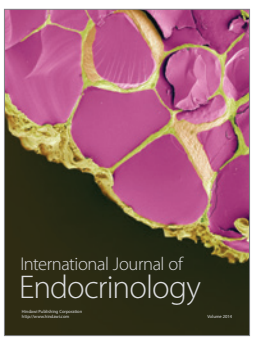
[40] S. Michiels, S. Koscielny, and C. Hill, "Prediction of cancer outcome with microarrays: a multiple random validation strategy," *The Lancet*, vol. 365, no. 9458, pp. 488–492, 2005.

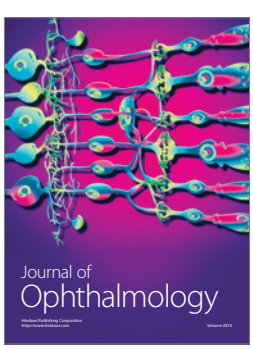
[41] L. Ein-Dor, O. Zuk, and E. Domany, "Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 15, pp. 5923–5928, 2006.

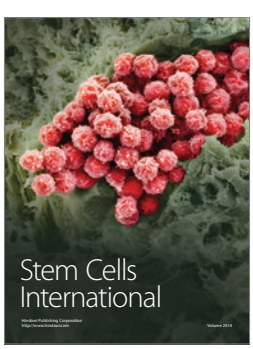


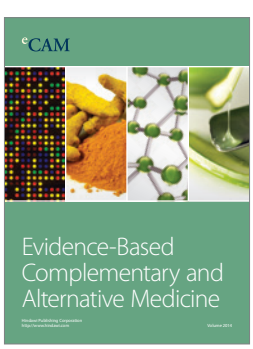


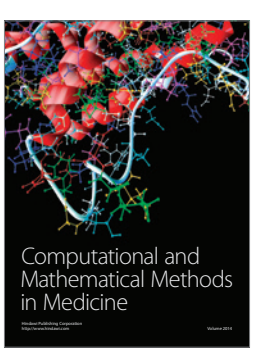


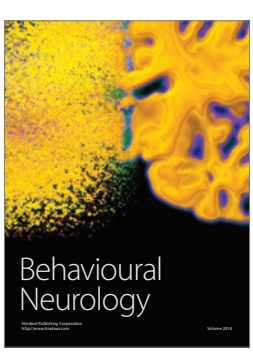


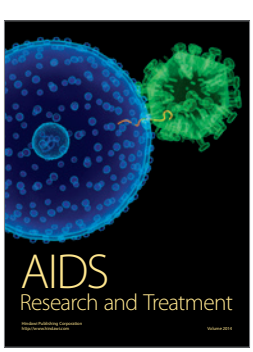


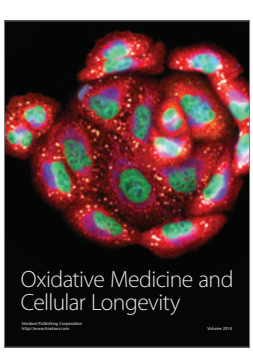












Hindawi

Submit your manuscripts at
http://www.hindawi.com