Single-Cell RNA-Seq of Bone Marrow-Derived Mesenchymal Stem Cells Reveals Unique Profiles of Lineage Priming

Brian T. Freeman[1], Jangwook P. Jung[1], Brenda M. Ogle[1,2,3,4]*

1 Department of Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota, 55455, United States of America, 2 Stem Cell Institute, University of Minnesota, Minneapolis, Minnesota, 55455, United States of America, 3 Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, 55455, United States of America, 4 Lillehei Heart Institute, University of Minnesota, Minneapolis, Minnesota, 55455, United States of America

* ogle@umn.edu

Abstract

The plasticity and immunomodulatory capacity of mesenchymal stem cells (MSCs) have spurred clinical use in recent years. However, clinical outcomes vary and many ascribe inconsistency to the tissue source of MSCs. Yet unconsidered is the extent of heterogeneity of individual MSCs from a given tissue source with respect to differentiation potential and immune regulatory function. Here we use single-cell RNA-seq to assess the transcriptional diversity of murine mesenchymal stem cells derived from bone marrow. We found genes associated with MSC multipotency were expressed at a high level and with consistency between individual cells. However, genes associated with osteogenic, chondrogenic, adipogenic, neurogenic and vascular smooth muscle differentiation were expressed at widely varying levels between individual cells. Further, certain genes associated with immunomodulation were also inconsistent between individual cells. Differences could not be ascribed to cycles of proliferation, culture bias or other cellular process, which might alter transcript expression in a regular or cyclic pattern. These results support and extend the concept of lineage priming of MSCs and emphasize caution for *in vivo* or clinical use of MSCs, even when immunomodulation is the goal, since multiple mesodermal (and even perhaps ectodermal) outcomes are a possibility. Purification might enable shifting of the probability of a certain outcome, but is unlikely to remove multilineage potential altogether.

Introduction

Mesenchymal/multipotent stem/stromal cells (MSCs) are utilized in stem cell therapy for treatment of a variety of diseases including myocardial infarction, cancer, lung fibrosis, spinal cord injury, bone and cartilage repair, and muscular dystrophy[1–4]. MSCs are clinically beneficial due in part to the ability to home to sites of injury[5, 6], differentiate to mesenchymal cell types, suppress immune responses[7] and modulate angiogenesis[8–10]. In addition, MSCs are

easy to isolate and expand and can be derived from multiple different tissue sources including bone-marrow, fat, placenta, synovium, periosteum, and tooth[2].

The large variety of tissue sources and species from which MSCs can be isolated have spurred efforts to characterize and compare each MSC isolate. The approach has been to identify a protein marker, or series of markers unique to MSCs and then to validate multipotency via differentiation protocols. For example, human MSCs are typically isolated from bone-marrow by selecting for adherent cells then confirming expression of $CD73^+/CD90^+/CD105^+/CD34^-/CD14^-/CD19^-/CD45^-$ via a variety of methods including flow cytometry or fluorescence microscopy[11]. Use of the entire panel is inconsistent, as are the subsets selected by individual investigators[12, 13]. A similar trend occurs with isolation and characterization of murine MSCs derived from bone marrow. In this case, more than thirty different surface markers have been used with varying subsets over the past 15 years[14]. It is challenging to determine whether subset selection indicates an assumption by investigators that each subset reflects the whole or that a given isolate does not in fact express certain markers. But we do know that inconsistent use of MSC biomarkers to isolate "pure" populations can lead to variable levels of differentiation potential and ability to self renew[15, 16].

More perplexing is the fact that *consistent* use of biomarkers, can also lead to variable *in vitro* and *in vivo* outcomes between research groups. For example, murine bone marrow-derived MSCs sorted via immunodepletion of CD11b and CD45 for treatment of acute lung injury showed increased survival across studies but associated mechanisms were varied and sometimes contradictory[17–20]. Gupta et al., showed elevation of IL-10 and no change in neutrophil infiltration relative to controls[18], while Xu et al., showed no change in IL-10 production, but a decrease in neutrophil infiltration[17]. In addition, retention of water in the lungs was significantly decreased only after 48 hours in the study by Gupta et al., while in Xu et al., water retention was only observed at 24 hours and lost by 48. Some of this variation has been attributed to age or disease of the organism at the time of MSC isolation[21]. More likely however, is the fact that a handful of proteins cannot adequately describe the varied members of MSCs between species, between tissues and even perhaps between cells of a given population. Advanced molecular approaches including microarray[22], qPCR[23] and RNAseq[12, 13, 24–29] have allowed extensive characterization of 100s to 10,000s of transcripts of MSC populations. Results of these studies suggest heterogeneity of MSC populations could be ascribed to populations containing a varied mixture of undifferentiated MSCs primed[12, 30] for multiple pathways. Lineage priming has been observed with hematopoietic stems cells using single cell RT-PCR[31] and microarray[32, 33]. Transcription factors from both erythroid and myeloid differentiation pathways were expressed at a variety of levels in hematopoietic stems cells, suggesting that these stem cells could differentiate effectively down either pathway. Delorme et al., proposed that a similar process might occur with MSCs, but the study was conducted on clonal populations using qPCR and therefore a limited number of transcripts[12].

Until recently, cell-by-cell, whole-transcriptome analysis of MSCs has not been possible. However, advances in microfluidics and small volume cDNA synthesis now allow single-cell RNA-seq of individual MSCs[34]. Using this approach we find that individual MSCs exhibit multilineage priming, but priming is not uniform and appears to favor one and sometimes two lineages even while maintaining multipotency. In addition, a limited, but measureable degree of heterogeneity is observed in expression of genes associated with immunomodulation in the absence of immune stimulation. These results point to an as yet unappreciated source of heterogeneity of MSCs from a single tissue source.

Materials and Methods

Cell culture

Mouse bone marrow mesenchymal stem cells (mMSCs) were purchased (3 month-old male C57BL mice, Georgia Reagents University, Augusta, GA, all procedures approved by the Institutional Care and Use Committee, Medical College of Georgia[35]) and expanded and cultured as previously described[36]. Briefly, MSCs were cultured on a 0.1% gelatin (Sigma Aldrich, St. Louis, MO) pretreated flask containing α-minimum essential medium (MEM) complete. Complete alpha-MEM consisted of α-MEM (Invitrogen, Carlsbad, CA), 10% fetal bovine serum (Hyclone, Logan, UT), 0.1 mM nonessential amino acids (Invitrogen), and 2 mM L-glutamine (Invitrogren). MSC cultures were allowed to grow to 60–70% confluence and were replated at a concentration of 1,500 cells/cm^2. Experiments were performed using passages 6–11 for mMSCs. As a control, HL-1 cardiomyocytes (HL-1cm) (a gift of Dr. William Claycomb) were also expanded and cultured as previously described[37]. All cultures were maintained at 37°C in 5% CO_2.

Single-cell capture and RNA-seq

mMSCs were trypsinized and suspended in phosphate buffered saline (PBS). The mMSCs were centrifuged and resuspended in 5 μL of complete alpha-MEM medium for addition to the capture chip. mMSCs were captured on a large-sized (17–25 μm cell diameter) chip using the Fluidigm C1 system. Cells were loaded into the chip at approximately 2000 cells/L and stained for viability (DEAD cell viability assay; Molecular Probes, Life Technologies, Grand Island, NY) and imaged with phase-contrast and fluorescence microscopy to confirm cell number and viability at each capture point. mMSCs and the HL-1cm controls were captured in the same Fluidigm C1 device with the HL-1cm labled with a green cytoplasmic dye (1 μm CellTracker Green CMFDA, Molecular Probes, Eugene, OR). 16 single, live mMSCs and 5 single, live HL1cm were selected for analysis. Once cells were captured in the device, cDNAs were prepared from each cell on the chip using the SMARTer Ultra Low RNA kit for Fluidigm C1 System (Clontech, Mountain View, CA). RNA spike-in Mix (Ambion, Life Technologies) was added to the lyses reaction and processed with cellular mRNA. mRNA library was constructed using the Illumina Nextera XT preparation kit (Illumina, San Diego, CA) according to the manufacturer's protocol and sequenced on the Illumina MiSeqv3 using paired end reads with a length of 75 bp to a depth of 18 to 22 million reads with the Multiplex on one MiSeq lane to create *. fastq files. A bulk population RNA control of both mMSCs and HL-1cm was run in parallel to the single-cell samples. RNA-seq data were submitted to NCBI Gene Expression Omnibus and can be accessed via the following link, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70930.

Gene expression analysis

Gene expression analysis was performed with Galaxy software (Minnesota Supercomputing Institute (MSI), University of Minnesota, Minneapolis, MN). Reads were processed and aligned to the mouse reference genome (mm10_genes_2012_05_23.gtf and canonical_mm10.fa) using Tophat (version 2.0.12, open source software, http://ccb.jhu.edu/software/tophat/index.shtml) [38]. See **S1 Table** for information about the total number of reads and percent concordant mapped reads for each cell. The default options supplied with the software were used and the aligned read files produced by Tophat were processed using Cufflinks software (version 2.2.1, open source software, http://cole-trapnell-lab.github.io/cufflinks/), for further analysis, including assembling transcripts, estimating their abundance, and testing for the differential

expression between single-cell RNA-seq samples[38]. Read counts were normalized to fragments per kilobase of exon per million mapped reads (FPKM) according to the gene length and total mapped reads. Genes with a $\log_2$ fold change greater than 1 from the control cells to the fusion products and had a *P* value of less than 0.05 were considered "differentially expressed" and further analyzed for gene ontology (**S1 Table**). Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed with DAVID informatics resources 6.7 of the National Institute of Allergy and Infectious Diseases (NIAID) and of the National Institutes of Health (NIH)(**S3 Table**)[39, 40].

Gene cluster analysis

Average linkage hierarchical clustering of gene expression intensity was performed using the Pearson distance to measure distance between gene and single cells. SingulaR (Fluidigm, San Francisco, CA) was used to compute and create the hierarchal clustering and principle component analysis plots (**S2 Table**).

Statistical analysis

Data were analyzed with Microsoft Excel (Microsoft, Redmond, WA, USA). RNA-seq data was analyzed with the Cuffdiff or SingulaR programs.

Quantitative real-time polymerase chain reactions (qRT-PCR)

Complementary DNA (cDNA) was synthesized following the instructions from the Maxima First Strand cDNA Synthesis Kit (cat# K1642, Thermo Scientific). The cDNA was amplified from the cDNA from the single-cell reaction performed in the Illumina chip. Primers of qRT-PCR were purchased from Biorad (Hercules, CA) and the sequence of *Gapdh* primers (FOR: CATGGCCTTCCGTGTTCCTA, REV: CCTGCTTCACCACCTTCTTGAT) was obtained from [41]. Primer efficiencies were extracted from RealPlex^2 software and verified with melting curves. The comparative C_t method[42] was employed to determine the relative changes in gene expression. For comparison to the single-cell RNA-seq results, each gene's FPKM values were normalized to each cells *Gapdh* FPKM value, which was then normalized to the control cells.

Results

Bone marrow-derived mMSCs express genes associated with MSC multipotency

To begin characterization of the transcriptome of bone marrow-derived MSCs, sixteen mouse mesenchymal stem cells (mMSC1-mMSC16) were individually sequenced using single-cell RNA-seq. Also, a population control (PC) containing thousands of cells (mMSC-PC) was sequenced in parallel with the single cells. Next, a gene set was assembled that included published markers indicative of MSC multipotency[14, 43, 44] (**S4 Table**). Hierarchal clustering (HC) and principal component analysis (PCA) analysis were then conducted using this gene set. HL-1 cardiomyocytes (HL1cm) were used as a negative control for this analysis and *Gapdh* was used to represent basal expression of a housekeeping gene. The mMSCs all clustered together far from the HL-1cm controls (**Fig 1A and 1B**). Of note, all mMSCs and mMSC-PC exhibited high expression of MSC stemness-associated markers previously shown to be upregulated in mouse MSCs (*Ly6a*, *Cd9*, *Cd44*, *Sdc4*, *Lamp2)* and low expression of MSC stemness-associated markers previously shown to be downregulated in mouse MSCs (*Anpep*, *Itgam*, *Eng*, *Nt5e*, *Pecam1*, *Nanog)[14]*. In addition, mMSCs exhibited inconsistent expression of markers
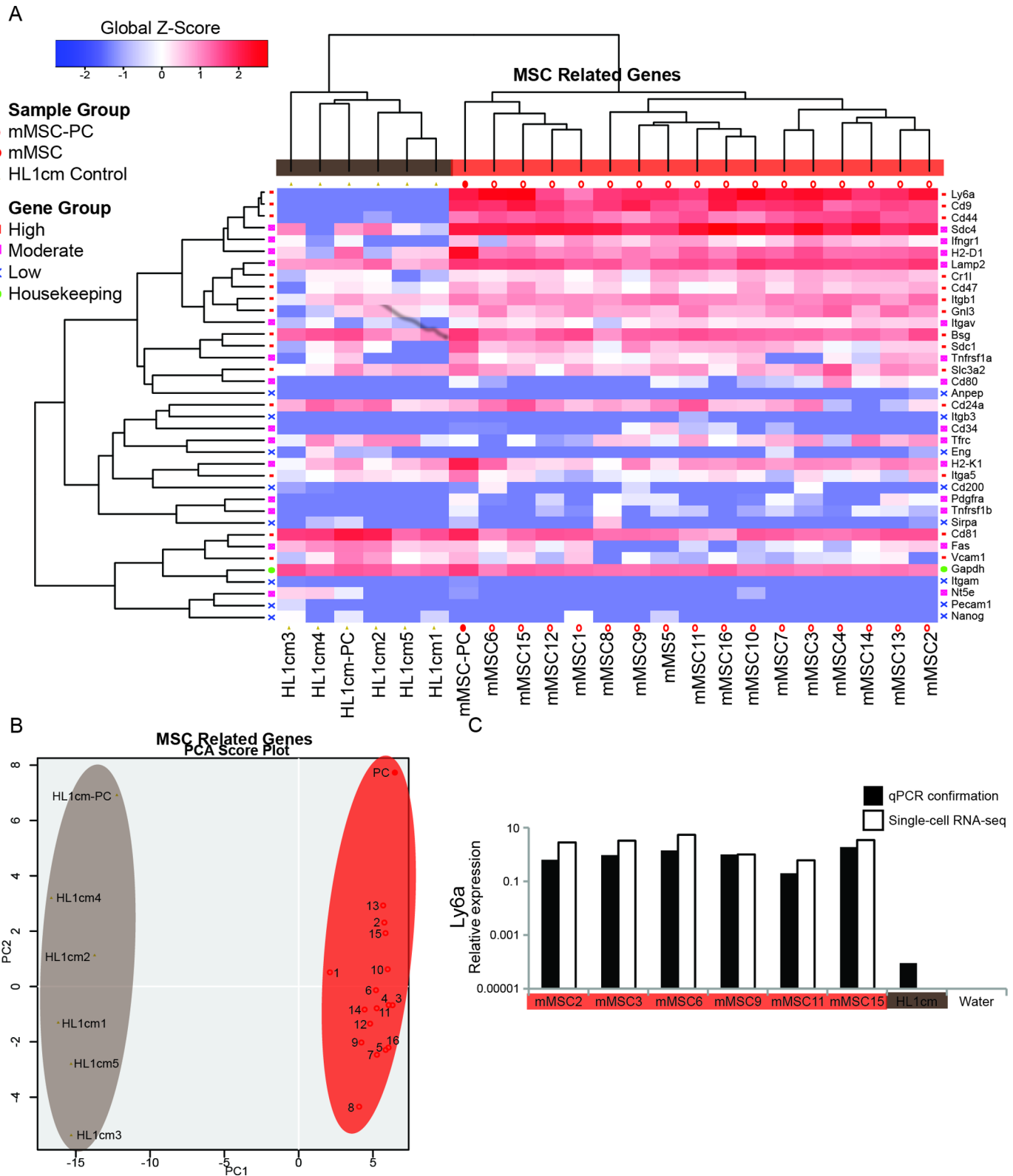
Fig 1. Single-cell mMSCs express genes associated with MSC multipotency (A) Hierarchal clustering (HC) of mMSCs for a set of genes related to MSC multipotency. Genes are organized either by genes expressed highly in mMSCs (red), genes expressed moderately/heterogeneously in mMSCs (pink), genes that are lowly expressed in mMSCs (blue) or housekeeping genes (green). Global Z-Score reflects the number of standard deviations away from the

mean of expression in the reference. **(B)** Principal component analysis (PCA) analysis of single cell mMSCs (1–16) (red), negative control single cells (HL1cm1-HL1cm5) (brown), and population controls (PC and HL1cmPC). **(C)** Quantitative reverse transcriptase polymerase chain reaction (qPCR) confirmation of RNA-seq data.

reported to have variable expression in mMSCs (*Pdgfra, Cd80, Cd34, Tfrc)*[14]. The HL1cm control cells expressed MSC markers at a much lower level, but with similar levels of *Gapdh* expression, demonstrating the specificity of the gene set for mMSCs. RNA-seq data was confirmed with qPCR analysis of *Ly6a* showing similar relative expression between mMSCs and low expression in the HL1cm control (**Fig 1C**).

Multilineage priming of <u>individual</u> bone marrow-derived mMSCs

To determine whether differentiation potential could account, at least in part, for varied *in vivo* outcomes, we generated gene sets corresponding to mesenchymal specification[12, 45–53]. The gene list included early stage (e.g. transcription factors) and late stage (e.g. functional proteins) markers for osteogenic, chondrogenic, adipogenic, vascular smooth muscle and neurogenic lineages. Using this set, HC and PCA analyses were conducted on single-cell mMSCs and the mMSC-PC (**Fig 2A–2C**). The mMSC-PC control correlated with the average of the single mMSCs supporting the accuracy of the single-cell data. In addition, the population control contained transcripts corresponding to early and late markers of multiple lineages at variable levels, a finding perhaps attributable to the mixing of multiple lineage-committed or lineage-primed populations. The profile of the mMSC-PC control was similar to that of each single-cell mMSC (**Fig 2A**). Individual mMSCs did not express markers of a single lineage, but instead showed expression of early stage markers of at least three differentiation pathways *simultaneously* (**Fig 2E** and **S1 Fig**). Since it was previously established that all mMSCs used in this study were still expressing high levels of multipotency-associated genes (**Fig 1A and 1B**), this was quite surprising. The separation of individual MSCs in the PCA plot was due to variable expression of genes from all five lineages. Cells at the top of the PCA score plot had higher levels of a few osteogenic (*Ogn* and *Wisp2*) and adipogenic (*Fabp4* and *Cebpb*) markers, cells on the right had higher levels of vascular smooth muscle markers (*Acta2, Cnn1* and *Tagln*), *Tubb* (neurogenic), *Slco2a1* and *Angptl4* (most often associated with adipogenesis) (**Fig 2C**). In an attempt to more easily visualize the varied linage propensities of individual mMSCs, one key transcription factor for each lineage (*Runx2, Sox9, Cebpb, Gata6* and *Nr4a2*) was selected and normalized to *Gapdh* expression for all sixteen cells and the population control (**Fig 2E**). The population control revealed a similar profile to that seen for the clonal populations of the Delorme study[12], but surprisingly each individual mMSCs exhibited much more varied profiles. Seven of the sixteen mMSCs expressed all five transcription factors, six mMSCs expressed four of the transcription factors and three mMSCs expressed three of the transcription factors. These data were confirmed with qPCR of a gene highly expressed in all mMSCs (*Ly6a*) (**Fig 1C**) and a gene with high variability of expression in individual mMSCs (*Fabp4*) (**Fig 2D**).

Basal immunomodulatory capacity of <u>individual</u> bone marrow-derived mMSCs

To determine whether basal immunomodulatory capacity could also account for varied MSC behavior *in vitro* and *in vivo*, we generated a gene set containing soluble, immunoregulatory factors known to be produced by MSCs *in vitro* and *in vivo*[54–56] (**Fig 3A, S4 Table**). Individual cells were largely the same with the exception of *Il6* expression. Indeed variability in the expression of this gene accounted almost solely for the population spread on the PCA score
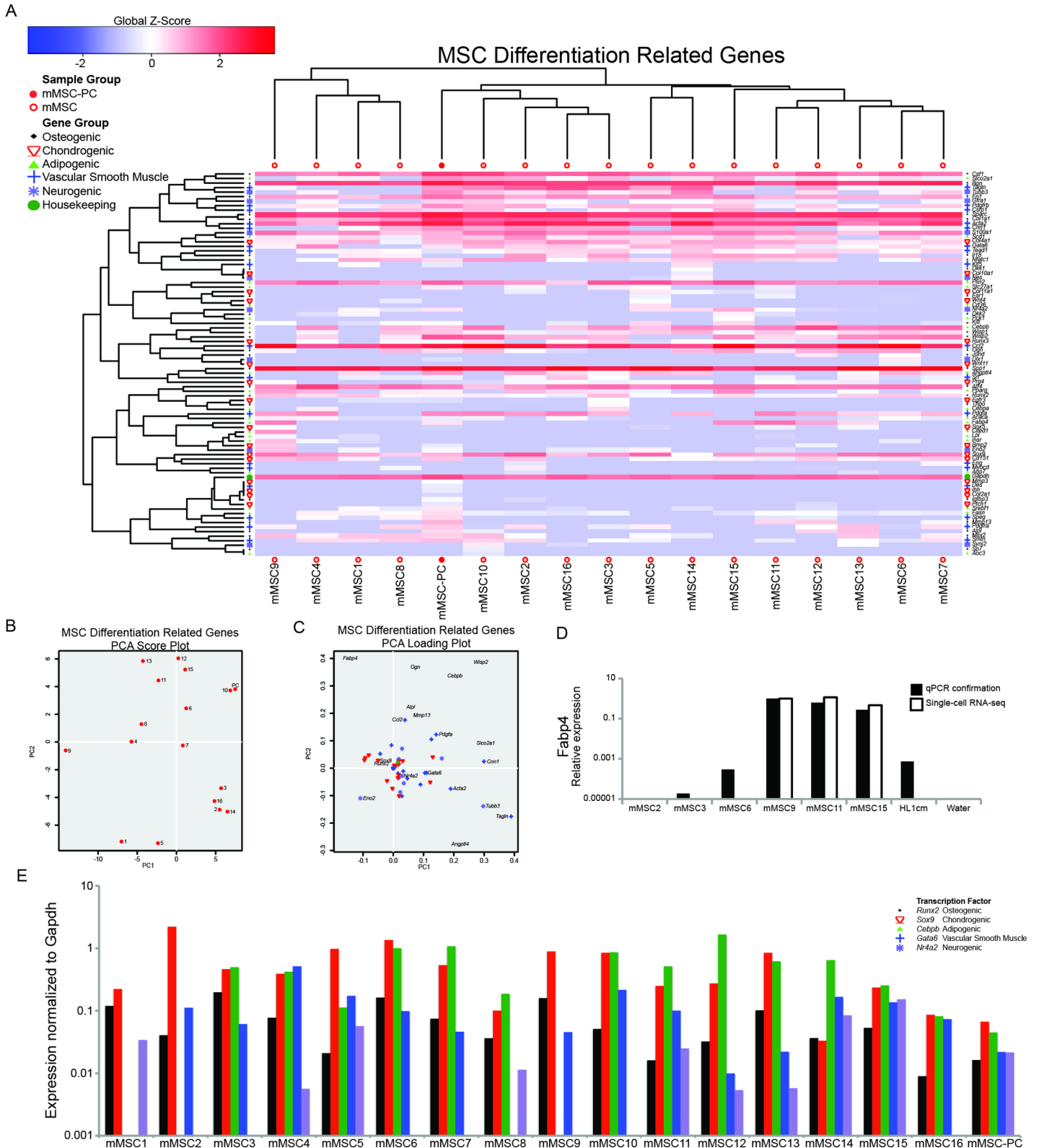
Fig 2. Individual mMSCs express genes from multiple differentiation pathways. (A) HC of single cell mMSCs and mMSC population control (mMSC-PC) for a set of genes related to osteogenesis, chondrogenesis, adipogenesis, vascular smooth muscle, neurogenesis, and housekeeping genes. **(B)** PCA analysis of single cell mMSCs (1–16) and population control (PC). **(C)** PCA loading plot showing individual differentiation genes represented in **(B)**.

(D) qPCR confirmation of RNA-seq data. **(E)** A key transcription factor was selected for each differentiation lineage and expression was normalized to *Gapdh* for each sample.

plot (**Fig 3B and 3C**). The significance of this difference was further probed via global analysis of differentially expressed genes.

Transcriptome heterogeneity of mMSCs

To more comprehensively delineate differences in the transcriptome that might augment findings related to lineage priming and support/refute subtle differences in immunomodulatory function, we identified all differentially expressed genes between each mMSC as well as the population control. Differential expression was defined as a $\log_2$ fold change of greater than 1 and a *P* value below 0.05. Over 9,000 differentially expressed genes were identified and displayed using HC (**Fig 4A**) and PCA score plot (**Fig 4B**). Results indicate global heterogeneity of single mMSCs arguably expected of mesenchymal progenitors (**Fig 4A and 4B**). Despite the heterogeneity, five different clusters emerged and are designated with different color labels. At first glance, we suspected clusters might represent different types and levels of lineage commitment of mesenchymal progenitors. To test this possibility, gene ontology analysis was conducted on the differentially expressed genes to determine whether each cluster aligned with specific mesenchymal differentiated cell types (e.g., osteogenic, adipogenic, chondrogenic).

DAVID bioinformatics resource was utilized to define function-related gene groups associated with each cluster of MSCs. Differentially expressed genes were isolated for each cluster of MSCs and separated according to up- or down-regulation (**S3 Table**). The significance of each concentrated gene group was plotted on a negative log scale with upregulated gene groups extending upward and downregulated gene groups extending downward (**Fig 4C**). Only gene groups with a *P* value below 0.05 were included in the graph and surprisingly only one was related to differentiation. The first cluster (green) contained upregulation of genes involved in cell division, ATP binding, DNA replication, and chromosomes which suggests that this cluster of five cells (along with the population control) was in active proliferation at the time of cell capture. The ratio of cells (5 of 16, ~30%) matches our data and literature reports of the percentage of cells actively dividing in a healthy population of bone marrow-derived MSCs[57]. The blue cluster showed upregulation of genes associated with DNA metabolic process, chromosomes, purine metabolism and nucleotide binding. The blue cluster did not contain genes upregulated with cell division, but these three cells do appear to be preparing DNA for future division. The cell of the yellow cluster (mMSC15) upregulated genes associated with regulation of transcription and protein catabolic process, while downregulating genes associated with cell division and regulation of neurogensis. The purple cluster also downregulated genes linked to cell division, but upregulated genes involved with growth factor activity, vesicle-mediated transport, lipid biosynthetic process and the extracellular region. This cluster could perhaps represent active paracrine signaling corresponding to the immunomodulatory capacity of MSCs. If active paracrine signaling is defined as augmented *Il6* production (**Fig 3**, mMSC 3, 6, 5, 11, 12, 13, 15), then there is an insignificant relationship between active paracrine signaling and association with the purple cluster (Chi-square statistic 0.15, *P* = 0.70). Indeed, a relationship with similar strength exists between proliferation and downregulation of *Il6* (Chi-square statistic 2.14, *P* = 0.14). Instead, the purple cluster could represent a subpopulation of mMSCs with advanced commitment to adipogenesis, since this cluster also shows an upregulation of the lipid biosynthetic process. However, cross reference of cells of this group (mMSC2, 3, 7, 10, 12 and 13) to the differentiation analysis (**Fig 2A–2B**), shows each falls in different regions of the PCA plot suggesting they each have unique lineage priming, not significantly tending
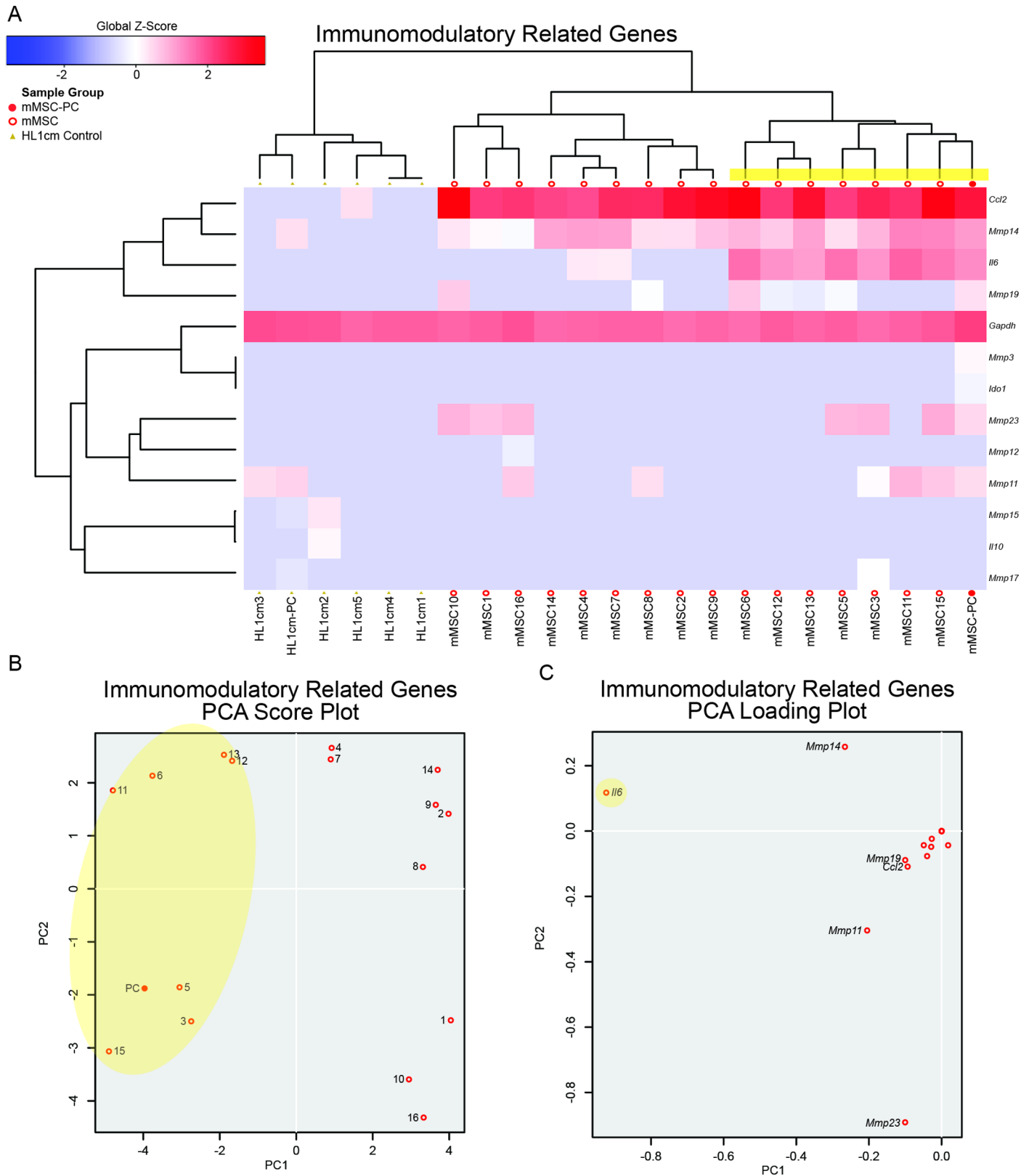
Fig 3. Single cell mMSCs exhibit minimal but measureable differences in expression of immunomodulatory genes. (A) HC of single cell mMSCs, negative control single cells (HL1cm1-5), and population controls (mMSC-PC and HL1cmPC) for a set of genes related to immunomodulatory function. (B)

PCA analysis of single cell mMSCs (1–16) and population control (PC). **(C)** PCA loading plot showing individual differentiation genes represented in **(B)**. *Il6* expressing cells clustered together (yellow).
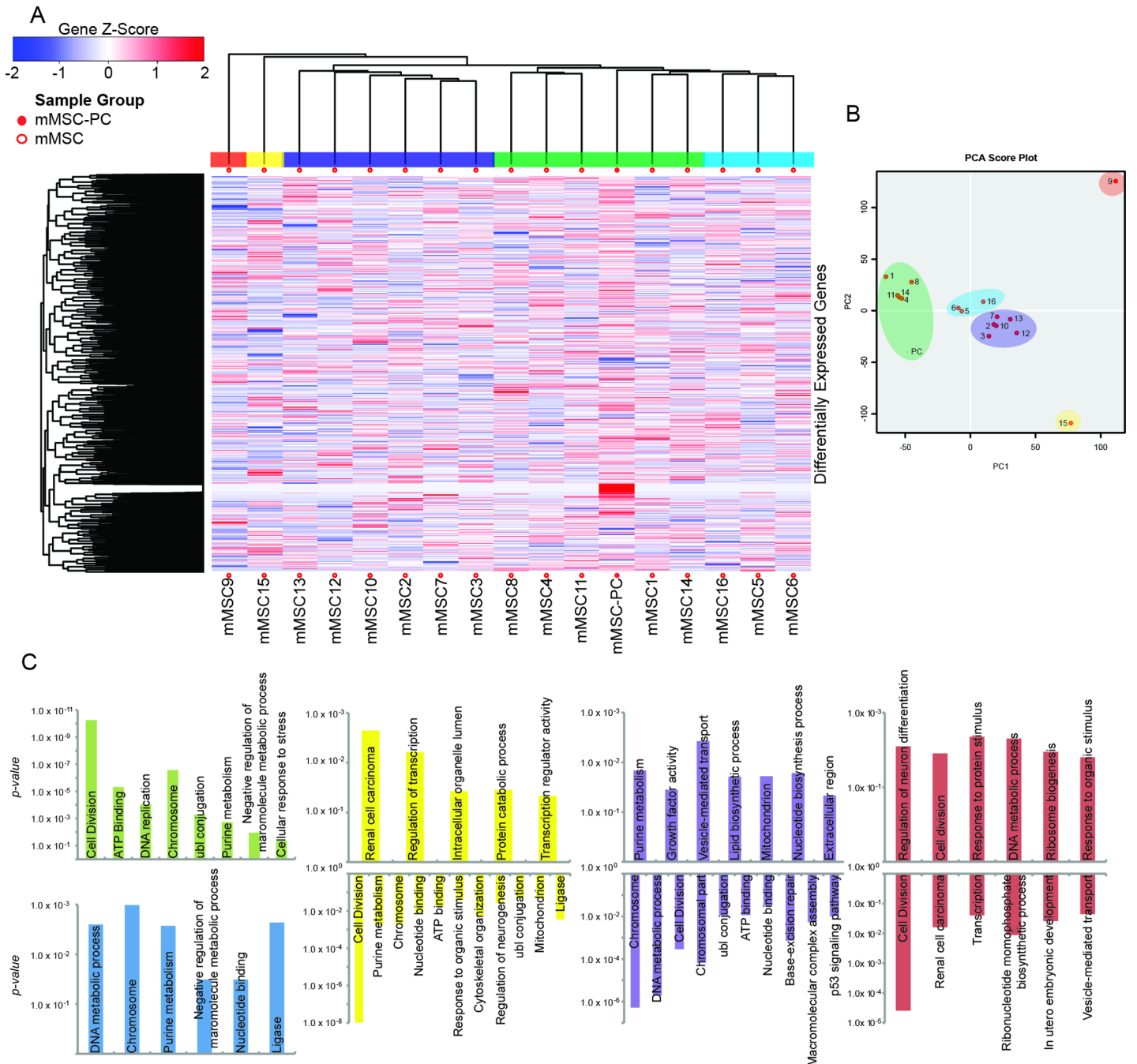
Fig 4. HC and PCA of mMSCs show transcriptome heterogeneity and gene ontology of subpopulations. (A) A global view of differential gene expression between single cell mMSCs and the population control (mMSC-PC). Gene Z-Score reflects the number of standard deviations away from the mean of expression in the reference. **(B)** PCA analysis of single cell mMSCs (1–16) and population control (PC). **(C)** Gene onotology of five subpopulations of single-cell mMSCs (green, blue, yellow, purple and red clusters). Ontology groups are plotted with the *P* value on the y-axis. Upregulated groups have bars extending in the positive y-direction and down regulated groups have bars extending in the negative y-direction.

toward adipogenesis. The cell of the red cluster (mMSC 9) showed an unexpected upregulation of genes involved with the regulation of neuronal differentiation and ribosome biogenesis, suggesting priming for neurogenesis which has been reported in isolated instances for MSCs[45–47]. In sum, mMSC clusters primarily correspond to proliferating cells (green and blue clusters) and quiescent cells (yellow, purple and red clusters). Importantly proliferation status does not correlate with the level or number of lineage commitment or immune regulatory genes expressed per cell.

Discussion

In this study the transcriptomes of individual, bone marrow-derived mMSCs were analyzed via RNA-seq. Using this approach, a new perspective on the heterogeneity of MSCs emerges. First, MSCs were found to express varied levels of early markers of multiple mesenchymal lineages extending the definition of lineage priming of this unique cell type. Importantly, differences in level or number of lineage commitment genes expressed did not correlate with proliferation or other cellular process, which might alter transcript expression in a regular or cyclic pattern. Second, basal expression of genes associated with immunomodulation were quite uniform aside from *Il6* in gene set analyses and categories related to immunomodulation did not emerge from ontology analyses suggesting cell-to-cell variation for immunomodulation is present but less pronounced than lineage commitment.

Since there was no clear functional clustering of the mMSCs based on gene ontology (aside from proliferating/nonproliferating), we also scanned for gene markers that could be used to represent the clustered populations (**Fig 2**) and perhaps correspond to lineage priming. A previous report on mouse MSCs subpopulations found that *Cd200* expression level indicated osteogenic potential and *Pdgfra* expression indicated adipogenic potential[14]. However, mMSC3, mMSC6, and mMSC9 were the only mMSCs with detectable *Cd200* expression and indeed these three mMSCs did not cluster in either the global analyses (**Fig 4**) or gene list analyses with markers of differentiation (**Fig 2**). *Pdgfra* expression also failed to aid clustering of the eleven mMSCs that had detectable expression (mMSC2, 3, 5, 6, 7, 8, 9, 10, 12, 13, 14). The discrepancy between our work and previous reports suggests an environmental trigger was necessary to promote particular cell fates and therefore that marker expression alone (at least in these two cases) is not enough to predict single cell propensities. We also note, based on the results presented here, that inter-lineage plasticity is likely very high. And that, even if an MSC "commits" to a particular lineage, conversion to another might not entail the set up of whole molecular programs, but instead upregulation of a few components[58].

Of note, one cell upregulated expression of neurogenic transcript for *Eno2*. This is not the first report of an MSC expressing transcripts considered specific to ectodermal lineage. In multiple reports Foudah *et al*. demonstrate that neuronal makers *Tubb* and *NeuN* are spontaneously expressed by a high percentage of undifferentiated MSCs from multiple tissues from both human and rat[45–47]. Thus, while the relative fraction of MSCs capable of expressing early transcripts of neurogenesis and of giving rise to bona fide neurons is not known, single cell analysis suggests plasticity to this lineage may be possible.

In sum, this study confirms and augments the concept of multi-lineage priming of MSCs by showing that MSCs typically express early (and sometimes even late) markers of more than one mesenchymal lineage *simultaneously*. Moreover, expression levels are quite distinct between cells, suggesting MSCs, despite being primed to multiple fates, might prefer one fate over the other. In other words, even if an MSC is environmentally triggered for one lineage, it might easily be switched if even a weak signal comes along for the preferred lineage. Immunomodulation of MSCs on the other hand, appears to be relatively consistent between cells of a

given population. Even so, these results emphasize caution for *in vivo* or clinical use of MSCs, even when immunomodulation is the goal, since multiple mesenchymal (and even perhaps ectodermal) outcomes are a possibility. Purification might enable shifting of the probability of a certain outcome, but can never remove multilineage potential altogether.

Supporting Information

S1 Fig. Expression of each lineage gene normalized to *Gapdh*. (A). Osteogenic gene profile. **(B)** Chondrogenic gene profile. **(C)** Adipogenic gene profile. **(D)** Vascular smooth muscle gene profile. **E**. Neurogenic gene profile.
(TIF)

S1 Table. Alignment statistics for single-cell with sequenced transcriptome, FPKM values and differentially expressed genes of mMSC clusters, Related to Fig 4. This first tab contains all of the alignment statistics for the captured mMSCs and HL1cm controls relative to the mouse genome. The number of aligned read pairs and the percent of concordant alignment (or reads with proper alignment) are listed for each single cell. The alignment data were generated from Tophat alignment results. The second tab is the single-cell RNA-seq expression data (FPKM values) for the mMSCs (16 cells and population control) and HL-1-cardiomyocytes controls (5 cells and population control) in an excel file as discussed in **Figs 1–4**. The remainder of the file contains all of the differentially expressed genes detected with the Cufflinks package analysis between the five mMSC clusters. The file contains the gene (columns A-C), locus (column D), samples compared (columns E and F), the FPKM values (H and I), the log$_2$ fold change (column J), *P* value (column L) and whether or not the differential expression is significant (column N). Genes are listed alphabetically and each mMSC cluster is separated onto different tabs.
(XLSX)

S2 Table. SingulaR script used to analyze single-cell RNA-seq data. This file contains the SingulaR script used with the single-cell RNA-seq data to create the hierarchal clustering and principal component analysis.
(XLSX)

S3 Table. Gene Ontology and KEGG pathway enrichment analysis, Related to Fig 4. This file contains all of the differentially expressed genes between the mMSC clusters, but organized into lists used for gene ontology on the first tab. The first tab is organized into upregulated and downregulated genes for each cluster of mMSCs. The remaining tabs contain the gene ontology and KEGG pathway enrichment analysis that was performed using DAVID informatics Resources 6.7 for upregulated or downregulated differentially expressed gene lists for mMSCs clusters as identified in the Cufflinks differential analysis.
(XLSX)

S4 Table. Genes Lists, Related to Fig 4. This file contains the genes associated with each gene list (MSC stemness[14, 43, 44], MSC Differentiation[12, 45–53], and immunomodulatory[54–56]. Included in this file are the gene symbol, full gene name, functional group and source for inclusion.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: BTF BMO. Performed the experiments: BTF JPJ. Analyzed the data: BTF BMO JPJ. Contributed reagents/materials/analysis tools: BMO. Wrote the paper: BTF BMO JPJ.

References

1. Barry FP, Murphy JM. Mesenchymal stem cells: clinical applications and biological characterization. Int J Biochem Cell Biol. 2004; 36(4):568–84. doi: 10.1016/j.biocel.2003.11.001 PMID: 15010324.

2. Deans RJ, Moseley AB. Mesenchymal stem cells: biology and potential clinical uses. Exp Hematol. 2000; 28(8):875–84. PMID: 10989188.

3. Ren G, Chen X, Dong F, Li W, Ren X, Zhang Y, et al. Concise review: mesenchymal stem cells and translational medicine: emerging issues. Stem Cells Transl Med. 2012; 1(1):51–8. doi: 10.5966/sctm.2011-0019 PMID: 23197640; PubMed Central PMCID: PMC3727691.

4. Williams AR, Hare JM. Mesenchymal stem cells: biology, pathophysiology, translational findings, and therapeutic implications for cardiac disease. Circ Res. 2011; 109(8):923–40. doi: 10.1161/CIRCRESAHA.111.243147 PMID: 21960725; PubMed Central PMCID: PMC3604746.

5. Mouiseddine M, Francois S, Semont A, Sache A, Allenet B, Mathieu N, et al. Human mesenchymal stem cells home specifically to radiation-injured tissues in a non-obese diabetes/severe combined immunodeficiency mouse model. Br J Radiol. 2007; 80 Spec No 1:S49–55. doi: 10.1259/bjr/25927054 PMID: 17704326.

6. Nagaya N, Fujii T, Iwase T, Ohgushi H, Itoh T, Uematsu M, et al. Intravenous administration of mesenchymal stem cells improves cardiac function in rats with acute myocardial infarction through angiogenesis and myogenesis. Am J Physiol Heart Circ Physiol. 2004; 287(6):H2670–6. doi: 10.1152/ajpheart.01071.2003 PMID: 15284059.

7. Djouad F, Bouffi C, Ghannam S, Noel D, Jorgensen C. Mesenchymal stem cells: innovative therapeutic tools for rheumatic diseases. Nat Rev Rheumatol. 2009; 5(7):392–9. doi: 10.1038/nrrheum.2009.104 PMID: 19568253.

8. Zacharek A, Chen J, Cui X, Li A, Li Y, Roberts C, et al. Angiopoietin1/Tie2 and VEGF/Flk1 induced by MSC treatment amplifies angiogenesis and vascular stabilization after stroke. J Cereb Blood Flow Metab. 2007; 27(10):1684–91. doi: 10.1038/sj.jcbfm.9600475 PMID: 17356562; PubMed Central PMCID: PMC2796470.

9. Bian S, Zhang L, Duan L, Wang X, Min Y, Yu H. Extracellular vesicles derived from human bone marrow mesenchymal stem cells promote angiogenesis in a rat myocardial infarction model. J Mol Med (Berl). 2014; 92(4):387–97. doi: 10.1007/s00109-013-1110-5 PMID: 24337504.

10. Nassiri SM, Rahbarghazi R. Interactions of mesenchymal stem cells with endothelial cells. Stem Cells Dev. 2014; 23(4):319–32. doi: 10.1089/scd.2013.0419 PMID: 24171705.

11. Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, et al. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy. 2006; 8(4):315–7. doi: 10.1080/14653240600855905 PMID: 16923606.

12. Delorme B, Ringe J, Pontikoglou C, Gaillard J, Langonne A, Sensebe L, et al. Specific lineage-priming of bone marrow mesenchymal stem cells provides the molecular framework for their plasticity. Stem Cells. 2009; 27(5):1142–51. doi: 10.1002/stem.34 PMID: 19418444.

13. Kaltz N, Ringe J, Holzwarth C, Charbord P, Niemeyer M, Jacobs VR, et al. Novel markers of mesenchymal stem cells defined by genome-wide gene expression analysis of stromal cells from different sources. Exp Cell Res. 2010; 316(16):2609–17. doi: 10.1016/j.yexcr.2010.06.002 PMID: 20599957.

14. Rostovskaya M, Anastassiadis K. Differential expression of surface markers in mouse bone marrow mesenchymal stromal cell subpopulations with distinct lineage commitment. PLoS One. 2012; 7(12): e51221. doi: 10.1371/journal.pone.0051221 PMID: 23236457; PubMed Central PMCID: PMC3517475.

15. Pevsner-Fischer M, Levin S, Zipori D. The origins of mesenchymal stromal cell heterogeneity. Stem Cell Rev. 2011; 7(3):560–8. doi: 10.1007/s12015-011-9229-7 PMID: 21437576.

16. Roobrouck VD, Ulloa-Montoya F, Verfaillie CM. Self-renewal and differentiation capacity of young and aged stem cells. Exp Cell Res. 2008; 314(9):1937–44. doi: 10.1016/j.yexcr.2008.03.006 PMID: 18439579.

17. Xu J, Woods CR, Mora AL, Joodi R, Brigham KL, Iyer S, et al. Prevention of endotoxin-induced systemic response by bone marrow-derived mesenchymal stem cells in mice. Am J Physiol Lung Cell Mol Physiol. 2007; 293(1):L131–41. doi: 10.1152/ajplung.00431.2006 PMID: 17416739.

18. Gupta N, Su X, Popov B, Lee JW, Serikov V, Matthay MA. Intrapulmonary delivery of bone marrow-derived mesenchymal stem cells improves survival and attenuates endotoxin-induced acute lung injury in mice. J Immunol. 2007; 179(3):1855–63. PMID: 17641052.

19. Rojas M, Xu J, Woods CR, Mora AL, Spears W, Roman J, et al. Bone marrow-derived mesenchymal stem cells in repair of the injured lung. Am J Respir Cell Mol Biol. 2005; 33(2):145–52. doi: 10.1165/rcmb.2004-0330OC PMID: 15891110; PubMed Central PMCID: PMC2715309.

20. Ortiz LA, Dutreil M, Fattman C, Pandey AC, Torres G, Go K, et al. Interleukin 1 receptor antagonist mediates the antiinflammatory and antifibrotic effect of mesenchymal stem cells during lung injury. Proc Natl Acad Sci U S A. 2007; 104(26):11002–7. doi: 10.1073/pnas.0704421104 PMID: 17569781; PubMed Central PMCID: PMC1891813.

21. Griffith LM, Pavletic SZ, Tyndall A, Gratwohl A, Furst DE, Forman SJ, et al. Target populations in allogeneic hematopoietic cell transplantation for autoimmune diseases—a workshop accompanying: cellular therapy for treatment of autoimmune diseases, basic science and clinical studies, including new developments in hematopoietic and mesenchymal stem cell therapy. Biol Blood Marrow Transplant. 2006; 12(6):688–90. doi: 10.1016/j.bbmt.2006.02.007 PMID: 16737943.

22. Jeong JA, Hong SH, Gang EJ, Ahn C, Hwang SH, Yang IH, et al. Differential gene expression profiling of human umbilical cord blood-derived mesenchymal stem cells by DNA microarray. Stem Cells. 2005; 23(4):584–93. doi: 10.1634/stemcells.2004-0304 PMID: 15790779.

23. Phinney DG, Hill K, Michelson C, DuTreil M, Hughes C, Humphries S, et al. Biological activities encoded by the murine mesenchymal stem cell transcriptome provide a basis for their developmental potential and broad therapeutic efficacy. Stem Cells. 2006; 24(1):186–98. doi: 10.1634/stemcells.2004-0236 PMID: 16100003.

24. De Luca A, Roma C, Gallo M, Fenizia F, Bergantino F, Frezzetti D, et al. RNA-seq analysis reveals significant effects of EGFR signalling on the secretome of mesenchymal stem cells. Oncotarget. 2014; 5 (21):10518–28. PMID: 25344915; PubMed Central PMCID: PMC4279390.

25. Roson-Burgo B, Sanchez-Guijo F, Del Canizo C, De Las Rivas J. Transcriptomic portrait of human Mesenchymal Stromal/Stem cells isolated from bone marrow and placenta. BMC Genomics. 2014; 15:910. doi: 10.1186/1471-2164-15-910 PMID: 25326687; PubMed Central PMCID: PMC4287589.

26. Tsai MS, Hwang SM, Chen KD, Lee YS, Hsu LW, Chang YJ, et al. Functional network analysis of the transcriptomes of mesenchymal stem cells derived from amniotic fluid, amniotic membrane, cord blood, and bone marrow. Stem Cells. 2007; 25(10):2511–23. doi: 10.1634/stemcells.2007-0023 PMID: 17556597.

27. Eirin A, Riester SM, Zhu XY, Tang H, Evans JM, O'Brien D, et al. MicroRNA and mRNA cargo of extracellular vesicles from porcine adipose tissue-derived mesenchymal stem cells. Gene. 2014; 551(1):55–64. doi: 10.1016/j.gene.2014.08.041 PMID: 25158130; PubMed Central PMCID: PMC4174680.

28. Jaager K, Fatkina A, Velts A, Orav E, Neuman T. Variable expression of lineage regulators in differentiated stromal cells indicates distinct mechanisms of differentiation towards common cell fate. Gene. 2014; 533(1):173–9. doi: 10.1016/j.gene.2013.09.094 PMID: 24103479.

29. Li J, Ma Y, Teng R, Guan Q, Lang J, Fang J, et al. Transcriptional profiling reveals crosstalk between mesenchymal stem cells and endothelial cells promoting prevascularization by reciprocal mechanisms. Stem Cells Dev. 2015; 24(5):610–23. doi: 10.1089/scd.2014.0330 PMID: 25299975; PubMed Central PMCID: PMC4333618.

30. Charbord P. Bone marrow mesenchymal stem cells: historical overview and concepts. Hum Gene Ther. 2010; 21(9):1045–56. doi: 10.1089/hum.2010.115 PMID: 20565251.

31. Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C, et al. Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. 1997; 11(6):774–85. PMID: 9087431.

32. Huang S. Reprogramming cell fates: reconciling rarity with robustness. Bioessays. 2009; 31(5):546–60. doi: 10.1002/bies.200800189 PMID: 19319911.

33. Huang S, Guo YP, May G, Enver T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev Biol. 2007; 305(2):695–713. doi: 10.1016/j.ydbio.2007.02.036 PMID: 17412320.

34. Kalisky T, Quake SR. Single-cell genomics. Nat Methods. 2011; 8(4):311–4. doi: 10.1038/nmeth0411-311 PMID: 21451520.

35. Zhang W, Ou G, Hamrick M, Hill W, Borke J, Wenger K, et al. Age-related changes in the osteogenic differentiation potential of mouse bone marrow stromal cells. J Bone Miner Res. 2008; 23(7):1118–28. doi: 10.1359/jbmr.080304 PMID: 18435580; PubMed Central PMCID: PMC2679384.

36. Trivedi P, Hematti P. Simultaneous generation of CD34+ primitive hematopoietic cells and CD73+ mesenchymal stem cells from human embryonic stem cells cocultured with murine OP9 stromal cells. Exp Hematol. 2007; 35(1):146–54. Epub 2007/01/03. S0301-472X(06)00584-4 [pii] doi: 10.1016/j.exphem.2006.09.003 PMID: 17198883.

37. Delcarpio JB, Lanson NA Jr., Field LJ, Claycomb WC. Morphological characterization of cardiomyocytes isolated from a transplantable cardiac tumor derived from transgenic mouse atria (AT-1 cells). Circ Res. 1991; 69(6):1591–600. PMID: 1954678.

38. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7(3):562–78. doi: 10.1038/nprot.2012.016 PMID: 22383036; PubMed Central PMCID: PMC3334321.

39. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37(1):1–13. doi: 10.1093/nar/gkn923 PMID: 19033363; PubMed Central PMCID: PMC2615629.

40. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956.

41. Tsujita Y, Muraski J, Shiraishi I, Kato T, Kajstura J, Anversa P, et al. Nuclear targeting of Akt antagonizes aspects of cardiomyocyte hypertrophy. Proc Natl Acad Sci U S A. 2006; 103(32):11946–51. doi: 10.1073/pnas.0510138103 PMID: 16882732; PubMed Central PMCID: PMC1567678.

42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25(4):402–8. doi: 10.1006/meth.2001.1262 PMID: 11846609.

43. Kolf CM, Cho E, Tuan RS. Mesenchymal stromal cells. Biology of adult mesenchymal stem cells: regulation of niche, self-renewal and differentiation. Arthritis Res Ther. 2007; 9(1):204. doi: 10.1186/ar2116 PMID: 17316462; PubMed Central PMCID: PMC1860068.

44. Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell. 2008; 2 (2):151–9. doi: 10.1016/j.stem.2008.01.004 PMID: 18371436; PubMed Central PMCID: PMC2276627.

45. Foudah D, Monfrini M, Donzelli E, Niada S, Brini AT, Orciani M, et al. Expression of neural markers by undifferentiated mesenchymal-like stem cells from different sources. J Immunol Res. 2014; 2014:987678. doi: 10.1155/2014/987678 PMID: 24741639; PubMed Central PMCID: PMC3987801.

46. Foudah D, Redondo J, Caldara C, Carini F, Tredici G, Miloso M. Expression of neural markers by undifferentiated rat mesenchymal stem cells. J Biomed Biotechnol. 2012; 2012:820821. doi: 10.1155/2012/820821 PMID: 23093867; PubMed Central PMCID: PMC3474592.

47. Foudah D, Redondo J, Caldara C, Carini F, Tredici G, Miloso M. Human mesenchymal stem cells express neuronal markers after osteogenic and adipogenic differentiation. Cell Mol Biol Lett. 2013; 18 (2):163–86. doi: 10.2478/s11658-013-0083-2 PMID: 23430457.

48. Funari VA, Day A, Krakow D, Cohn ZA, Chen Z, Nelson SF, et al. Cartilage-selective genes identified in genome-scale analysis of non-cartilage and cartilage gene expression. BMC Genomics. 2007; 8:165. doi: 10.1186/1471-2164-8-165 PMID: 17565682; PubMed Central PMCID: PMC1906768.

49. James CG, Stanton LA, Agoston H, Ulici V, Underhill TM, Beier F. Genome-wide analyses of gene expression during mouse endochondral ossification. PLoS One. 2010; 5(1):e8693. doi: 10.1371/journal.pone.0008693 PMID: 20084171; PubMed Central PMCID: PMC2805713.

50. Ng F, Boucher S, Koh S, Sastry KS, Chase L, Lakshmipathy U, et al. PDGF, TGF-beta, and FGF signaling is important for differentiation and growth of mesenchymal stem cells (MSCs): transcriptional profiling can identify markers and signaling pathways important in differentiation of MSCs into adipogenic, chondrogenic, and osteogenic lineages. Blood. 2008; 112(2):295–307. doi: 10.1182/blood-2007-07-103697 PMID: 18332228.

51. Park KW, Waki H, Kim WK, Davies BS, Young SG, Parhami F, et al. The small molecule phenamil induces osteoblast differentiation and mineralization. Mol Cell Biol. 2009; 29(14):3905–14. doi: 10.1128/MCB.00002-09 PMID: 19433444; PubMed Central PMCID: PMC2704753.

52. Tanaka T, Yoshida N, Kishimoto T, Akira S. Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. EMBO J. 1997; 16(24):7432–43. doi: 10.1093/emboj/16.24.7432 PMID: 9405372; PubMed Central PMCID: PMC1170343.

53.   Zaidi M. Skeletal remodeling in health and disease. Nat Med. 2007; 13(7):791–801. doi: 10.1038/
       nm1593 PMID: 17618270.

54.   English K, Barry FP, Mahon BP. Murine mesenchymal stem cells suppress dendritic cell migration,
       maturation and antigen presentation. Immunol Lett. 2008; 115(1):50–8. doi: 10.1016/j.imlet.2007.10.
       002 PMID: 18022251.

55.   English K, French A, Wood KJ. Mesenchymal stromal cells: facilitators of successful transplantation?
       Cell Stem Cell. 2010; 7(4):431–42. doi: 10.1016/j.stem.2010.09.009 PMID: 20887949.

56.   English K, Mahon BP. Allogeneic mesenchymal stem cells: agents of immune modulation. J Cell Bio-
       chem. 2011; 112(8):1963–8. doi: 10.1002/jcb.23119 PMID: 21445861.

57.   Baksh D, Yao R, Tuan RS. Comparison of proliferative and multilineage differentiation potential of
       human mesenchymal stem cells derived from umbilical cord and bone marrow. Stem Cells. 2007; 25
       (6):1384–92. doi: 10.1634/stemcells.2006-0709 PMID: 17332507.

58.   Chateauvieux S, Ichante JL, Delorme B, Frouin V, Pietu G, Langonne A, et al. Molecular profile of
       mouse stromal mesenchymal stem cells. Physiol Genomics. 2007; 29(2):128–38. doi: 10.1152/
       physiolgenomics.00197.2006 PMID: 17179208.