METHOD

Correlating Bladder Cancer Risk Genes with Their Targeting MicroRNAs Using MMiRNA-Tar

Yang Liu 4,a, Steve Baker 3,b, Hui Jiang 5,c, Gary Stuart 1,2,d, Yongsheng Bai 1,2,*,e

1 Department of Biology, Indiana State University, Terre Haute, IN 47809, USA
2 The Center for Genomic Advocacy, Indiana State University, Terre Haute, IN 47809, USA
3 Department of Math and Computer Science, Indiana State University, Terre Haute, IN 47809, USA
4 Department of Electrical and Computer Engineering, Rose-Hulman Institute of Technology, Terre Haute, IN 47803, USA
5 Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Received 8 May 2015; accepted 27 May 2015
Available online 10 July 2015
Handled by Luonan Chen

Abstract The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov) is a valuable data resource focused on an increasing number of well-characterized cancer genomes. In part, TCGA provides detailed information about cancer-dependent gene expression changes, including changes in the expression of transcription-regulating microRNAs. We developed a web interface tool MMiRNA-Tar (http://bioinf1.indstate.edu/MMiRNA-Tar) that can calculate and plot the correlation of expression for mRNA–microRNA pairs across samples or over a time course for a list of pairs under different prediction confidence cutoff criteria. Prediction confidence was established by requiring that the proposed mRNA–microRNA pair appears in at least one of three target prediction databases: TargetProfiler, TargetScan, or miRanda. We have tested our MMiRNA-Tar tool through analyzing 53 tumor and 11 normal samples of bladder urothelial carcinoma (BLCA) datasets obtained from TCGA and identified 204 microRNAs. These microRNAs were correlated with the mRNAs of five previously-reported bladder cancer risk genes and these selected pairs exhibited correlations in opposite direction between the tumor and normal samples based on the customized cutoff criterion of prediction. Furthermore, we have identified additional 496 genes (830 pairs) potentially targeted by 79 significant microRNAs out of 204 using three cutoff criteria, i.e.,
false discovery rate (FDR) < 0.1, opposite correlation coefficient between the tumor and normal samples, and predicted by at least one of three target prediction databases. Therefore, MMiRNA-Tar provides researchers a convenient tool to visualize the co-relationship between microRNAs and mRNAs and to predict their targeting relationship. We believe that correlating expression profiles for microRNAs and mRNAs offers a complementary approach for elucidating their interactions.

### Methods

#### Data source

The test datasets were downloaded from TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/). The type of cancer studied in this paper is bladder urothelial carcinoma (BLCA). Illumina HiSeq data were acquired based on the availability of expression profile for both miRNA, which was produced by Baylor College Human Genome Sequencing Center (BCGSC), and mRNA, which was produced by University of North Carolina at Chapel Hill (UNC). Specifically, TCGA level 3 miRNASeq data were produced on Illumina HiSeq 2000 sequencers and its public release date is 04/30/2012. Read counts and reads per kilobase per million (RPKM) per composite gene (UCSC genes Dec 2009 build) were calculated using the SeqWare framework via the RNASeqAlignmentBWA workflow (http://seqware.sourceforge.net). The miRNA analyses of TCGA level 3 BLCA samples were produced by Illumina HiSeq as well. Normalized expression per miRNA gene (Reads per million miRNA mapped or RPM) was reported as miRNAs expression measurement unit. The public release date of miRNA data used in this study is 10/09/2014. To make measurement units between two sequencing data sets consistent, we converted RPKM expression values for mRNA samples into transcripts per million (TPM) values. A total of 53 tumor and 11 normal samples from seven batches (batch No. 86, 113, 128, 150, 170, 175, and 192) were downloaded and tested for both miRNA and mRNA data. The normalized mRNA and miRNA expression data of both the tumor and normal samples are shown in Tables S1 and S2, respectively.

#### Data pre-processing

Expression profiles of BLCA datasets for a total of 20,532 mRNAs were downloaded. We excluded 29 genes that do not have their gene symbols available (gene names marked as “?” in the annotation table) from the list. We also excluded SLC35E2 because it is doubly reported. Thus, a total of 20,501 genes were used to check against a miRNA expression file, in which 1046 miRNAs were available.

#### Correlation coefficient calculation and target prediction

Calculations of linear (positive) or inverse (negative) correlation (Pearson correlation) for each miRNA–mRNA pair across samples and the prediction of miRNA and mRNA target relationship were implemented in C language. All three...
databases including TargetProfiler [20], TargetScan [21], and miRanda [22] were precompiled for the search of targeting relationship between miRNA and mRNA. We claimed the existence of the targeting relationship if a target prediction outcome is supported by at least one of the three databases mentioned above. The FDR multiple testing [23] control and normalization steps were implemented using a customized R script.

Figure 1 shows the workflow of selecting potential targeting miRNAs and additional targeted genes. *MMiRNA-Tar* is available at [http://bioinf1.indstate.edu/MMiRNA-Tar](http://bioinf1.indstate.edu/MMiRNA-Tar) and the software source code is freely available upon request for non-commercial purposes.

**Results**

**Correlation of expression profiles of miRNAs and mRNAs**

We took five genes that have been reported as common bladder cancer risk genes in multiple studies and the National Institutes of Health (NIH) Genetic Home Reference website ([http://ghr.nlm.nih.gov/condition/bladder-cancer](http://ghr.nlm.nih.gov/condition/bladder-cancer)) and set out to identify their potential targeting miRNAs using three popular target prediction databases mentioned in the Method section. These genes include *FGFR3*, *HRAS*, *RB1*, *TSC1*, and *TP53*. We calculated correlations (Pearson correlation) between each of the five genes and all miRNAs reported in 53 tumor and 11 normal samples from the aforementioned TCGA datasets. We then selected the pairs with correlation values in opposite directions between the tumor and normal samples and with targeting relationship predicted by at least one of three databases using *MMiRNA-Tar*. As shown in Figure 2, three prediction databases showed similar density distribution patterns for calculated correlation values in the tumor samples, although the density distribution by miRanda was slightly different from the other two in the normal samples. We concluded that requiring a prediction outcome from any of these databases would be reasonable.

Using these five genes, 204 miRNAs in total were obtained based on the cutoff criteria of opposite correlation direction between the tumor and normal samples and by at least one database prediction ([Table 1](#) and [Table S3](#)). These 204 miRNAs are presumed to have targeting relationships with five bladder cancer risk genes. The expression information in heatmap format for 204 miRNAs (259 pairs) across 53 tumor and 11 normal samples is shown in Figure S1. We noticed that miRNAs targeting the same gene(s) were often grouped together using hierarchical clustering with the Pearson correlation distance measure method of multiple array viewer ([http://sourceforge.net/projects/mev-tm4/](http://sourceforge.net/projects/mev-tm4/)).

**The expression profile correlation analysis for 79 selected miRNAs and their targeting mRNAs**

We then calculated correlations and predicted targeting relationships for all possible pair combinations of 204 miRNAs and 20,501 mRNAs in 53 tumor and 11 normal samples of BLCA data. We obtained 830 additional miRNA–mRNA pairs (comprising of 79 miRNAs and 496 genes) showing opposite correlated relationships between the tumor and normal samples and having at least one database prediction outcome with FDR < 0.1. Figure 3 is a Venn diagram showing...
prediction results derived by applying the three target prediction database filters. The additional list of miRNA-gene target pairs, along with their correlation values and target prediction result using the aforementioned cutoff criteria, is shown in Table S4. We noticed, among the 830 pairs, half of the genes seem to have targeting relationships with at least two of the 79 identified miRNAs. Thus, in addition to the five initial genes, we obtained another 496 genes having at least one predicted targeting relationship with 79 selected miRNAs.

**Gene functional enrichment analysis**

We searched the Database for Annotation, Visualization and Integrated Discovery (DAVID) [24,25] for functional information about the 496 genes with their predicted targeting miRNAs identified above. Enrichment of these genes was found in several GO biological processes. Some of genes are involved in chromatin remodeling complex, some of genes are associated with cell cycle regulation, and some genes are involved in protein kinase signaling pathways. These biological processes (cell cycle regulation, kinase signaling, chromatin remodeling) are frequently dysregulated in bladder cancer [26]. Genes associated with aforementioned biological processes and their associated GO terms are shown in Table 2.

**Discussion**

In this study, we computed the correlation coefficients for all available combinations of miRNA and mRNA pairs using TCGA BLCA sequencing data. Performing multivariable correlation analysis on a genome scale would be our future research strategy. Under the assumption of an opposite correlation of miRNA and mRNA (gene) expression levels between the tumor and normal samples as an indicator for the miRNA–mRNA target relationship, we used five previously reported bladder cancer risk genes to obtain a list of 204 potential targeting miRNAs by applying several state-of-the-art target prediction algorithms. We then used this list of miRNAs to identify other potential targeted pairs (genes), which could be bladder cancer risk candidate genes, and perform GO functional analysis on these genes. Fewer pairs with negative correlation coefficients between the tumor samples and normal samples were then predicted as potential target pairs for 79 selected miRNAs.
correlation were reported in tumor samples than in normal samples, suggesting that these miRNAs possibly lose their functions in tumor samples, under the assumption that miRNAs often anti-correlate with their gene targets.

Target prediction tools employed in our study for predicting miRNA targets likely contain false positives since the intersection of the predictions by TargetProfiler, TargetScan, and miRanda are low (Figure 3). In our effort, to identify more targets, further analysis with at least one prediction selection criteria was performed.

Conclusion

We have developed a web-based tool, MMiRNA-Tar, to plot the correlation relationships and to report target prediction outcomes between miRNAs and mRNAs across multiple samples and time course data. We used the complete TCGA BLCA dataset currently available to test the tool and identified 204 potential targeting miRNAs and many additional targeted genes by 79 selected miRNAs. We believe our tool is the first to utilize miRNA and mRNA correlation plotting combined with multiple target prediction tools for the analysis of miRNA contributions to transcription regulation in cancer. Although the current work was limited to BLCA, the tool developed in this study should also be valuable for studies of functional miRNAs for other cancer datasets as well. The future work will be extended to enhance our web-based tool by incorporating the functionality of matching seed regions of miRNA to the mRNA targets. We would also like to incorporate other

Table 2 Predicted target genes along with their associated GO terms enriched

| Gene                          | GO ID     | Biological process                        |
|-------------------------------|-----------|-------------------------------------------|
| SHPRH, RSFI, MLL, NAP1L1, WRN, MLH3, SIRT1, TAF5L, HUWE1, BRPF3, SUPT16H, PHF21A, KDM3B, PARP1, USP16, MYSM1, RERE, EP400, APC | 0051276   | Chromosome organization                    |
| TAF5L, HUWE1, BRPF3, USP16, SIRT1, MYSM1, EP400 | 0016570   | Histone modification                       |
| TAF5L, MLL, RSFI, HUWEI, BRPF3, PHF21A, KDM3B, USP16, SIRT1, MYSM1, EP400, RERE | 0016568   | Chromatin modification                      |
| TAF5L, HUWE1, BRPF3, USP16, SIRT1, MYSM1, EP400 | 0016569   | Covalent chromatin modification            |
| SHPRH, RSFI, MLL, NAP1L1, SIRT1, TAF5L, BRPF3, HUWE1, SUPT16H, PHF21A, KDM3B, USP16, RERE, EP400, MYSM1 | 006325     | Chromatin organization                      |
| UHRF2, ZAK, SMAD3, PPP1CB, PTPN11, APC | 0051726   | Regulation of cell cycle                   |
| BCAT1, TAF1, MLL, ZAK, SMAD3, MLH3, PPP1CB, TACC1, JMY, CUL5, PSMC6, UHRF2, HSPA2, CASP8AP2, MAPK4, PTEN1, TUBE1, TNKS, MAPRE2, MAPRE1, USP16, DST, APC | 0000749   | Cell cycle                                 |
| BCAT1, TAF1, ZAK, SMAD3, MLH3, PPP1CB, JMY, CUL5, PSMC6, HSPA2, TUBE1, MAPRE2, TNKS, MAPRE1, USP16, DST, APC | 0022402   | Cell cycle process                         |
| BCAT1, TAF1, CUL5, PPP1CB | 0000082   | G1/S transition of mitotic cell cycle      |
| BCAT1, TAF1, CUL5, PPP1CB | 0051329   | Interphase of mitotic cell cycle           |
| BCAT1, TAF1, PSMC6, CUL5, TNKS, MAPRE2, MAPRE1, USP16, PPP1CB, APC | 0051325   | Interphase                                  |
| BCAT1, TAF1, CUL5, HSPA2, TNKS, MAPRE2, MAPRE1, MLH3, USP16, PPP1CB, APC | 0000278   | Mitotic cell cycle                          |
| TNX1, ZAK, MAPK8IP1 | 0022403   | Cell cycle phase                           |
| PHIP, UTP11L, EPHA7, GRB10, BAIAP2, SOCS7, RAF1, SOCS5, PTPN11 | 0031098   | Stress-activated protein kinase signaling pathway |
| TWSG1, SMAD9, ID1, SMAD5, SMAD3 | 0007169 | Transmembrane receptor protein tyrosine kinase signaling pathway |
|                              | 0007178   | Transmembrane receptor protein serine/threonine kinase signaling pathway |
available TCGA cancer datasets and identify interesting signatures of miRNA–mRNA pairs for other datasets as well. We also plan to develop a visualization tool to present the relationships between miRNAs and mRNAs for comparing tumor and normal expression data sets.

Authors’ contributions

YL designed the web interface and deployed the software on the server. SB wrote C code to perform pairwise correlation calculation and target database prediction. HJ wrote R code for statistical filtering and offered statistical advice. GS was involved in interpreting findings and critically reviewing this manuscript for content and accuracy. YB designed and supervised the project, performed the analysis, provided biological interpretation, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

The results published here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. This research was supported by the startup funds of Indiana State University, USA to YB. We thank Cameron Meyer and Joshua Stolz for helping with data preparation. We also thank Norman Miller for offering manuscript editing help.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2015.05.003.

References

[1] Ambros V. MicroRNAs: tiny regulators with great potential. Cell 2001;107:823–6.
[2] Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 2002;21:4663–70.
[3] Ruike Y, Ichimura A, Tsuchiya S, Shimizu K, Kunimoto R, Okuno Y, et al. Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines. J Hum Genet 2008;53:515–23.
[4] Wang YP, Li KB. Correlation of expression profiles between microRNAs and mRNA targets using NCI-60 data. BMC Genomics 2009;10:218.
[5] He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 2004;5:522–31.
[6] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116:281–97.
[7] Nunez YO, Truitt JM, Gorini G, Ponomareva ON, Blednov YA, Harris RA, et al. Positively correlated miRNA–mRNA regulatory networks in mouse frontal cortex during early stages of alcohol dependence. BMC Genomics 2013;14:725.
[8] Miles GD, Seiler M, Rodriguez L, Rajagopal G, Bhanot G. Identifying microRNA/mRNA dysregulations in ovarian cancer. BMC Res Notes 2012;5:164.
[9] van Rooij E, Sutherland LB, Liu N, Williams AH, McNally J, Gerard RD, et al. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci U S A 2006;103:18255–60.
[10] Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 2004;64:3753–6.
[11] Iorio MV, Ferracin M, Liu CG, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065–70.
[12] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
[13] Bertz S, Abe C, Schwartz-Furlan S, Aller J, Hofstadter F, Stoehr R, et al. Increased angiogenesis and FGFR protein expression indicate a favourable prognosis in bladder cancer. Virochows Arch 2014;465:687–95.
[14] Beukers W, Hercegovac A, Zwarthoff EC. HRAS mutations in bladder cancer at an early age and the possible association with the Costello Syndrome. Eur J Hum Genet 2014;22:837–9.
[15] Malekzadeh K, Sobti RC, Nikbakht M, Shekari M, Hosseini SA, Tandamani DK, et al. Methylation patterns of Rb1 and Casp-8 promoters and their impact on their expression in bladder cancer. Cancer Invest 2009;27:70–80.
[16] Guo Y, Chekaluk Y, Zhang J, Du J, Gray NS, Wu CL, et al. TSC1 involvement in bladder cancer: diverse effects and therapeutic implications. J Pathol 2013;230:17–27.
[17] Smal MP, Rolevich AJ, Polyakov SL, Krasny SA, Goncharova R, FGFR3 and TP53 mutations in a prospective cohort of Belarusian bladder cancer patients. Exp Oncol 2014;36:246–51.
[18] Groebeil PJ, Knowles MA. Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. Urol Oncol 2010;28:409–28.
[19] Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011;39:D945–50.
[20] Oulas A, Karathanasis N, Louloupis I, Iliopoulos I, Kalantidis K, Poirazi P. A new microRNA target prediction tool identifies a novel interaction of a putative miRNA with CCND2. RNA Biol 2012;9:1196–207.
[21] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.
[22] John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. PLoS Biol 2004;2:e63.
[23] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
[24] Huang D, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
[25] Huang D, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucelic Acids Res 2009;37:1–13.
[26] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature 2014;507:315–22.