Genomics Data 2 (2014) 320–324

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Genome-wide mRNA and miRNA expression data analysis to screen for markers involved in sarcomagenesis in human chondrosarcoma cell lines

Biju Issac [a,*], Karina Galoian [b], Toumy Guettouche [c], Loida Navarro [d], H. Thomas Temple [e]

[a] Biostatistics and Bioinformatics Core/Division of Bioinformatics, Sylvester comprehensive Cancer Center, University of Miami, 1550 NW 10th Ave, Fox 304A, Miami, FL 33136, USA
[b] Department of Orthopaedic Surgery, Miller School of Medicine, University of Miami, 1600 NW 10th Ave, S 8012, Miami, FL 33136, USA
[c] Children's Hospital of Philadelphia, Center of Applied Genomics, 3615 Civic center Blvd, Philadelphia, PA 19104-4318, USA
[d] Oncogenomics Core, Sylvester Comprehensive Cancer Center, University of Miami, 1501 NW 10th Ave, BRB 542 B.02, Miami FL 33136, USA
[e] Department of Orthopaedic Surgery and University of Miami Tissue Bank Division, Miller School of Medicine, University of Miami, 1400 NW 12th Ave, UMH, East Bldg 4036, Miami, FL 33136, USA

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form 30 September 2014
Accepted 1 October 2014
Available online 13 October 2014

Keywords:
Human chrondrosarcoma
mRNA expression
miRNAs
Microarray
Bioinformatics

ABSTRACT

Genes and miRNAs involved in sarcomagenesis related pathways are unknown and therefore signaling events leading to mesenchymal cell transformation to sarcoma are poorly elucidated. Exiqon and Illumina microarray study on human chondrosarcoma JJ012 and chondrocytes C28 cell lines to compare and analyze the differentially expressed miRNAs and their gene targets was recently published in the Journal *Tumor Biology* in 2014. Here we describe in details the contents and quality controls for the miRNA and gene expression data associated with the study that is relevant to this dataset.

Specifications	
Organism/cell line/tissue	Human Chondrosarcoma JJ012 and Chondrocytes C28 Cell lines
Sex	Not applicable
Sequencer or array type	Exiqon miRNA array and Illumina Human HT-12 Gene expression array
Data format	Raw
Experimental factors	Tumor vs normal
Experimental features	Whole genome coverage transcriptional coverage of characterized and uncharacterized RNAs using arrays
Consent	Not applicable
Sample source location	Miami, Florida, USA

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48420.

Experimental design, materials and methods

Study cell line data

Human chondrosarcoma (JJ012) and chrondrocytes (C28) cell lines were obtained from the University of Miami Tissue bank and cultured in monolayer until they were semiconfluent as described previously [1,2]. Cells were lysed and 100 μg of total RNA from the sample was prepared using the Qiagen RNeasy mini kit. NanoDrop 8000 spectrometer was used to check for RNA integrity and the presence of small RNA fractions was determined using Agilent Bioanalyzer 2100.

miRNA expression data

20 ng of total RNA per panel was reverse transcribed using the miRCURY LNA™ Universal RT microRNA PCR and polyadenylation and cDNA synthesis kit (Exiqon, Woburn, MA). For each sample, the cDNA was diluted, and each reaction was combined with SYBR Green Master Mix (Exiqon) and added to the Ready-to-Use PCR panels. Human microRNA Ready-to-Use PCR panels I and II hold 743 different miRNA targets and six reference gene assays (mirBase13). They were tested on a Roche LightCycler 480 real-time PCR instrument.

Abbreviations: SD, standard deviation; PCA, principal component analysis.

* Corresponding author. Tel.: +1 305 243 6741; fax: +1 305 243 9342.
E-mail address: bissac@med.miami.edu (B. Issac).

http://dx.doi.org/10.1016/j.gdata.2014.10.001

Quality control

We used the GenEx version 5 software from MultiD Analyses AB to analyze the RT-PCR data of the human miRNAs. Layout files for human panels I and II version 2 with miRBase 16 annotations was loaded on the software. miRNA expression data was imported and sample and group classification columns were appropriately modified to define the replicates. The groups were labeled as 'SARC' for human chondrosarcoma and 'C28' for chondrocytes. The data was pre-processed as provided by the software guidelines. Interplate calibration was performed using mean values between plates. Outlier data points were deleted from analysis using Grubb's test on replicates with default confidence level (0.95) and cut-off SD (0.25). All values which were larger than 37 (default) were replaced with a blank and all empty rows and undetected miRNAs were removed. Missing 'NaN' values were also replaced and the data was validated to remove data points for miRNAs which had less than 75% values from replicates within a group. All remaining missing values were now imputed based on groups and empty cells were filled with a value of 38 as recommended.

Normalization

As recommended by the software user guide for screening studies, variability arising from differences in sample handling was minimized using global mean normalization with mean expression of all genes. Markers with Cq larger than 34 were not included as directed. Replicates were averaged and values were converted to relative quantities on the linear scale prior to log2 conversion.

Basic analysis

The processed miRNA data was analyzed using hierarchical clustering (Fig. 1) and principal component analysis (Fig. 2) methods to visually interpret the differential expression between groups. The groups were noted to have sufficient differences for a more robust statistical significance analysis. The C28 and SARC were compared using a 2-tailed unpaired *T*-test and the resulting P-values were corrected for false discovery using the well documented Benjamini–Hochberg method (Fig. 3).

Gene expression data

Expression data was obtained by hybridizing extracted RNA from samples to HumanHT-12 V4_0_R1 array platform from Illumina. Care was taken to obtain RNA aliquots of same harvested cells as with miRNA expression data.

Quality control

The mRNA expression data were loaded on GeneSpring® 12.5 GX (Agilent Technologies, Santa Clara, CA, USA). The data for all probes including those with missing values were included in the preprocessing steps. Quality assessments were done visually using various plots including PCA (Fig. 4). All samples were found to be of high quality with intra-group correlation of expression data within 98–99% match.

Normalization

Normalization of the mRNA expression data was performed using the 'Quantile' normalization method available in GeneSpring and the data was log2 transformed to the mean of all samples before statistical analyses.

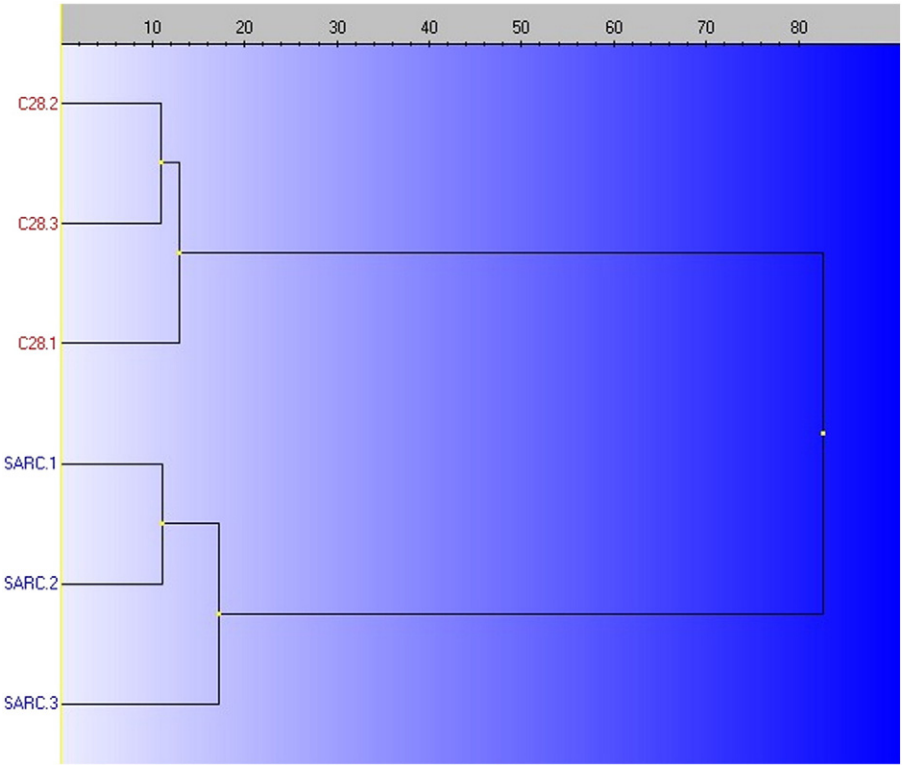

Fig. 1. Hierarchical clustering of miRNA expression data from human chondrosarcomas JJ012 (SARC) and human chondrocytes (C28) cell lines.

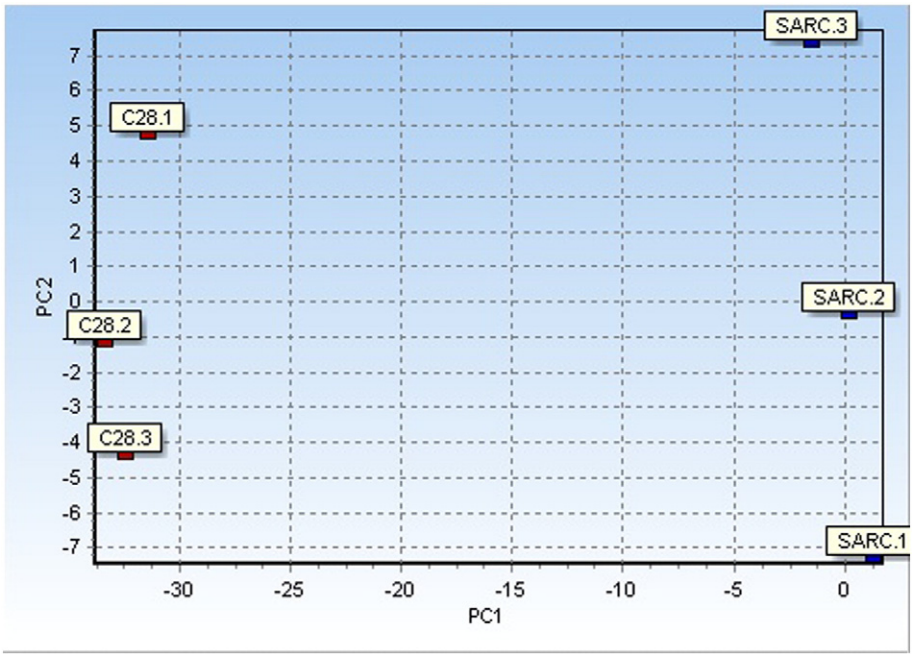

Fig. 2. Principal component analysis (PCA) plot on miRNA expression data from human chondrosarcoma (SARC) and chondrocytes (C28) using the first two components.

Basic analysis

Since one objective of the study was to cross-compare significant miRNA expression within groups to their target mRNA expression, statistical significance test was performed between human chondrosarcoma (SARC) and chrondrocytes (C28) as with miRNA expression data. Multiple testing corrections on nominal P-values were performed using Benjamini–Hochberg method (Fig. 5). Predicted and validated gene targets for miRNAs were obtained from miRWalk database (http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/index.html). Gene expression data for targets of significant miRNAs were matched from the differential mRNA expression obtained above.

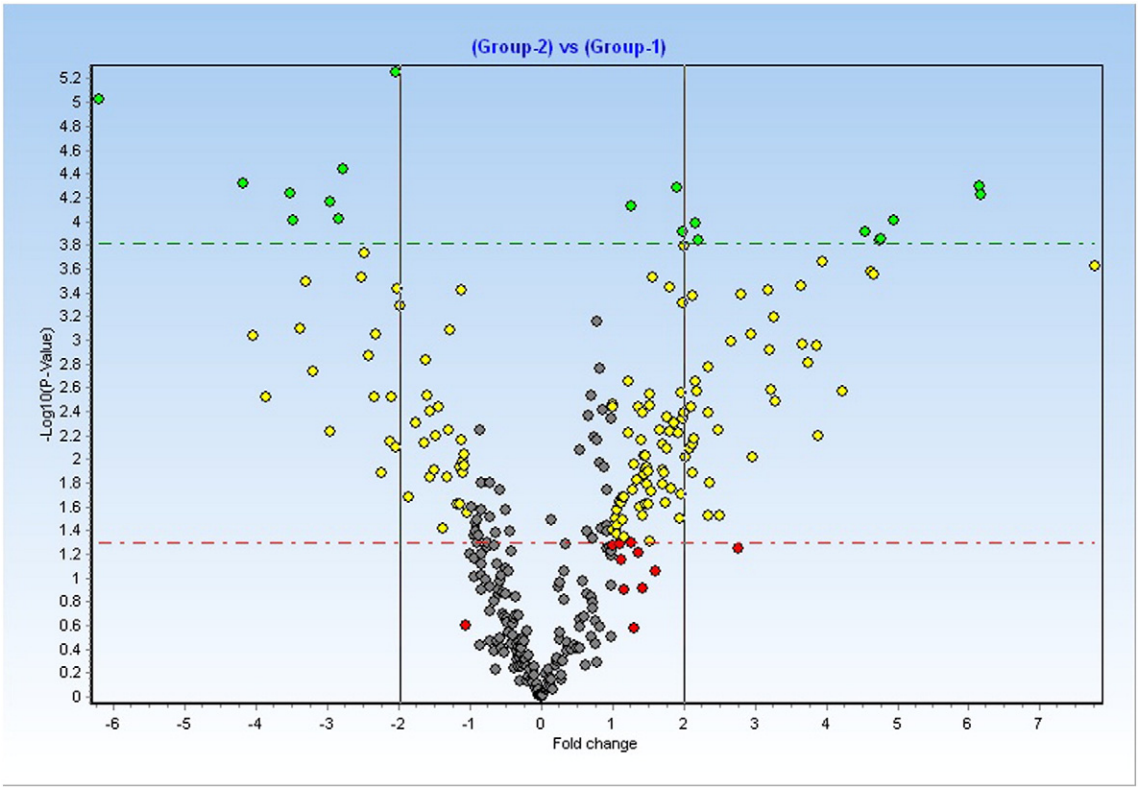

Fig. 3. miRNAs with statistical significance after a *T*-test are shown in green on a volcano plot above. Those with marginal significance are shown in yellow, while those with low or no significance are shown in red and gray respectively.

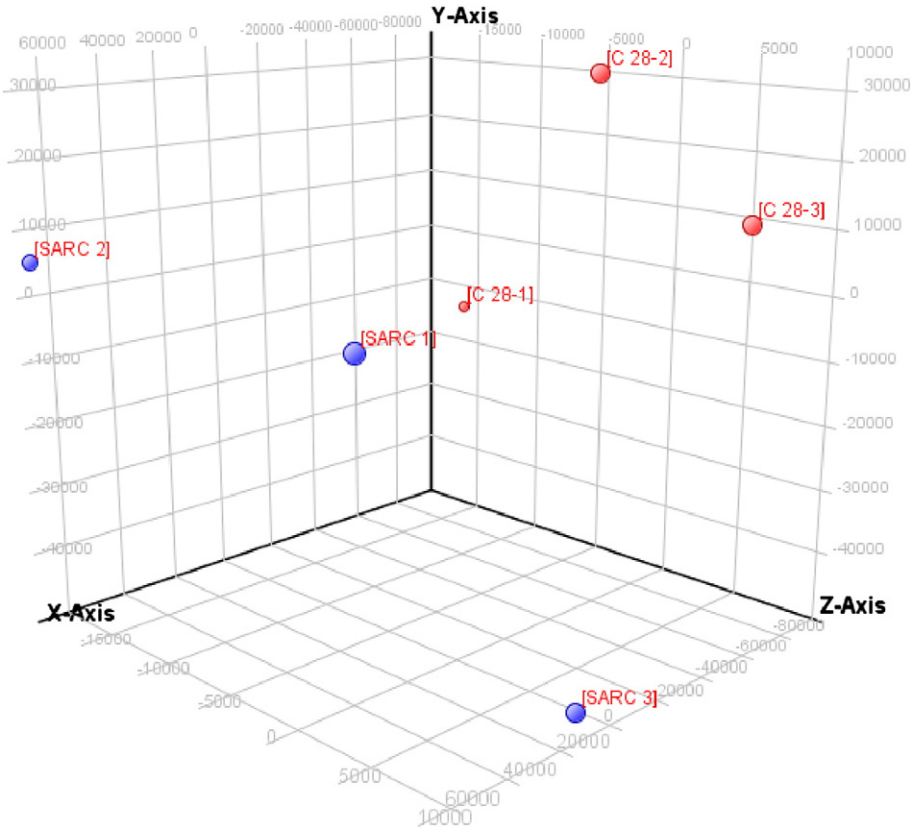

Fig. 4. Principal component analysis (PCA) plot using first 3 components on mRNA expression data from human chondrosarcoma (SARC) in blue and human chondrocytes (C28) in red.

Discussion

We describe here a unique miRNA and mRNA expression dataset from human chondrosarcoma (JJ012) and chondrocytes (C28) cell lines. This dataset comprises only array data from Exiqon miRNA array and Illumina mRNA array. The expression data submitted here is of highest quality and has been used in studies published recently [3] in journals with wide subscription and impact, therefore would benefit wider biomedical research through further sharing.

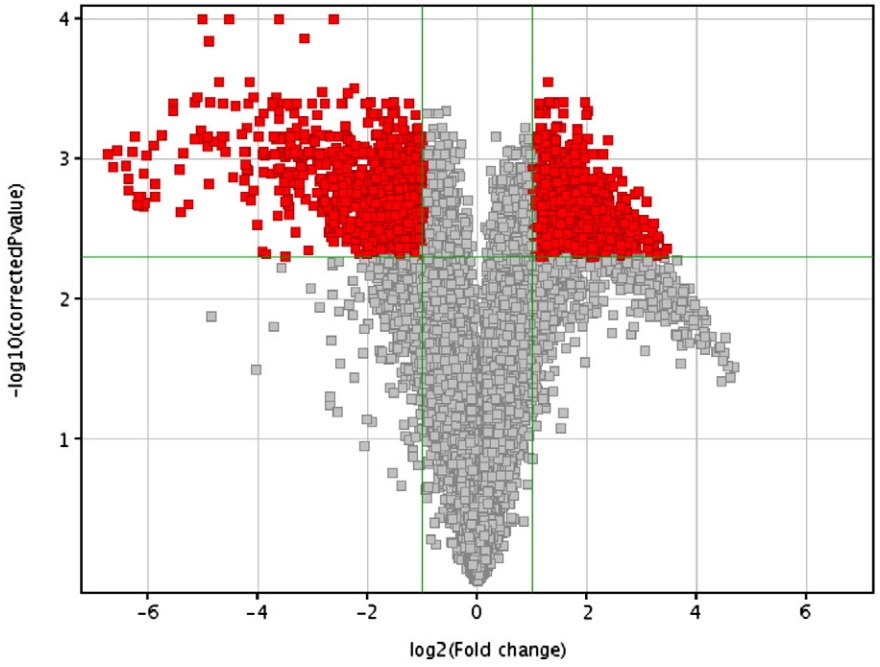

Fig. 5. Volcano plot showing significantly (fold change > 1.5 and P-values < 0.05) differentially expressed genes in red after an unsupervised *T*-test on human chondrosarcoma (SARC) and chondrocytes (C28).

Conflict of interests

None.

Acknowledgments

The study was financially supported by Sylvester Comprehensive Cancer Center (SCCC) and University of Miami Tissue Bank.

References

[1] K.A. Galoian, H.T. Temple, A.A. Galoyan, mTORC1 inhibition and ECM–cell adhesion-independent drug resistance *via* PI3K–AKT and PI3K–RAS–MAPK feedback loops. Tumour Biol. 33 (3) (2012) 885–890.
[2] K.A. Galoian, N. Garamszegi, S. Garamszegi, S. Scully, Molecular mechanism of tenascin-C action on matrix metalloproteinase-1 invasive potential. Exp. Biol. Med. 232 (4) (2007) 515–522.
[3] K.A. Galoian, T. Guettouche, B. Issac, L. Navarro, Lost miRNA surveillance of Notch, IGFR pathway—road to sarcomagenesis. Tumor Biol. 35 (2014) 483–492.