



Data in Brief

Gene expression profiling of human fibrocytic myeloid-derived suppressor cells (f-MDSCs)



Emilia Maria Cristina Mazza^a, Alessia Zoso^b, Susanna Mandruzzato^{c,d}, Vincenzo Bronte^e, Paolo Serafini^{f,b}, Luca Inverardi^b, Silvio Biciato^{a,*}

^a Department of Life Sciences, University of Modena and Reggio Emilia, Italy

^b Diabetes Research Institute, University of Miami, FL, USA

^c Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy

^d Istituto Oncologico Veneto IOV – IRCCS, Padova, Italy

^e Department of Pathology, Verona University Hospital, Immunology Section, Verona, Italy

^f Department of Microbiology and Immunology, University of Miami, Miami, FL, USA

ARTICLE INFO

Article history:

Received 13 October 2014

Received in revised form 24 October 2014

Accepted 25 October 2014

Available online 31 October 2014

Keywords:

Microarrays

Myeloid cells

Transcriptional data

ABSTRACT

Myeloid-derived suppressor cells (MDSCs) have been shown to control self-reactive and anti-graft effector T-cells in autoimmunity and transplantation, but their therapeutic use is limited by their scarce availability in the peripheral blood of tumor-free donors. We isolated and characterized a novel population of myeloid suppressor cells, named fibrocytic MDSC (f-MDSC), which are differentiated from umbilical cord blood (UCB) precursors (Zoso et al., 2014). This MDSC subset promotes regulatory T-cell expansion and induces normoglycemia in a xenogeneic model of type 1 diabetes. Here we describe in details the experimental design and the bioinformatics analyses of the gene expression dataset used to investigate the molecular mechanisms at the base of MDSC tolerogenic and suppressive properties. We also provide an R code to easily access the data and perform the quality controls and basic analyses relevant to this dataset. Raw and pre-processed data are available at Gene Expression Omnibus under accession [GSE52376](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52376).

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications	
Organism/cell line/tissue	<i>Home sapiens Umbilical Cord Blood derived CD33⁺IL4Rα⁺ cells (f-MDSC); HEK-293 cell line ATCC # CRL-1573</i>
Sex	n/a
Sequencer or array type	Affymetrix Human Gene 2.0 ST Array
Data format	Raw data (CEL files)
Experimental factors	<i>f-MDSCs; f-MDSCs co-cultured with PHA activated CD3⁺T-cells in cell to cell contact; f-MDSCs co-cultured with PHA activated CD3⁺T-cells in a transwell device</i>
Experimental features	<i>Experiment to compare gene expression in f-MDSCs co-culture with CD3⁺T-cells (before contact), after 3-day co-culture with T-cells in conventional dishes (contact), after 3-day co-culture with T-cells using trans-well membranes (transwell), and in HEK-293 cells.</i>
Consent	n/a
Sample source location	Diabetes Research Institute, University of Miami 1450 NW 10th Ave Miami 33136 Florida, USA

Direct link to deposited data

Deposited data are publicly available at Gene Expression Omnibus (GEO) under [GSE52376](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52376): <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52376>.

Experimental design, materials and methods

Cell culture

Fibrocytic myeloid-derived suppressor cells (f-MDSCs) were differentiated from umbilical cord blood (UCB) precursors by a 4-day culture in the presence of recombinant human (rh)-GM-CSF and rh-G-CSF after Ficoll-Paque gradient separation and red cell lysis [1]. Cultures in the presence of both cytokines result in the differentiation of a population of myeloid cells that co-express markers of MDSC (as CD33, IL4Rα, CD13, CD14, and CD15), markers of dendritic cells (e.g., CD11c, HLA-DR, CD86 and CD40) and molecules that characterize fibrocytes, fibroblasts, and other cell types involved in tissue remodeling and repair. To functionally characterize this population, the CD33⁺IL4Rα⁺ sub-fraction was FACS-sorted and tested for their capacity to suppress the proliferation of autologous PHA (Phytohemagglutinin,

* Corresponding author at: Department of Life Sciences, University of Modena and Reggio Emilia, via G. Campi, 287, 41125 Modena, Italy. Tel.: +39 059 2055 219; fax: +39 059 2055 410.

E-mail address: silvio.biciato@unimore.it (S. Biciato).

0.6%, Gibco) activated, carboxyfluorescein-succinimidyl-ester (CFSE, Molecular Probe) labeled CD3⁺T-cells negatively isolated with human Pan-T-cell magnetic beads and an LS column (Miltenyi-Biotec). Suppressive assays were performed in either conventional dishes (contact) or using trans-well membranes (Corning) separating f-MDSC and CD3⁺T-cells. HEK-293 human fibroblasts were used as negative controls.

Microarray study design

We extracted RNA from f-MDSC CD33⁺IL4R α ⁺ cells under the following conditions: 1) before co-culture with CD3⁺T-cells (*before contact*; n = 7), 2) after 3-day co-culture with T-cells in conventional dishes (*contact*; n = 7), and 3) after 3-day co-culture with T-cells using trans-well membranes (*transwell*; n = 7). RNA was also prepared from two replicates of HEK-293 cells maintained in complete RPMI-140 10% FBS complete media (Gibco). The RNA samples were subjected to microarray analysis using a Human Gene 2.0 ST Array (Affymetrix, Santa Clara, CA).

RNA labeling and hybridization

RNA samples were stored in Trizol (Invitrogen), purified using the miRNeasy mini kit (Qiagen), analyzed with an Agilent Bioanalyzer 6000

(Agilent Technologies), and hybridized by the Gene Expression Core Facility at the University of Miami using Affymetrix GeneChip WT Terminal Labeling and Hybridization, Wash, and Stain kits. Briefly, sense-strand cDNA was fragmented with a combination of uracil DNA glycosylase and apurinic/aprimidinic endonuclease 1 and its yield further labeled by terminal deoxynucleotidyl transferase (TdT) with the Affymetrix proprietary DNA Labeling Reagent that is covalently linked to biotin. Following labeling, 10 μ g of cRNA was hybridized for 16 h at 45 °C on GeneChip Human Gene 2.0 ST arrays. Arrays were washed and stained (streptavidin, R-phycoerythrin conjugate and biotinylated anti-streptavidin) in the Affymetrix Fluidics Station 400 and scanned using standard Affymetrix scanning procedures on a Scanner 3000 7G. Raw intensity data (CEL files) were generated using Affymetrix GeneChip Command Console Software (AGCC). CEL files and sample annotation data were deposited in GEO under accession number GSE52376.

Quality controls, normalization, and signal quantification

Quality controls were performed using the *oligo* and *AffyQCControl* Bioconductor packages on an ExpressionFeatureSet object generated from CEL files. As can be seen from the distribution of all PM intensities (Fig. 1A), from the heat map of the array–array Spearman rank correlation coefficients (Fig. 1B) and from Relative Log Expression (RLE; Fig. 1C) and Normalized Unscaled Standard Error (NUSE; Fig. 1D)

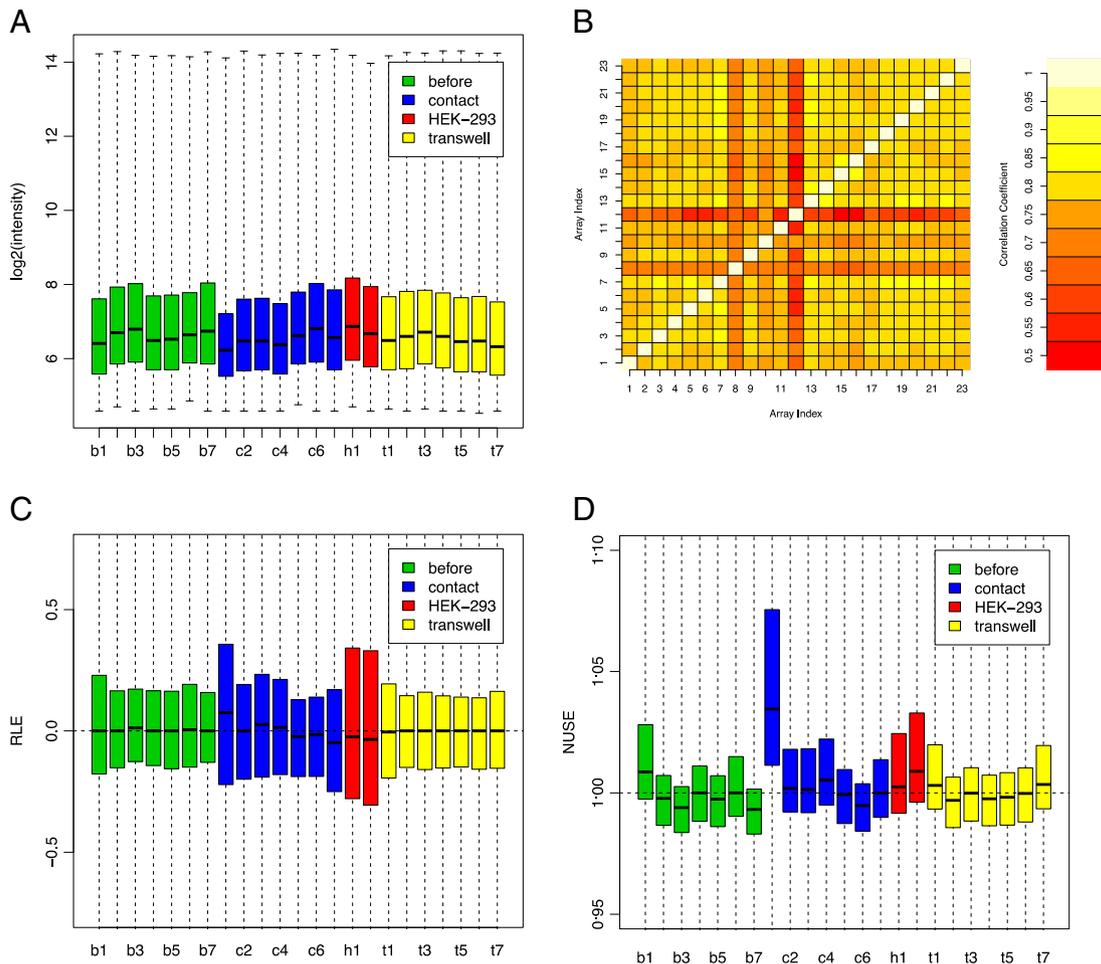


Fig. 1. Quality controls for the Affymetrix raw data. (A) Distribution of all raw PM intensities; (B) heat map of the array–array Spearman rank correlation coefficients; (C) Relative Log Expression (RLE) values computed for each probe set by comparing the expression value on each array against the median expression value for that probe set across all arrays; (D) Normalized Unscaled Standard Errors (NUSE), i.e., the standard error estimates obtained for each gene on each array by the *fitProbeLevelModel* function of the Bioconductor *oligo* package and standardized across arrays so that the median standard error for that gene is 1 across all arrays. Green bars: f-MDSC before co-culture with CD3⁺T-cell (b = *before*), blue bars: f-MDSC after co-culture with CD3⁺T-cell in conventional dishes (c = *contact*); red bars: HEK-293 cell line (h); yellow bars: f-MDSC after co-culture with CD3⁺T-cell using trans-well membranes (t = *transwell*).

plots, all arrays in the dataset have good and reproducible quality metrics according to standard guidelines.

CEL file fluorescence signals were converted to expression values using the robust multi-array average procedure RMA of the Bioconductor *affy* package [2]. Specifically, probe fluorescence intensities were background adjusted, normalized using quantile normalization, and log2 expression values for a total of 23,786 custom probe sets calculated using median polish summarization and the custom chip definition files for Human Gene 2.0 ST arrays based on Entrez genes (hugene20st_Hs_ENTREZG version 17.1.0; [3]). All data analyses were performed in R version 3.0.2 using Bioconductor libraries of BioC 2.13 and R statistical packages.

Unsupervised analysis

To identify samples that display similar gene expression profiles in an unbiased manner, we used unsupervised hierarchical clustering on the subset of probe sets that change their expression level among the various samples. Global unsupervised clustering was performed using the function *hclust* of the R *stats* package with the Pearson correlation as distance metric and average agglomeration method (Fig. 2A). In order to reduce the effect of noise from non-varying genes, those probe sets with a coefficient of variation smaller than the 90th percentile of the coefficients of variation in the entire dataset, were removed before the unsupervised clustering. The filter retained 2509 probe sets that are more variable across samples. To assess cluster-specific reproducibility, we calculated p-values for sample clusters using the multiscale bootstrap resampling method coded in the R *pvclust* package. Briefly, hierarchical clustering was first performed, using the correlation matrix of row-wise standardized expression values and the average agglomeration method, on the original data and on $B = 1000$ replications of different resampling of the given dataset (bootstrapped data). Then, p-values were computed for all clusters of the original data as the frequency that any cluster appears in the bootstrap replicates (Bootstrap Probability). To confirm the relationship between sample clusters, we used Principal Component Analysis (PCA) coded by the *prcomp* function of the R *stats* package. PCA is a mathematical technique to reduce data dimensionality while retaining most of its variation. Data reduction is accomplished by identifying directions, named principal components, along which the data variability is maximal. By using a few components, each sample can be projected in a space of reduced dimension making it possible to determine if samples form clusters or groups. Considering the filtered data matrix (2509 probe sets), the first 3 principal

components accounted for 45.73% (PC1), 29.93% (PC2) and 16.28% (PC3) of the total variance (Fig. 2B). Both unsupervised clustering and PCA indicate that f-MDSCs are completely different from HEK-293 fibroblasts and that f-MDSCs co-cultured with CD3⁺T-cell in contact condition (*contact*) slightly segregate from both *before contact* and *transwell* samples.

Discussion

Here we described the characteristics of the gene expression dataset used to investigate a novel population of myeloid suppressor cells, named fibrocytic MDSC (f-MDSC), that are differentiated from umbilical cord blood (UCB) precursors [1]. In particular, this dataset provided clues to elucidate the molecular mechanisms at the base of f-MDSC tolerogenic and suppressive properties, indicating that fibrocytic MDSCs require direct contact with activated T-cells to exert their pro-tolerogenic function. This new myeloid subset may represent an important tool for the in vitro and in vivo production of T regulatory cells for the treatment of autoimmune diseases and the prevention or control of allograft rejection.

Acknowledgments

We thank Dr. Mattia Forcato for the critical reading of the manuscript and for the revisions on the Sweave document. This work was supported by the Italian Ministry of Education, Research and Universities (FIRB grant RBAP11T3WB Accordi di Programma 2011 and EPIGEN: Italian Flagship Project Epigenomics), by AIRC Special Program Molecular Clinical Oncology '5 per mille' (grant 10016), and by the Diabetes Research Institute Foundation, Stacy Joy Goodman Memorial Foundation, Diabetes Research & Wellness Foundation, Foundation for Diabetes Research, Frank Strick Foundation, and by M0800359 YCMI award from FAMRI and the KG090350 award from Susan G. Komen for the Cure Foundations.

Appendix A. Supplementary data

Supplementary data to this article include a Sweave file to reproduce the analysis and the Sweave report generated using R version 3.0.2 and Bioconductor libraries of BioC 2.13. Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.gdata.2014.10.018>.

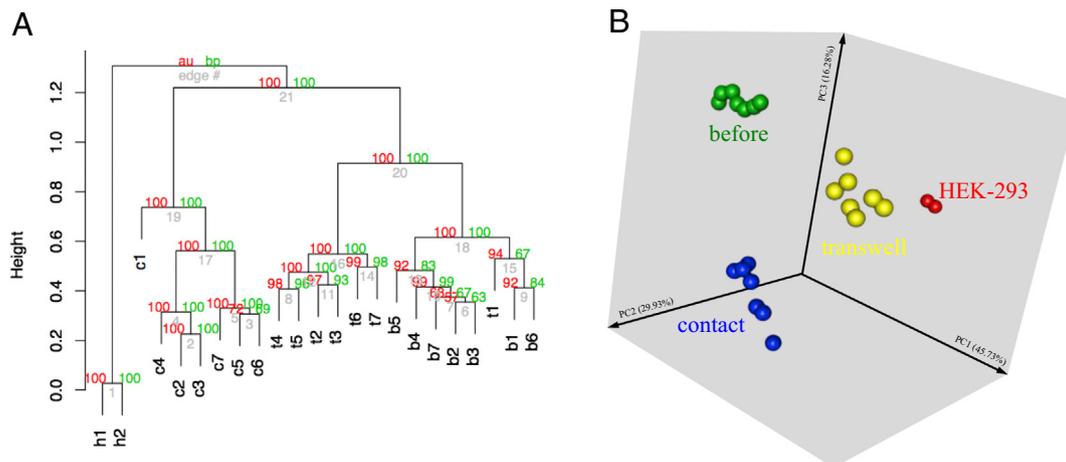


Fig. 2. Unsupervised analysis. (A) Unsupervised hierarchical clustering of the f-MDSC and HEK-293 samples using gene expression data from 2509 probe sets that are more variable across samples. AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) values are indicated in red and green, respectively. AU p-value is computed by multiscale bootstrap resampling while BP value is computed by normal bootstrap resampling. (B) Principal component analysis (PCA) of the gene expression data presented in (A). Colors and abbreviation are as in Fig. 1.

References

- [1] A. Zoso, E.M. Mazza, S. Bicciato, S. Mandruzzato, V. Bronte, P. Serafini, L. Inverardi, Human fibrocytic myeloid-derived suppressor cells express IDO and promote tolerance via Treg-cell expansion. *Eur. J. Immunol.* (2014), <http://dx.doi.org/10.1002/eji.201444522> (Aug 11).
- [2] R.A. Irizarry, B. Hobbs, F. Collin, Y.D. Beazer-Barclay, K.J. Antonellis, U. Scherf, T.P. Speed, Exploration, normalization, and summaries of high-density oligonucleotide array probe level data. *Biostatistics* 4 (2) (2003) 249–264 (Apr).
- [3] M. Dai, P. Wang, A.D. Boyd, G. Kostov, B. Athey, E.G. Jones, W.E. Bunney, R.M. Myers, T.P. Speed, H. Akil, S.J. Watson, F. Meng, Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33 (20) (2005) e175 (Nov 10).