Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer

Yifei Tang†, Wenying Yan†, Jiajia Chen1, Cheng Luo1,2, Antti Kaipia3, Bairong Shen1*

From Asia Pacific Bioinformatics Network (APBioNet) Twelfth International Conference on Bioinformatics (InCoB2013)
Taicang China. 20-22 September 2013

Abstract

Background: MicroRNAs (miRNAs) are potential regulators that contribute to the pathogenesis of cancer. Microarray technologies have been widely used to characterize aberrant miRNA expression patterns in cancer. Nevertheless, the miRNAs expression signatures identified for a same cancer differs among laboratories due to the cancer heterogeneity. In addition, how the deregulated miRNAs coordinately contribute to the tumourigenic process of prostate cancer remains elusive.

Results: We evaluated five outlier detection algorithms that take into account the heterogeneity of cancer samples. ORT was selected as the best method and applied to four prostate cancer associated microRNA expression datasets. After microRNA target prediction and pathway enrichment mapping, 38 Gene Ontology terms, 16 KEGG pathways and 99 GeneGO pathways are found putative prostate cancer associated. Comparison with our previous studies, we identified two putative novel pathways important in prostate cancer. The two novel pathways are 1) ligand-independent activation of ESR1 and ESR2 and 2) membrane-bound ESR1: interaction with growth factors signalling.

Conclusions: We proved that expression signatures of at the pathway level well address the cancer heterogeneity and are more consistent than at the miRNA/gene levels. Based on this observation, we identified putative novel microRNA regulatory pathways which will help us to elucidate the cooperative function of different microRNAs in prostate cancer.

Background

MicroRNAs (miRNAs) are small non-coding RNAs of approximately 22-nucleotides. They play important roles in gene regulation at post-transcriptional level. They are able to repress the activity of complementary mRNAs by targeting the 3’-untranslated regions [1]. Release 19 of the miRBase database contains more than 2200 mature miRNA sequences for human [2]. Aberrant miRNA expression was shown related to the generation of cancer stem cells and the tumour genesis [3-5].

Microarray-based technologies have routinely been used for profiling molecular expression in cancer. Microarray allows simultaneous expression profiling of tens of thousands of genes in normal versus malignant cells. The growing number of microarray expression datasets has necessitated the integrative analysis approaches to identify significant molecular patterns across multiple datasets.

Many efforts have been made in search of common molecular signatures, however without obvious success. This is partly due to the highly heterogeneous nature of cancer. Tumour samples often comprise of subpopulations with different genomic alterations. However, the most popular outlier detection algorithm, t-test or its
analogue, simply removes heterogeneity between sub-
types, and fail to identify the subgroup-specific gene
alterations [6-8]. Recently novel statistical methods were
developed to identify patterns only existed in the sub-
groups of the studied samples [9-13].

In this study, we applied these outlier detection meth-
ods to analyze our collection of four miRNA expression
microarray datasets to identify differentially expressed
miRNAs (DE-miRNAs). The DE-miRNAs were then
compared among the four data sets at both gene and
gene set (i.e., the functional gene set or pathway) levels
for comparison. By considering the cancer heterogeneity,
we applied different statistical methods to identify the
consistent prostate cancer (PCa) associated pathways
that are coordinately targeted by miRNAs.

Results
Comparison of heterogeneous feature detection
algorithms
Most of the previous expression data studies used fold-
change, t-test and other statistics alike to detect cancer-
related genes. Recently, it has been recognized that
many oncogenes show altered expression in only a small
proportion of cancer samples [11]. Such features will be
removed when using t-test or t-test like methods
because they average gene expression levels in all the
studied samples. Tomlins et al. concluded that t-tests
were not adequate for detecting heterogeneous patterns
of oncogenes [14].

To address this complexity, a series of new heteroge-
nous detection algorithms have been proposed in
recent years. Among these methods are Least Sum of
Ordered Subset Squared (LSOSS) [10], Cancer Outlier
Profile Analysis (COPA) [9], Maximum Ordered Subset
T-statistics (MOST) [11], Outlier Robust T-statistics
(ORT) [13], and Outlier Sum (OS) [12].

The performance of the above algorithms and the
traditional t-test were compared on the detection of the
outliers in our collection of prostate cancer (PCa)
associated microRNA expression data. The outliers
erve refer to the differentially expressed microRNAs
(DE-miRNAs). For all these methods applied to the dif-
ferent data sets with different numbers of samples, we
set the quantile of outliers to 0.05 (5%). Those DE-
miRNAs detected by at least three methods were con-
sidered to be putative PCa associated outliers, and then
the percentages of the putative outliers in the original
result of each method were calculated to determine the
method's accuracy (see Figure 1). In most of the cases,
these heterogeneity feature detection algorithms per-
formed better than the traditional t-test. In most of
this comparison, ORT performed better than the other
methods. For these four studied datasets, ORT had the
biggest median observation and smallest standard
development. Therefore, we take the result by ORT for the
downstream analyses.

The outlier miRNA targets in prostate cancer
As miRNAs play a role in post-transcriptional regulation
by targeting complementary miRNAs, we collection their
putative targets and subsequently mapped these target
genes to pathways or gene sets for enrichment analysis.
Target genes were retrieved from both TargetScan data-
base and our integrative prediction (see methods section
for detail). Additional file 1 shows the target genes of
the PCa associated DE-miRNAs. At last, 1236, 3566,
1520 and 4749 target genes of the DE-miRNAs of four
different datasets were obtained respectively.

The identification of the microRNA regulatory pathways
in prostate cancer
The collection of the four different datasets are from
different platforms, the overlapping of miRNA probes
between these data are about 40–60% while the detected
differently expressed miRNA profiles only have 3% over-
lapping [15]. We aim to identify the consistent pattern
at high level. First, the target genes of DE-miRNAs
found by at least 3 datasets were extracted, then mapped
to function and pathway databases, e.g. GO [16], KEGG
[17,18] and GeneGO (GeneGo, Inc), to identify PCA-
associated functions and pathways. In this process, we
identified 1221 target genes of the PCa associated DE-
miRNAs, among which 253 were shared by all the four
target gene datasets, and 968 overlapped in three of the
four datasets. As shown in Figure 2, the ligand-indepen-
dent activation of ESR1 and ESR2 is the most significant
GeneGO pathway (See Additional file 2 for the notation
of the symbols in this figure). In Figure 2, insulin-like
growth factor-1 (IGF-1) encodes the protein involved in
mediating growth and development. In this pathway,
IGF-1 binds to IGF-1 receptor on the membrane and
activates signal transduction through Shc, SOS, Mek1,
and ERK2, finally mediating the production of ESR1 and
ESR2. Genes involved in the signal transduction above
are all target genes of highly expressed miRNAs in pros-
tate cancer samples; therefore, the expression of ESR1
and ESR2 will be down-regulated which is in accordance
with the previous report by Gamba and his co-authors
[19].

Figure 3 illustrates various biological themes enriched
in the gene list. The left side of the figure is a bar plot
of enriched GO terms, KEGG pathways, and GeneGO
pathways against -log10 (p value); the top five terms of
each biological theme were shown in the right. The
details are also available in Additional files 3, 4, and 5.
In these files, the pathway or GO terms were sorted by
p value. Overall, we identified 38 GO terms (FDR <
0.001), 16 KEGG pathways (p < 0.001), and 99 GeneGO
pathways (FDR < 0.001) that are enriched with target genes of the PCa associated DE-miRNAs.

**Analysis and validation of the putative microRNA regulatory pathways in prostate cancer**

Among the 99 enriched GeneGO pathways, 67 (67.7%) pathways were also significantly enriched in our previous study in which we processed 10 mRNA microarray datasets [20]. In the set of top 15 GeneGO pathways in our previous work, 11 (73.3%) were also detected in the 99 pathways in this study (see Additional file 5).

To identify potential microRNA regulatory pathways in prostate cancer, the 15 most significantly enriched (i.e., with the lowest p value) pathways were chosen for the analysis. Of those, four had previously been reported to be related to prostate cancer in PubMed citations. We verified the other 11 pathways indirectly by analysis of the component genes in PubMed citations although the wet-lab experiments can direct validate them (Table 1). Among the top 15 pathways reported by both the previously and the present studies, 3 pathways are the same in both studies, 2 of the 3 pathways are novel ones i.e., 1) ligand-independent activation of ESR1 and ESR2, this is the most significant pathway we mentioned in the last section, and 2) membrane-bound ESR1: interaction with growth factors signalling.

PubMed citation counts of corresponding genes in each potential pathway can be found in Additional file 6. According to PubMed citation results, the percentages of reported PCa related genes in each pathway range from 25.0% to 71.4%. These percentages will be changed with the PubMed update, since more researches were performed to investigate the cancer hallmarks related pathways, some pathways may be overrepresented in the PubMed database while others may have less citations. The results of PubMed citations indirectly verified the link between the pathways and the prostate cancer, although experimental validation is needed for further confirmation.

**Discussion**

In this study, we collected four prostate cancer miRNA microarray datasets. These datasets were processed with outlier detection statistical methods considering cancer heterogeneity. This is the first work to compare the performance of heterogeneity feature detection statistical methods with real miRNA datasets. The analysis indicates these novel algorithms generally perform better than the t-test. All the methods are important and they may show different performance for different data sets, we could select the best methods based on the consensus analysis.
Figure 3 illustrates the GO terms or pathways (both from KEGG and GeneGO) that are enriched with the overlapped target genes from the PCa DE-miRNAs of the four datasets. The top 5 enriched GO terms are all related to transcription and its regulation, which are in accordance with the observation of the abnormal gene expression in prostate tumours. Most of the identified significant KEGG or GeneGO pathways are important for cancer developing and usually involved in the gene expression or tumour metastasis. Neurotrophins exert their functions by engaging Trk tyrosine kinase receptors or p75 neurotrophin receptor (p75 NTR), a metastasis and tumour suppressor in prostate cancer [21,22]. ESR1 inhibits cell migration and the repression of ESR1 expression enhances cell migration and accelerates tumour formation and metastasis. All the evidence above corroborates our findings in the present study.

The comparison of the previous study [20] with the present one indicates the high consistency between the integrative analysis of the microRNA and the mRNA microarray expression datasets. We here identified 11 novel PCa associated pathways (see Table 1). Two novel pathways among the top 15 in both studies are identified. These overlapping pathways can be potential key pathways contributing to prostate carcinogenesis. Among the key genes in these two novel pathways, histone deacetylases (HDACs) was reported abnormally expressed in prostate cancer [23]. Additionally, the IGF family is involved in the regulation of prostate growth and bone metastasis [24]. In prostate cancer cells, the
IGF-1 receptor, a tyrosine kinase receptor related to tumour progression and metastasis, is highly expressed with MT1-MMP, a metalloproteinase involved in prostate cancer metastasis [25]. Abnormal HIF expression mediates vital processes such as cell survival, proliferation, and angiogenesis [26,27]. Activin A inhibits prostatic branching and growth [28] and enhances prostate cancer cell migration [29]. Additionally, IL15 activates neutrophils and dendritic cells and generates cytotoxic T lymphocytes against cancer cells [30], so the blocking of the IL15 signalling pathway weakens the immune system’s ability to resist cancers. Additional file 6 shows the PubMed citation counts of corresponding genes of each potential pathway in prostate cancer. More wet-lab experiments are suggested to verify the functions of these pathways in prostate cancer.

Conclusions
In this study, heterogeneity feature detection methods were evaluated and applied to the identification of the novel microRNA regulatory pathways in prostate cancer and 11 novel PCa associated pathways were identified. Comparing the present study on PCa microRNA expression data with our previous work on PCa gene expression data, we identified two important novel pathways among the top 15 of the two studies.

Methods
Data collection
We retrieved four miRNA expression datasets from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), which is a public functional genomics data repository supporting MIAME-compliant data
submissions. The datasets were downloaded in single-matrix file format, and named according to the name of the first authors of the references (Table 2). The miRNA probes of these datasets were designed by using Sanger miRBase release 16.0. Because of the diverse platforms of the datasets, a local Blast search [31] was performed by mapping probe sequences to the miRNA precursors of miRBase (release 16.0 [2]) to identify the concordant miRNA names. Figure 4 displays the pipeline of the whole procedure used in this study.

Comparison of detection algorithms and detect the differentially expressed miRNAs

In this study, outliers of microRNA expression in PCa microarray datasets were detected by using six statistical methods: LSOSS, COPA, MOST, ORT, OS and t-test. All these methods were implemented in R packages written by Wang [10] and Lian [11]. The quantile of outlier extraction for all the methods was set to 0.05 (5%) by default.

We compared the performance of the six methods in obtaining the PCa associated DE-miRNAs. We considered the DE-miRNAs detected by at least three methods to be putative outliers. The percentage of these putative outliers in the original result of each method was calculated to measure the method’s accuracy. We selected ORT to be the best method for these PCa microRNA expression datasets considering the consensus analysis results.

Reliable prediction of targets for PCa DE-miRNAs

Targets of DE-miRNAs were retrieved from TargetScan database by a series of in-house Perl scripts. For those miRNAs unavailable in the TargetScan database, the putative targets were manually predicted by performing a genome-wide, sequence-based bioinformatics procedure
with three of the most popular tools, i.e., miRanda [32], RNAhybrid [33], and TargetSpy [34]. Only the overlapped targets of the prediction were kept as reliable result.

**PubMed Search and the citation counts**

PubMed citation count was calculated by searching PubMed in the fields of title and abstract, such as for the “ligand-independent activation of ESR1” pathway, we use “ligand-independent activation of ESR1 [tiab] AND prostate cancer [tiab]” as the search term, and the search term “SP1 [tiab] AND prostate cancer [tiab]” was applied to the search of the link between SP1 gene and prostate cancer. This citation counts may change with the update of PubMed.

**GO and pathway enrichment analysis**

To study the function of the PCa DE-miRNAs, we mapped their target genes to GO, KEGG and GeneGO databases. To decrease the number of the false positives pathways, we first identified target genes shared by at least three PCa DE-miRNAs datasets, which were then mapped to GO, KEGG pathway database by DAVID, and GeneGO pathway database by MetaCore (Gene, Inc.). Both DAVID and MetaCore use hypergeometric distribution to calculate the significance level (i.e. the \( p \) value) for each pathway and adjust it using the FDR value as the threshold. In MetaCore databases, \( p \) value means the probability of a random intersection of two gene sets, with low \( p \) values indicating a high potential of non-randomness of the finding.

**Additional material**

Additional file 1: Entrez IDs and Official Symbols of targeting genes for DE-miRNAs of each datasets

Additional file 2: The notations of all the symbols in Figure 2.
Additional file 3: Enriched GO terms with P-value and FDR. Gene list was generated by extraction of target genes from at least 3 datasets. (FDR < 0.001)

Additional file 4: Enriched KEGG pathways with P-value. Gene list was generated by extraction of targeting genes from at least 3 datasets. (P-value < 0.001)

Additional file 5: Enriched GeneGO pathways with P-value. Gene list was generated by extraction of targeting genes from at least 3 datasets. (FDR<0.001)

Additional file 6: Citation counts of corresponding genes of each potential GeneGo pathway. Citation count was calculated by searching PubMed in the fields of title and abstract, this may change with the update of PubMed.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
YT, JC, and WY carried out the calculation and analysis, AK and CL participated in the discussion of the project and drafted the manuscript; BS conceived the idea and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We gratefully acknowledge financial support from the National Natural Science Foundation of China grants (91230117, 31170795), the Specialized Research Fund for the doctoral Program of Higher Education of China (201132110105), National S&T Cooperation Program of Sichuan (SH201120) and the National High Technology Research and Development Program of China (863 program, Grant No. 2012AA02A601).

Declarations
Publication of this article was funded by the National Natural Science Foundation of China grants (91230117, 31170795). This article has been published as part of BMC Systems Biology Volume 7 Supplement 3, 2013: Twelfth International Conference on Bioinformatics (InCoB2013): Systems Biology. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcsystbiol/supplements/7/s3.

Authors' details
1. Center for Systems Biology, Soochow University, Suzhou 215006, China.
2. Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China.
3. Department of Urology, Tampere University Hospital, 33521 Tamper, Finland.

Published: 16 October 2013

References
1. Bartel DP: MicroRNAs: target recognition and regulatory functions. Cell 2009, 136:215-233.

2. Koizumia A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 2010, 39:D152-157.

3. Lin SL, Chang DC, Ying SY, Lee D, Wu DT: MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells, at coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways. Cancer Research 2010, 70:947-9482.

4. Liu C, Kelnar K, Liu B, Chen X, Calhoun-Davis T, Li H, Patrawala L, Yan H, Jeter C, Honorio S, et al: The microRNA miR-34a inhibits prostate cancer stem cell metabolism and metastasis by directly repressing CD44. Nat Med 2011, 17:211-215.

5. Mallick B, Chakraborti J, Ghosh Z: MicroRNA reis in embryonic and cancer stem cells. RNA Biol 2011, 8(3):415-426.

6. Fisher R, Puziatli L, Swanton C: Cancer heterogeneity: implications for targeted therapeutics. British journal of cancer 2013, 108:1479-1485.

7. Samuel N, Hudson T J: Translating genomics to the clinic: implications of cancer heterogeneity. Clinical chemistry 2013, 59:127-137.

8. Wang ZA, Mitrofanova A, Bergen SK, Abate-Shen C, Cardiff RD, Califano A, Shen MM: Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. Nature cell biology 2013, 15:274-283.

9. MacDonalki JW, Gaoa D: COPA-cancer outlier profile analysis. Bioinformatics 2006, 22:2950-2951.

10. Wang Y, Rekaya R: LSOS: Detection of Cancer Outlier Differential Gene Expression. Biomarks Insights 2010, 5:69-78.

11. Lian H: MOST: detecting cancer differential gene expression. Biostatistics 2008, 9:41-44.

12. Tibshirani R, Hastie T: Outlier sums for differential gene expression analysis. Biostatistics 2007, 8:2-8.

13. Wu B: Cancer outlier differential gene expression detection. Biostatistics 2007, 8:566-575.

14. Tomlin SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005, 310:644-648.

15. Tang Y, Chen J, Luo C, Kalia P, Shen B: MicroRNA Expression Analysis Reveals Significant Biological Pathways in Human Prostate Cancer. IEEE International Conference on Systems Biology 2011, 2-4 Sept. 2011 Zhuji 2011, 203-210.

16. Aslburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000, 25:29-28.

17. Kanehsa M, Goto S, Furumichi M, Tanabe M, Hirakawa M: KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010, 38:D355-360.

18. Kanehsa M, Araki M, Goto S, Hattori M, Hirakawa M, Hoh M, Katayama T, Kawasaki S, Okuda S, Tomikata T, Yamanishi Y: KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008, 36:D480-484.

19. Gamba L, Cubedo N, Gysen A, Gamba D. Cytogenetics. Cytogenet Cytogenet 2010, 9:479-485.

20. Wu B: Cancer outlier differential gene expression detection. Biostatistics 2007, 8:566-575.

21. Kanehsa M, Goto S, Furumichi M, Tanabe M, Hirakawa M: KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 2010, 38:D355-360.

22. Wu B: Cancer outlier differential gene expression detection. Biostatistics 2007, 8:566-575.

23. Tomlin SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005, 310:644-648.

24. Gamba L, Cubedo N, Ghysen A, Gamba D. Cytogenetics. Cytogenet Cytogenet 2010, 9:479-485.

25. Wu B: Cancer outlier differential gene expression detection. Biostatistics 2007, 8:566-575.

26. Tomlin SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005, 310:644-648.

27. Shin J, Lee HJ, Jung DB, Jung JH, Lee EO, Lee SG, Shim BS, Choi SH, Ko SG, Lee JH, et al: MOST: detecting cancer differential gene expression. Bioinformatics 2008, 24:411-418.

28. Wu B: Cancer outlier differential gene expression detection. Biostatistics 2007, 8:566-575.

29. Tomlin SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005, 310:644-648.

30. Tomlin SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al: Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 2005, 310:644-648.
generates cytotoxic T lymphocytes against cancer cells. *Cell Immunol* 2011, 271:89-96.

31. Camacho C, Couteuris GS, Avagyan V, Ma N, Papadopoulou J, Bealer K, Madden TL: BLAST+: architecture and applications. *BMC Bioinformatics* 2009, 10:421.

32. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: Human microRNA targets. *PloS Biol* 2004, 2:e363.

33. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R: Fast and effective prediction of microRNA/target duplexes. *RNA* 2004, 10:1507-1517.

34. Sturm M, Hackenberg M, Langenberger D, Frishman D: TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics* 2010, 11:292.

35. Ambs S, Prueitt RL, Yi M, Hudson RS, Howe TM, Petrocca F, Wallace TA, Liu CC, Volinia S, Calin GA, et al: Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer Res* 2008, 68:6162-6170.

36. Schaefer A, Jung M, Mollenkopf HJ, Wagner I, Stephan C, Jentzmik F, Miller K, Lein M, Kristiansen G, Jung K: Diagnostic and prognostic implications of microRNA profiling in prostate carcinoma. *Int J Cancer* 2010, 126:1166-1176.

37. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al: Integrative genomic profiling of human prostate cancer. *Cancer Cell* 2010, 18:11-22.

38. Wach S, Nolte E, Saczyba J, Stohr R, Hartmann A, Orntoft T, Dynskot L, Eltze E, Wieland W, Keck B, et al: MiRNA profiles of prostate carcinoma detected by multi-platform miRNA screening. *Int J Cancer* 2012, 130(3):611-621.

Cite this article as: Tang et al.: Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer. *BMC Systems Biology* 2013 7(Suppl 3):S6.