Multilineage gene expression in human bone marrow stromal cells as evidenced by single-cell microarray analysis

Beerelli Seshi,* Sanjay Kumar,1 and Debra King2

Department of Interdisciplinary Oncology, H. Lee Moffitt Cancer Center and Research Institute, University of South Florida, 12902 Magnolia Drive, Tampa, FL 33612-9497, USA

Submitted 20 May 2003; revised 30 May 2003
(Communicated by M. Lichtman, M.D., 5/30/03)

Abstract
The nonhematopoietic stromal cells of the bone marrow are critical for the development of hematopoietic stem cells into functionally competent blood cells. This study addresses the question of whether bone marrow stromal cell cultures in the Dexter system propagate multiple different mesenchymal stromal cell types or one stromal cell type that expresses multiple phenotypes. Results show that isolated single stromal cells simultaneously express transcripts associated with osteoblast, fibroblast, muscle, and adipocyte differentiation. Furthermore, isolated single stromal cells simultaneously express transcripts characteristic of epithelial cells, endothelial cells, and neural/glial cells. Isolated single stromal cells also express transcripts for CD45, CD19, CD10, CD79a, and representative proto-oncogenes and transcription factors, which are typically associated with normal and neoplastic hematopoietic cells. These findings suggest that the nonhematopoietic mesenchymal cells and the hematopoietic B-lymphocytes have a common progenitor. This is consistent with the idea that progenitor cells express genes that are characteristic of the multiple lineage paths that such cells may be capable of adopting. This study demonstrates the technical feasibility of transcriptome analysis of individual primary cell-culture grown stromal cells and supports the concept that bone marrow stromal cells are relatively homogeneous and show a phenotypic signature of potential multilineage differentiation capacity.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Bone marrow stromal cell; Mesenchymal progenitor cell; Single-cell microarray analysis; Cellular plasticity; Transdifferentiation; Pluridifferentiated; Multidifferentiated; Pluripotentiality

Introduction
Dexter-type stromal cell appears to be more homogeneous in morphology and phenotype than previously reported, and this study is designed to explore the expression of lineage-related genes in single Dexter stromal cells. Seeking to understand the molecular mechanism of cellular plasticity, this study addresses the related question: Is there an intermediate stage between the stem cell and a single-lineage cell that is destined to perform specialized functions? Theoretically, (a) a stem cell can directly become a terminally differentiated cell or (b) a stem cell can enter a phase of multilineage differentiation before becoming a

* An abstract of this work was presented at the NCI Workshop on Hematologic Malignancies and the Marrow Microenvironment, the National Cancer Institute, September 23–25, 2002, Rockville, MD. An abstract of this work (abstract 226) was also presented at the 44th American Society of Hematology Annual Meeting, December 6–10, 2002, Philadelphia, PA. Dr. Seshi conceived the study, put together the methods, analyzed the data, and wrote the article. Dr. Kumar performed microdissection of cells, isolation and amplification of RNA, and in vitro transcription. Ms. King performed cell culture and microdissection. All authors read and approved the article.

†† Supplementary data associated with this article can be found on ScienceDirect at doi: 10.1016/S1079-9796(03)00150-5.

* Corresponding author. Department of Pathology, Box 12, Harbor-UCLA Medical Center and Research and Education Institute, David Geffen School of Medicine at UCLA, 1000 West Carson Street, Torrance, CA 90509-2910. Fax: +1-310-222-8002.
E-mail address: BSeshi@rei.edu (B. Seshi).

1 Current address: Department of Pathology, University of Alabama at Birmingham, 701, 19th Street South, Birmingham, AL 35294.
2 Current address: USF Center for Biological Defense, College of Public Health/EOH, 3602 Spectrum Boulevard, Tampa, FL 33612.

doi:10.1016/S1079-9796(03)00150-5
single-lineage, mature cell. To determine between these alternatives, single marrow stromal cells were analyzed as a model. The present study uses microarray analysis at the single-cell level with direct comparison to enriched and unfractionated stromal cell samples.

Materials and methods

The present study involved microarray analysis of 23 samples and a corresponding number of chips. The marrow samples were obtained from 4 normal healthy adult human subjects and consisted of mixtures of unfractionated stromal cell samples. The marrow samples were obtained from 4 normal healthy adult human subjects and consisted of mixtures of unfractionated stromal cell samples. The marrow samples were obtained from 4 normal healthy adult human subjects and consisted of mixtures of unfractionated stromal cell samples.

Materials and methods

The present study involved microarray analysis of 23 samples and a corresponding number of chips. The marrow samples were obtained from 4 normal healthy adult human subjects and consisted of mixtures of unfractionated stromal cells (collective USCs or cUSCs; 8 samples), Percoll gradient-purified mesenchymal progenitor cells (collective MPCs or cMPCs; 5 samples), and single-cell MPCs (sMPCs; 10 samples) obtained by laser-capture microdissection (LCM) [1], ensuring adequate controls and replicates. The isolated single stromal cells were selected on the basis of morphology. Wright–Giemsa (or hematoxylin) stained cytospin preparation revealed characteristically large nonhematopoietic cells with a relatively irregular nucleus and cytoplasm compartmentalized into ectoplasm and endoplasm [2]. Hematoxylin stain is simpler to use, provides morphologic detail sufficient to allow recognition and isolation of these cells by laser capture microdissection, and does not interfere with downstream microarray testing (see Detailed materials and methods at the end of the article). The photomicrographs of 10 stromal cells that have been subjected to microarray testing are shown in Fig. 1. As characterized earlier using immunocytochemical staining [2], the stromal cells targeted for microarray analysis were CD45-negative cells, thus separating them from contaminating hematopoietic cells. To serve as controls and facilitate comparison, we analyzed side by side 8 samples of unfractionated stromal cells that were “contaminated” by up to 35% macrophages and 5% hematopoietic cells (cUSC) and 5 samples of Percoll-gradient purified stromal cells that were up to 95% pure (cMPC) [2]. RNA isolated from sMPC samples was subjected to two rounds of amplification using a RiboAmp kit (Arcturus, Inc., Mountain View, CA) before in vitro transcription (IVT). In contrast, RNA samples isolated from cUSCs and cMPCs were used without amplification for IVT. Except for this difference, the steps of microarray testing were standard for all three types of samples and are outlined as follows: preparation of total RNA → generation of cDNA → preparation of ds cDNA → in vitro transcription into cRNA → fragmentation of cRNA → hybridization of target RNA to a microarray of known genes (Affymetrix U95Av2 oligonucleotide microarray, with 12,625 probe sets) → signal quantification and first-tier analysis using the microarray quantification software (Microarray Suite MAS v.5, Affymetrix, Inc., Santa Clara, CA). According to the statistical expression analysis algorithm implemented in MAS v.5, the presence of a gene within a given sample was determined at a detection p value of < 0.05 and was graded as absent (A), marginal (M), or present/positive (P).

Results

Morphologic and phenotypic characterization of cell populations in Dexter cultures

BM stromal cell cultures grown under Dexter conditions (i.e., in the presence of hydrocortisone and horse serum) are generally considered to be heterogeneous. Earlier published work, however, showed that Dexter cultures are not heterogeneous based on light microscopic, ultrastructural, phenotypic, and molecular biological characteristics of non-hematopoietic stromal cells isolated from these cultures [2]. Detailed characteristics of constituent cell populations in Dexter cultures were published previously [2] and show that nonhematopoietic stromal cells (sMPCs) are morphologically and phenotypically uniform. Their morphologic characteristics are summarized as follows: The sMPCs are large cells with relatively large irregular nucleus and abundant cytoplasm that is uniquely compartmentalized into ectoplasm and endoplasm. Macrophages are large cells as well; however, they have a very small round bulletlike nucleus and foamy cytoplasm. In contrast, hematopoietic cells are small cells with minimal amount of cytoplasm. Our earlier study used Wright–Giemsa stain; comparable data are presented here using hematoxylin staining (Fig. 1) before laser capture microdissection (LCM). To further characterize these cells, a Percoll gradient technique was devised for enrichment of nonhematopoietic stromal cells (Fig. 8, under Detailed materials and methods). Whereas the unfractionated samples contained on average 40% contaminating cells (35% macrophages + 5% hematopoietic cells), the Percoll gradient-enriched samples contained on average 5% contamination (macrophages + hematopoietic cells).

Assessment of RNA amplification method

The single-cell microarray data were reviewed for reproducibility and validity. Two important statistics, reflecting on the reliability of the RNA amplification step, were evaluated.

3′:5′ ratios of housekeeping control genes

As shown in Table 1, these ratios were close to 1 in the standard unamplified samples, but were increased in the amplified samples. This may reflect preferential amplification toward the 3′ end since amplification may not proceed all the way to the 5′ end. Alternatively, it may reflect sample degradation. The 3′:5′ ratios were highly variable across single-cell MPC samples in the case of ACTB (β-actin), but were relatively close in the case of GAPD (glyceraldehyde-3-phosphate dehydrogenase), suggesting that sample degradation did not occur. Furthermore, both GAPD and ACTB
gene probes used as part of the standard gene probe set yielded relatively stable signals across replicates in each sample type, providing further evidence of intactness of RNA samples targeted for microarray analysis.

As outlined in Table 1, the amplified single-cell MPC RNA samples expectedly showed significantly lower numbers of genes compared to the standard RNA samples (av-

**Fig. 1.** Photomicrographs of single-cell MPCs that were isolated by laser capture microdissection (LCM) and subsequently targeted for microarray analysis. sMPCs are readily distinguishable from macrophages as shown for comparison.
Table 1
Summary of human bone marrow stromal cell samples targeted for microarray analysis with an outline of the corresponding indicators of assay quality performance

| Subject A | UNFR A | MPC A | SCA1, SCA2, SCA3 |
| Subject B | UNFR B, UNFR B R1, UNFR B R2 | MPC B R2, MPC C R2 | SCB1, SCB3 |
| Subject C | UNFR C R1, UNFR C R2 | MPC D R1, MPC D R2 | SCC1, SCC3 |
| Subject D | UNFR D R1, UNFR D R2 | No | Two rounds |

Amplification of RNA before IVT
Amplification of RNA before IVT

Number of genes present (% of 12,625)
Mean: 46.63
SD: 5.95
Mean: 46.54
SD: 3.66
Mean: 33.93
SD: 3.94
Mean: 33.93
SD: 3.94

3'5' ratio, GAPD M33197 (Probe used as part of housekeeping control gene probe set)
Mean: 0.89
SD: 0.33
Mean: 1.23
SD: 0.53
Mean: 6.76
SD: 2.97

GAPD signal 35905_s_at (Probe used as part of standard gene probe set)
Mean: 130.723
SD: 36.990
Mean: 164,587
SD: 40,204
Mean: 43,235
SD: 14,143

3'5' ratio, ACTB X00351 (Probe used as part of housekeeping control gene probe set)
Mean: 1.44
SD: 0.60
Mean: 2.29
SD: 1.57
Mean: 57.92
SD: 67.82

ACTB signal 32318_s_at (Probe used as part of standard gene probe set)
Mean: 86.104
SD: 18,458
Mean: 100,383
SD: 28,427
Mean: 884
SD: 90

Note. Replicate samples shown under each sample type as indicated correspond to each subject. The replicates of collective USC and collective MPC represent cell-culture or biological replicates of stromal cells grown in parallel flasks (instead of technical replicates). Of 27 samples, 2 collective MPC samples and 2 single cell MPC samples failed either at the test chip stage or produced unusual results in terms of the number of genes present and/or 3'5' ratios and were therefore excluded as outliers. The remaining 23 samples that were targeted for the data mining analysis are represented in this table. The statistics shown (means and SDs) were based on the number of sample replicates indicated in the top row of the table except for ACTB-signal for single cell MPC samples. Thirteen of these genes were present in at least 20 of 23 stromal cell samples, and 19 genes were detected in at least in 9 of 10 sMPC samples (i.e., in at least 20 of 23 stromal cell samples investigated). The main conclusions of the present report are based on what is referred to as "master list of stromal cell genes" that is broadly representative of all three types of stromal cell samples investigated and not on a gene list that is selective to sMPCs.

In many cases, microarray analysis is used to identify genes differentially expressed in different sample groups (i.e., treated vs untreated or normal vs diseased). In contrast, the goal in this study is to identify genes that are relatively uniformly expressed across normal untreated bone marrow stromal cell samples, regardless whether they are of single-cell type or populations of cells, purified or unpurified. As described under Detailed materials and methods, we used GeneSpring to achieve the following data-analysis objectives: (a) filtering for genes reliably detected in each sample group by eliminating the genes showing weak expression statistically close to the background estimate, (b) filtering for genes that are positive (present) across replicates in each sample group, (c) exclusion of genes with weak expression from genes present in each sample group, and (d) preparation of a master list of stromal cell genes by intersecting gene lists from step c (Fig. 2). These steps have led to identification of a list of 2,755 genes that are detected in at least 7 of 8 cUSC samples and 4 of 5 cMPC samples and 9 of 10 sMPC samples (i.e., in at least 20 of 23 stromal cell samples investigated). The main conclusions of the present report are based on what is referred to as "master list of stromal cell genes" that is broadly representative of all three types of stromal cell samples investigated and not on a gene list that is selective to sMPCs.

Not all of the 2,755 positive probes are nonredundant, as there are multiple probes for many individual genes on the chip employed. Because it was not possible to determine the actual number of genes that they represent, probe sets and genes are used interchangeably. The list of 2,755 genes in
the “master list of stromal cell genes” represents 88% of the
genomes expressed by single cells (3,124), 58% of genomes expressed
by unfractuated samples (4,761), and 53% of genomes expressed by Percoll-enriched samples (5,204). The
remaining genes expressed by collective cell samples are
probably due to contaminating cells as well as to genes whose transcripts failed to be amplified in single cell samples
by the amplification method. As indicated in the pre-
ceding section, the amplified single-cell samples detected
only ~34% of the genomes tested (12,625), as opposed to
unamplified cell samples, which detected about 46% of the
genomes tested. The “remaining genes” list contains genes
associated with myelomonocytic cells, which is consistent
with contaminating cells. The “remaining genes” list also
includes a number of mesenchymal-associated and other
genomes that failed to be amplified. As previously indicated,
the enriched samples contained only 5% contaminating cells
as opposed to unfractuated stromal cell samples, which
contained 40% contamination. In light of the high sensitiv-
ity of microarray analysis, 5% contamination is probably
still sufficient to detect some genes associated with the
contaminating cells. Contamination is recognized as a con-
 founding factor in the analysis of gene expression results
involving populations of cells; however, single cell expres-
sion profiling, as used here, is free from this artifact.

Hierarchical clustering analysis was used to construct a
bone marrow stromal cell tree for visualizing global gene
expression patterns across replicates and conditions. As
shown in Fig. 3, stromal cell genes that are expressed at a
relatively low level in amplified samples (sMPCs) are clus-
tered to the left of the gene tree; genes that are more
strongly expressed in sMPCs are prominently configured in
the middle of the gene tree; and genes that are expressed
approximately at the same level as unamplified samples
cMPCs and cUSCs) are clustered to the right of the gene
tree. Most importantly, as evident on the sample or exper-
iment tree, hierarchical clustering segregated the members
each sample type into a separate group (cMPC, cUSC,
and sMPC). Within each sample type corresponding subject
replicates clustered together (with minor exceptions). This
suggests a fairly high level of reproducibility within the data
set.

The data in Fig. 4 show that different transcripts amplify
to different extents. The expression of genes within the
stromal cell gene list ranges from 0.2 to 6 (on the log scale)
in unamplified samples (cMPC and cUSC) and from 0.02 to
36 in amplified samples (sMPC), thus showing much
greater variability in the amplified samples. The effect of
differential amplification is represented graphically using
color coding. Gene expression curves are colored (follow-
ing the linear color bar shown on the right) according to the
gene expression level in a particular single-cell sample.
SCA1. The genes detected at a low level in this sample (as
indicated in blue) are not necessarily expressed at a low
level in unamplified samples (as read by the log scale on y
axis). In fact, a significant number of them are expressed at
a high level in the unamplified samples. This finding to-
gether with the observation that amplified samples detected
about 34% of genes as opposed to unamplified samples
detecting about 46% of genes tested (Table 1), suggests that
some genes do not amplify at all by the method used,
whereas other genes amplify to a sufficient degree to be
detectable (shown in blue), whereas some other genes am-
plify to a degree equal to (in yellow) or surpassing (in red)
the amounts in the collective samples. (The curve shown in
white is the housekeeping gene, GAPD.) The statistical
algorithm utilized in the latest version of Microarray Anal-
ysis Suite (MAS v.5) determined that a gene within a given
sample was positive, regardless of grading.

Because different transcripts amplify variably, it is not
possible to make a quantitative comparison across trans-
scripts involving the amplified products. However, this does
not preclude the usefulness of the amplification method for
quantitative comparison of a particular transcript across
amplified single-cell samples. In fact, the data points of a
given expression curve in Fig. 4 are comparable within the
amplified samples, suggesting that expression of a particular
gene can be compared in different samples (i.e., normal vs
disease-associated MPCs). The reproducibility and the fi-
delity of linear amplification have been characterized pre-
viously [3]. It was observed that the spot intensities between
replicate amplified samples showed a correlation of \( r = 0.959 \) and that amplified and unamplified gene expression
ratios of mouse testis/brain showed a correlation of \( r = 0.913 \) [3]. These findings suggest that quantitative compar-
ison of differential gene expression is possible in cases
where some but not all RNA samples are amplified.

**Documentation of statistical variation in expression of the
master list of stromal cell genes vs the complete list of
genomes present on the chip**

There are multiple ways in which the genes of interest
can be selected for further study after microarray testing. As
outlined above, the stromal cell genes in the master list were
selected on the basis of their positive calls in at least 20 of
23 samples investigated. By plotting the mean expression
levels vs the standard deviation of the log-transformed data,
the statistical relationship between the expression levels vs
the background variation was determined for the master list
of genes and for the complete list of genes tested. As shown
in Fig. 5, the overall variation in the complete list of genes
showed a negative trend with decrease in the variation as the
mean signal strength increased. This result was observed
with all three types of samples investigated. In contrast,
similar plots involving the master list of stromal cell genes
showed flat curves with the random variation or error being
relatively constant, suggesting greater reliability of their
measurements. Also, majority of genes with weak expres-
sion have been excluded from the master list, as evident
from contrasting the mean expression levels shown on the
horizontal axes for complete list of genes vs master list of
genes for all three types of samples. These observations would agree with the fact that the genes within the master list were to begin with uniformly present or expressed in at least 20 of 23 samples tested.

**Multilineage gene expression in single stromal cells**

A stromal cell gene list, generated as outlined above, is expected to be representative of typical stromal cell gene expression profile. Such master list of genes forms the basis for derivation of all other stromal cell gene lists, organized in accordance with lineage or functional categories. As depicted in Fig. 6A and B and outlined in Table 2 (A–D), these findings show that isolated single cells simultaneously express genes associated with diverse mesenchymal cell lineages (namely osteoblast, muscle, fibroblast, and adipocyte), suggesting the existence of a pluridifferentiated mesenchymal progenitor cell type. An alternative interpretation of these findings is that the sensitive amplification/microarray approach detects levels of transcripts that are not physiologically relevant and may therefore detect “leaky” transcriptional regulation in these cells. Although “leaky” transcriptional regulation is possible, it is unlikely to be the case with sMPCs because the genes that formed the major basis for our conclusions are active not only in the amplified samples but also in the unamplified samples, ensuring that the results were not unduly biased by low-level expression occurring only in the single-cell samples.

As evident from the other gene lists (Table 2, E–G), an isolated single stromal cell simultaneously expresses transcripts for epithelial, endothelial, and neural cell types as well, widening its transcriptomic repertoire. Furthermore, as shown in Fig. 7, and Table 2 H–J, and Table 3, an isolated single stromal cell expresses transcripts that are typical of hematopoietic cells, in particular precursor B cells. This result supports the idea that the MPCs within the Dexter system might represent a form or stage of the progenitor cell that is common to nonhematopoietic and hematopoietic cells. That BM stromal cells express CD10 (CALLA) is not novel because BM stromal cells [4] as well as endometrial stromal cells [5] and normal breast myoepithelial cells [6] are known to express CD10. However, the simultaneous expression of CD19, CD79A, and immunoglobulin enhancer binding factors E12/E47 (proto-oncogene TCF3) by BM stromal cells is an unforeseen finding and forms the basis for postulating the existence of a common progenitor with B-cell lineage. B-cell progenitors typically display the phenotype (CD45 +/−, CD34 +/−, CD20 +/−), (CD10 +, CD19 +, CD79A +, HLA-DR +), which is also displayed by isolated single stromal cells at least at the transcriptome level (Table 3). Primitive CD34 + B-cell precursors (so-called Whitlock–Witte initiating cells [7]) express the human homolog of the Drosophila Polycomb group gene, BMI1, which appears to be essential for the maintenance and proliferation of hematopoietic stem cells [8,9]. As reported here, isolated single stromal cells also express BMI1 gene (Table 3 and Table S12).

CD45 positivity by cMPC and cUSC samples is attributable to coexisting or “contaminating” hematopoietic cells in these samples as evidenced by concurrent positivity for myelomonocytic markers CD13, CD33, and CD14 (Table 3). However, a similar explanation cannot be used in the case of isolated single stromal cells. Despite expression of numerous myeloid-associated proto-oncogenes and transcription factors, none of the typical myelomonocytic markers (e.g., CD13, CD33 and CD14) was identified in isolated single stromal cells. Similarly, other than CD4, no typical pan T-cell lineage markers (e.g., CD5 and CD7) were detected in stromal cells. CD3 α- and β-genese were not part of the gene chip used and therefore not tested.

The protein products or transcripts for CD45 and CD19 are most likely present in stromal cells at a basal level that...
is beyond the detection limits of conventional techniques, e.g., immunocytochemistry and Northern blotting. Correlation between transcriptome and proteome is estimated to be 0.48–0.76 [10], accounting for the discrepancy in findings by conventional techniques vs sensitive amplification/microarray analysis. Conceivably, two rounds of amplification prior to IVT sufficiently increased their transcript levels to be detected by microarray analysis. In fact, the CD45 levels were severalfold lower in cMPC and cUSC compared to CD45 levels in sMPC, and CD19 was undetectable in uncontrolled stromal cell samples (cMPC), collective unpurified stromal cell samples (cUSC), and single-cell stromal cell samples (sMPC). Individual samples are represented on the x axis (no attempt has been made to identify them by specific names). Signal intensity of a transcript in log scale (normalized across 23 samples) is shown on the y axis.

Table 2
Stromal cell gene lists associated with diverse cellular lineages

<table>
<thead>
<tr>
<th>Cell lineage</th>
<th>Representative examples of associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Osteoblast (Table S3)</td>
<td>Cadherin 11 (type 2, OB-cadherin, osteoblast), osteonectin, osteopontin, osteoblast specific factor 2 (fascinil-1-like), chondroitin sulfate proteoglycan 2 (versican), biglycan, hanamcan, collagen, type I, α2 (osteogenesis imperfecta)</td>
</tr>
<tr>
<td>B. Muscle (Table S4)</td>
<td>Various types of myosin, tropomyosin 1 and 2, transfelin, transfelin 2, caldesmon 1, dystrophin, dystroglycan 1, Fukuyama type congenital muscular dystrophy (fukutin), ATPase (Ca(^{2+}) transporting, cardiac muscle, slow twitch 2/Darier disease), capping protein (actin filament) muscle Z-line (α2 and β)</td>
</tr>
<tr>
<td>C. Fibroblast (Table S5)</td>
<td>Prolyl 4-hydroxylase, fibronectin, fibrillin 1, fibrillin, α1 collagen type IV gene, fibroblast growth factor 7 (keratinocyte growth factor) and periodontal ligament fibroblast protein</td>
</tr>
<tr>
<td>D. Adipocyte (Table S6)</td>
<td>Adipose differentiation-related protein (adipophilin), adipins, lipid desaturase, ECM protein 2 (adipocyte specific), vigilin, neadin and perilipin</td>
</tr>
<tr>
<td>E. Epithelial cell/carcinoma</td>
<td>Cytokeratin 10, keratin (hair, basic, 6, monleithrix), epithelial membrane protein 1, epithelial membrane protein 2, bullous pemphigoid antigen 1, milk fat globule-EGF factor 8 protein (lactadherin), breast epithelial mucin-associated antigen, prothymosin α and thymosin β4 and β10</td>
</tr>
<tr>
<td>F. Endothelial cell/angiogenesis/vasculogenesis (Table S8 and footnote 1)</td>
<td>Angiopoietin, VEGF, VCA1M, Factor VIII-associated gene, EDF1 (endothelial differentiation-related factor 1) and EDG2 (endothelial differentiation, G-protein-coupled receptor 2)</td>
</tr>
<tr>
<td>G. Neural cell (Table S9)</td>
<td>Neuron-specific γ enolase, GABA receptor-associated proteins, NCAM, N-cadherin, presenilin 1, Huntingtin-interacting protein, adenomedullin, azothophin, brain-derived neurotrophic factor (BDNF), syntaxin binding protein 1, peripheral myelin protein 22, ankyrin 3 (node of Ranvier, ankyrin G), glial maturation factor, β</td>
</tr>
<tr>
<td>H. Myeloid cell/myeloid leukemia (Table S10 and footnote 2)</td>
<td>MLLT2 (mixed-lineage leukemia trithorax homolog, Drosophila), CBFB (core-binding factor, β-subunit), ABL proto-oncogene, MCL1 (myeloid cell leukemia sequence 1), and DEK oncogene</td>
</tr>
<tr>
<td>I. T cell/NK cell/leukemia (Table S11 and footnote 3)</td>
<td>CD4, TAX1 binding protein 1, natural killer-tumor recognition sequence, RAP1, GTP-GDP association stimulator 1</td>
</tr>
<tr>
<td>J. B-cell/B-cell neoplasms (Table S12)</td>
<td>Bruton’s tyrosine kinase-associated protein, 135 kDa (BAP135), inhibitor of Bruton’s tyrosine kinase, pre-B-cell leukemia transcription factor 3 (PBX3), B-cell RAG-associated protein, cyclin D1, BCL6, TCF 3 (E12/E47), CALLA (CD10), (CD79A, COPEB (core promoter element binding protein, expression limited to CD19+B cells and testis), protein tyrosine phosphatase, receptor type, F (similar to CD45, which is PTPRC), restin (expressed in Reed-Sternberg cells in Hodgkin’s lymphoma, known as a type of B-cell lymphoma)</td>
</tr>
</tbody>
</table>

* The above gene lists are sublists of the master list of stromal cell genes consisting of 2,755 genes. Detailed lineage-associated gene lists (Table S3–Table S12) as well as the master list of stromal cell genes with the associated Affymetrix primary data (Table S15) are presented as Excel files in online supplemental documents. Expression of no one gene defines the phenotype of a particular cell type. Simultaneous expression of a panel of lineage-related genes in single isolated cell may be viewed as the harbinger of a potential cell type. Representative examples of genes corresponding to each cell lineage are outlined in the above table.

1. The stromal cell gene list contains a number of genes that are potentially capable of causing endothelial differentiation and vasculogenesis within the narrow microenvironment; however, these genes may not necessarily be endothelial cell markers. In fact, stromal cells express a gene, EDF1, the expression of which inversely correlates with endothelial cell differentiation within the stromal cells, indicating that the endothelial cell pathway is being actively “turned off” in these cells.

2. Despite expression of numerous myeloid-associated protooncogenes and transcription factors, none of the typical myelomonocytic markers (e.g., CD13, CD33, and CD14) was identified in stromal cells.

3. Despite expression of T-cell leukemia associated protooncogenes/transcription factors, no typical pan T-cell lineage markers (e.g., CD5 and CD7), other than CD4 and occasional CD2 and CD3ε, were identified in stromal cells. CD3α and -β genes were not part of the gene chip used and therefore not tested.

Fig. 3. Two-dimensional hierarchical clustering of 2,755 stromal cell genes based on the expression profiles of 23 samples. The gene tree is displayed on top and the experiment or sample tree is shown on left. Each column represents a particular gene on the chip and each row represents a separate stromal cell sample. The log-transformed, normalized fluorescence intensity of each gene (see Detailed materials and methods for specifics) was color coded as indicated by the color bar.

Fig. 4. Composite gene-expression plots of 2,755 stromal cell genes comparing collective purified stromal cell samples (cMPC), collective unpurified stromal cell samples (cUSC), and single-cell stromal cell samples (sMPC). Individual samples are represented on the x axis (no attempt has been made to identify them by specific names). Signal intensity of a transcript in log scale (normalized across 23 samples) is shown on the y axis.
amplified samples. Note that CD45 and CD19 are not isolated examples in this regard, because we have identified at least 200 other genes that are uniquely present across sMPC samples but undetectable in cMPC and cUSC (see red circle in Fig. 2). These findings could alternatively be interpreted as evidence of lack of fidelity of the amplification method. However, 200 genes represent 1.58% the total genes tested (12,625) and 4.67% of the 4,283 genes (on average) de-
tected in the amplified samples. Even assuming this alternative interpretation is correct, the fidelity of amplification measures over 95%. We used only one gene, namely CD19, from list of genes selective to sMPCs. Even if CD19 were excluded from consideration, the conclusions would still remain unchanged because they are not based on expression of any one particular gene but rather on simultaneous expression of a panel of lineage-associated genes.

Finally, the master list of stromal cell genes contained as many as 66 human homologs of *Drosophila*/*homeotic* genes. Some of these genes are ubiquitously expressed in *Drosophila*, whereas other genes are known for their association with specific cellular pathways. As shown in Table 4, the human homologs of *Drosophila* genes, representing diverse cellular pathways, are simultaneously active in a stromal cell. This finding represents additional evidence supporting the existence of a pluridifferentiated mesenchymal progenitor cell type.

**Discussion**

A recent study from our laboratory suggested the existence of a single unique pluridifferentiated stromal mesenchymal progenitor cell (MPC) type [2]. However, the mesenchymal lineage markers used previously are difficult to assess in the same cell. A later study by Tremain et al. applied micro-serial analysis of gene expression (micro-SAGE) to determine the “transcriptome” of a single colony-forming “unit-fibroblast” derived from a population of mesenchymal stem cells (MSCs) from the Friedenstein system [11]. These MSCs (that are relatively less differentiated in comparison to MPCs in Dexter system) also contained transcripts common to bone, cartilage, muscle, epithelium, and neural cells, which supports the concept that BM stromal cells express a pluridifferentiated mesenchymal phenotype. However, the study by Tremain et al. [11] only analyzed a single colony of BM fibroblasts, CFU-F, consisting of approximately 10,000 cells. Because such a large
colony of cells is not necessarily clonal, it could potentially contain multiple discrete singly differentiated mesenchymal cell types. Another study examined a clonally derived marrow stromal cell line that expressed the genes representative of all three germ layers [12], supporting the idea of a pluripotential progenitor cell [2].

The stromal cell–B cell relationship

Evidence for a progenitor cell common to stromal cells and hematopoietic cells has been emerging in fragments from isolated reports. Singer et al. in 1984 [13], while investigating bone marrow cultures from human patients with clonal myeloproliferative disorders, showed that the nonhematopoietic stromal cells were derived from the same clonal progenitors that were involved by the hematopoietic neoplasm, revealed by G6PD marker analysis. Huss et al. in 1995 [14] by studying a canine BM stromal cell line showed that the adherent stromal cells had “turned” into nonadherent hematopoietic cells, especially when the latter were cultured in the presence of stem cell factor. Pessina et al. in 1997 [15] showed that a particular murine stromal cell line, upon stimulation with bFGF, expressed a B-cell phenotype, including CD45R and surface immunoglobulin. The present report shows for the first time that isolated single stromal cells express transcripts that are typically associated with hematopoietic lineage, namely CD45 and CD19, as well as relevant proto-oncogenes and transcription factors. These results strongly support the existence of a progenitor cell common to bone marrow stromal cells and hematopoietic cells, particularly the bone marrow-derived (B) lymphocytes. Even though many of the above genes are not unique to B cells, our conclusions are not based on expression of any one gene. Simultaneous expression of a panel of genes (CD10 +, CD19 +, CD79A +, HLA-DR +) is indeed unique to pre-B cells. To our knowledge, only pre-B cells and BM stromal cells express this composite phenotype.

The experiments presented here use gene expression analysis of isolated, single, primary, normal human bone marrow stromal cells, which are known to have a broad capacity for multilineage differentiation. The isolated cells that are the targets of the present analysis are pictured in Fig. 1. The findings support the idea that progenitor cells express genes that are characteristic of any of the lineage fates that such cells are capable of adopting. Although conversion of stromal cells into hematopoietic or B cells has not been achieved, this work complements the work by earlier investigators outlined above [13–15] and provides new evidence involving gene expression patterns for possible lineage relationship between stromal cells and hematopoietic cells. In addition, this study may provide researchers with the tools and information to facilitate a search for cell culture conditions that permit development of B cells from an isolated single stromal cell.

Pluripotentiality vs pluridifferentiation

A number of investigators have recently shown that hematopoietic stem cells and nonhematopoietic stem cells alike have the capability to transdifferentiate by turning stem cells into variety of tissues revealing their extraordinary pluripotentiality [16–22]. The technical foundations of the studies that led to the excitement about transdifferentiation or plasticity of stem or progenitor cells have been recently vigorously challenged [23–30]. Two technical artifacts that could potentially provide misleading results are as follows: (a) donor cells can adopt the phenotype of other cells by spontaneous cell fusion, making them masquerade as transdifferentiated cells [23,24], and (b) heterogeneity of

Fig. 7. Gene-expression plots of representative precursor B-lymphocyte-associated genes by collective MPCs and single-cell MPCs. Individual samples are represented on x axis. Signal intensity of a transcript in log scale (normalized across 15 samples) is shown on y axis. The CD markers that are traditionally associated with hematopoietic cells, CD45 (probe ID 40518_at), CD19 (ID 1116_at), and CD34 (ID 538_at), are expressed by sMPCs. CD45, when present, is more abundantly detected in single MPCs than in collective MPCs and is particularly noticeable by wide range of log scale for CD45. The other pre-B cell associated markers that are expressed by sMPCs are CD10 (ID 1389_at), HLA-Dr (ID 33261_at), and CD79A (ID 34391_at). Samples 1–5, respectively, represent MPC A, MPC B R2, MPC C R2, MPC D R1, and MPC D R2. Samples 6–15, respectively, represent SCA1, SCA2, SCA3, SCB1, SCB3, SCC1, SCC3, SCD1, SCD2, and SCD3.
stem cell types that preexist within different tissues also can provide misleading results [29–31]. As noted above, the present investigation involves isolated single stromal cells, specifically 10 cells from four different individuals (Fig. 1). The cell culture system has been earlier characterized at the light microscopic level, ultrastructural level and by karyotypic analysis; these analyses revealed no evidence for spontaneous cell fusion or stem cell heterogeneity [2]. Spontaneous cell fusion most likely involves monocyte/macrophages forming multinucleated giant cells [23]: however, we observed no expression of myelomonocytic marker genes by isolated single stromal cells (Table 3). The interpretation of in vivo transdifferentiation studies involves localization of different lineage cells in different tissues or organs; such a situation requires only fusion between two cell types (one donor cell and one recipient cell) for investigators to believe the artifact as transdifferentiation. The probability of an array of different cell types fusing into 1 cell, which then masquerades as a pluridifferentiated cell, in 10 of 10 cells studied, is very low.

Although numerous reviews exist on the technology of single-cell genomics, few studies have applied this technology [32–36] and, to our knowledge, this is the first report of successful application of the Affymetrix microarray analysis at the single-cell level. These experiments were facilitated by the fact that sMPCs are uniquely large cells with abundant, tightly packed cytoplasm and conceivably contain relatively large amount of starting mRNA, as, for example, compared to a lymphocyte.

Investigators have shown that BM stromal cells under select culture conditions can be turned into bona fide bone cells, muscle cells, fat cells [16,37,38], glial cells [39], and nerve cells [40,41], demonstrating their pluripotentiality. By suggesting a molecular mechanism for stromal cell plasticity, the present data support the existence of a common precursor for MPC/neural and other lineages. These results provide an independent validation of the studies on transdifferentiation, such as the extraordinary multilineage potency of BM-derived stem or progenitor cells, reported by Krause’s group [21] and Verfaillie’s group [22]. “Lineage burst” characterized by simultaneous activation of diverse differentiation pathways within the same cell appears to be the signature profile of the stromal cell, which indicates that a “pluripotent” cell is “pluridifferentiated” at least at the molecular level. These results also imply that conversion of a stromal progenitor cell into a terminally differentiated cell (such as bone cell, muscle cell, fat cell, fibroblast, etc.) would need to “turn off” the diverse cellular pathways that are simultaneously active in a stem or progenitor cell. A recent study showing a clonally derived BM stromal cell line expressed the genes representative of all three germ layers [12] provides independent support to the concept of a pluridifferentiated stromal progenitor cell [2]. Support also comes from the observation that multilineage gene expression precedes unilineage commitment in the hematopoietic system [42].

It is likely that the multipotential cells in the marrow are rare, occurring at an estimated frequency of 1 in 10^4 nucleated cells [43]. However, these cells have been culture expanded over 4 weeks. Cultured stromal cells represent the progeny of the stromal cell and not necessarily the stromal cell itself, for which no in vivo assay as yet exists.

### Table 3
Stromal cells showing expression of genes typically associated with B-cell progenitors

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Probe ID</th>
<th>Genbank ID</th>
<th>cUSC</th>
<th>cMPC</th>
<th>sMPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD45</td>
<td>40518_at</td>
<td>Y00062</td>
<td>Positive in 8/8 samples</td>
<td>Positive in 4/5 samples</td>
<td>Positive in 6/10 samples</td>
</tr>
<tr>
<td>CD34</td>
<td>538_at</td>
<td>S53911</td>
<td>5/8</td>
<td>4/5</td>
<td>4/10</td>
</tr>
<tr>
<td>CD19</td>
<td>1116_at</td>
<td>M28170</td>
<td>0/8</td>
<td>0/5</td>
<td>10/10</td>
</tr>
<tr>
<td>CD20</td>
<td>619_s_at</td>
<td>M27394</td>
<td>0/8</td>
<td>0/5</td>
<td>3/10</td>
</tr>
<tr>
<td>CD22</td>
<td>38521_at</td>
<td>X59350</td>
<td>2/8</td>
<td>0/5</td>
<td>1/10</td>
</tr>
<tr>
<td>*CD10 (CALLA)</td>
<td>1389_at</td>
<td>J03779</td>
<td>8/8</td>
<td>5/5</td>
<td>10/10</td>
</tr>
<tr>
<td>*TF3 (E2A)</td>
<td>1373_at</td>
<td>M31523</td>
<td>8/8</td>
<td>5/5</td>
<td>9/10</td>
</tr>
<tr>
<td>*CD79A (IGBP1)</td>
<td>34391_at</td>
<td>Y08915</td>
<td>8/8</td>
<td>5/5</td>
<td>9/10</td>
</tr>
<tr>
<td>*HLA class II, Dr α</td>
<td>37039_at</td>
<td>J00194</td>
<td>8/8</td>
<td>5/5</td>
<td>9/10</td>
</tr>
<tr>
<td>*HLA class II, Dr β1</td>
<td>33261_at</td>
<td>M16941</td>
<td>8/8</td>
<td>5/5</td>
<td>10/10</td>
</tr>
<tr>
<td>*β2 microglobulin</td>
<td>34644_at</td>
<td>AB021288</td>
<td>8/8</td>
<td>5/5</td>
<td>10/10</td>
</tr>
<tr>
<td>*BMI1</td>
<td>41562_at</td>
<td>L13689</td>
<td>8/8</td>
<td>5/5</td>
<td>10/10</td>
</tr>
<tr>
<td>CD2</td>
<td>40738_at</td>
<td>M16336</td>
<td>2/8</td>
<td>1/5</td>
<td>2/10</td>
</tr>
<tr>
<td>CD3e</td>
<td>36277_at</td>
<td>M23323</td>
<td>4/8</td>
<td>3/5</td>
<td>3/10</td>
</tr>
<tr>
<td>CD5</td>
<td>32953_at</td>
<td>X04391</td>
<td>0/8</td>
<td>0/5</td>
<td>0/10</td>
</tr>
<tr>
<td>CD7</td>
<td>771_s_at</td>
<td>D00749</td>
<td>0/8</td>
<td>0/5</td>
<td>0/10</td>
</tr>
<tr>
<td>CD13</td>
<td>39365_at</td>
<td>M22324</td>
<td>8/8</td>
<td>5/5</td>
<td>0/10</td>
</tr>
<tr>
<td>CD33</td>
<td>36802_at</td>
<td>M23397</td>
<td>4/8</td>
<td>2/5</td>
<td>0/10</td>
</tr>
<tr>
<td>CD14</td>
<td>36661_s_at</td>
<td>X06882</td>
<td>8/8</td>
<td>3/5</td>
<td>0/10</td>
</tr>
</tbody>
</table>

*Note.* Genes marked with asterisk (*) met the criteria for inclusion in the master list of stromal-cell genes. The table also shows that typical myelomonocytic markers (e.g., CD13, CD33, and CD14) and typical pan T-cell lineage markers (e.g., CD5 and CD7) were not detected in single stromal cells, except for occasional CD2 and CD3e. The Affymetrix primary data corresponding to the genes listed in Table 3 can be found in Table S14 (under Online supplemental documents).
suggestion that nonhematopoietic mesenchymal cells and B lymphocytes share a common precursor is based on expression of a panel of genes (CD45, CD34, CD20, CD10, CD19, CD79A, HLA-Dr) and not expression of CD19 alone. Similar ideas were expressed in a recent article [44] and the accompanying commentary [45]. Whereas this article reports that the hematopoietic stem cells of varying potential express the genes associated with a variety of nonhematopoietic cell types, the present study reports nonhematopoietic stromal progenitor cells that express the genes associated with hematopoietic cells, in particular B cells. These two reports raise the question as to how hematopoietic stem cells and nonhematopoietic stromal cells are related in terms of ontogeny.

Acknowledgments

Dr. William E. Janssen kindly made available the filters of Fenwal Bone Marrow Collection System consisting of leftover normal marrow cells, after complete filtration of the marrow samples for standard marrow transplantation. We thank Dr. Shrikant Mane for advice on RNA isolation and in vitro transcription. Microarray testing was performed at the Joint Shands Cancer Center/ICBR Microarray Core Facility at the University of Florida, Gainesville. Our thanks are due to Dr. Mick Popp and Brian Dill for performance of microarray testing and helpful discussions. Robert Sprinkle and Federico Canton provided help with the use of Microsoft Access. We thank Professor A.N.V. Rao for helpful discussions with statistical analysis. Dr. Daniel H. Ryan critically reviewed the article and made valuable comments. We thank Dr. Robert Morin for helpful editorial suggestions. This work was funded by grants awarded to Dr. Seshi from the National Heart, Lung and Blood Institute of the National Institutes of Health (R29 HL059683); the H. Lee Moffitt Cancer Center and Research Institute; and financial support from the Department of Pathology, University of South Florida.

Appendix

Detailed materials and methods

Dexter-type bone marrow stromal cell culture

This study involved bone marrow samples obtained from four healthy adults (three women and one man) ranging in age from 43 to 50 years. The subjects were qualified to donate bone marrow for transplantation in a standard clinical BMT setting. Stromal cells were cultured using BM mononuclear cells as the starting cells and following stan-

Table 4
Human homologs of Drosophila genes, representing diverse cellular pathways, are simultaneously active in a stromal cell

<table>
<thead>
<tr>
<th>Cell lineage</th>
<th>Representative examples of associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Brief description</td>
</tr>
<tr>
<td>Neural</td>
<td>SNAI2 Snail homolog 2 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>NOTCH3 Notch homolog 3 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>NUMB Numb homolog (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>HOXB2 Homeo box B2</td>
</tr>
<tr>
<td></td>
<td>TWIST Twist homolog (acrocephalosyndactyly 3) (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>ZIC1 Zic family member 1 (odd-paired homolog, Drosophila)</td>
</tr>
<tr>
<td></td>
<td>ZFHX1B Zinc finger homeobox 1b</td>
</tr>
<tr>
<td>Muscle</td>
<td>MADH3 MAD, mothers against decapentaplegic homolog 3 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>SIX1 Sine oculis homeobox homolog 1 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>FOXO1A Forkhead box O1A (rhabdomyosarcoma)</td>
</tr>
<tr>
<td></td>
<td>PMX1 Paired mesoderm homeo box 1</td>
</tr>
<tr>
<td></td>
<td>FZD7 Frizzled homolog 7 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>MBNL Muscleblind-like (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>MEOX2 Mesenchyme homeobox 2 (important regulator of myogenesis)</td>
</tr>
<tr>
<td>Adipocyte</td>
<td>DEGS Degenerative spermatocyte homolog, lipid desaturase (Drosophila); adipocyte associated</td>
</tr>
<tr>
<td>Epithelial</td>
<td>DLG5 Disks, large (Drosophila) homolog 5</td>
</tr>
<tr>
<td>Endothelial</td>
<td>MADH7 MAD, mothers against decapentaplegic homolog 7 (Drosophila)</td>
</tr>
<tr>
<td>Fibroblast</td>
<td>EYA2 Eyes absent homolog 2 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>C3F Putative protein similar to nessy (Drosophila)</td>
</tr>
<tr>
<td>Hematopoietic</td>
<td>MLLT2 Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to 2</td>
</tr>
<tr>
<td></td>
<td>SNL Singed-like (fascin homolog, sea urchin) (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>ARH2 Ariadne homolog 2 (Drosophila)</td>
</tr>
<tr>
<td></td>
<td>PBX3 Pre-B-cell leukemia transcription factor 3</td>
</tr>
<tr>
<td></td>
<td>MADH5 MAD, mothers against decapentaplegic homolog 5 (Drosophila)</td>
</tr>
</tbody>
</table>

Note. A list of human homologs of Drosophila genes, outlining more detailed descriptions of possible lineage associations is presented as an Excel file (Table S13) under Online supplemental documents. The Affymetrix primary data corresponding to the genes listed in Table 4 and Table S13 can be found in Table S15 (under Online supplemental documents).
dard protocols as have been ongoing in our laboratory, i.e., in presence of hydrocortisone and horse serum [2, 46, 47]. The stromal cells representing mesenchymal progenitor cells (MPCs) (~95% pure) were purified or enriched as described using a discontinuous Percoll gradient after selective killing of the macrophages in stromal cultures with L-leucine methyl ester (LME, Sigma) [2]. Detailed protocols used were published earlier [2]. Briefly, the BM mononuclear cells were cultured for 4 weeks, monolayers (Fig. 8A) were trypsinized, and nonhematopoietic cells were purified by Percoll gradient (Fig. 8B) before they were cyto spun in preparation for laser capture microdissection (LCM). All samples were treated identically. The unfractionated samples contained on average 40% contaminating cells (35% macrophages + 5% hematopoietic cells), whereas Percoll gradient-enriched samples contained 5% contaminating cells (macrophages + hematopoietic cells).

**Isolation of individual MPCs using LCM**

Strict laboratory precautions were observed to ensure preservation of RNA. All buffers and solutions, e.g., phosphate-buffered saline (PBS) and ethanol solutions, contained DEPC-treated water. Before microdissection of individual stromal cells, cytopsin of Percoll-purified MPCs were prepared by attaching dispersed BM stromal cells to uncoated glass slides by low-speed (400 rpm) cytocentrifugation using a Shandon cytospin centrifuge. The cytopsins were fixed in 95% ethanol for 10 min and stained for 30 s using hematoxylin QS (Vector, Burlingame, CA) followed by washing in DEPC water. The cytopsins were then dehydrated in increasing concentration of ethanol and treated in xylene. This is a single-step staining method without involving a bluing protocol; it provided sufficient morphologic detail and did not interfere with downstream microarray analysis. The MPCs selected on the basis of morphology, as visualized on the microscope monitor, were microdissected [1] using PixCell II (Arcturus, Inc.) and captured on CapSure LCM Caps (Arcturus), followed by extraction of RNA (see next).

**Microarray sample preparation and testing**

Unless mentioned otherwise, sample preparation and microarray testing were performed according to the protocols outlined in the GeneChip Expression Analysis Technical Manual (Affymetrix, Inc., Santa Clara, CA). The RNeasy mini protocol kit (Qiagen, Valencia, CA) was used for isolation of total RNA from unfractionated stromal cells (USCs) and from Percoll-purified MPCs. Superscript II cDNA synthesis kit (Invitrogen) primed with a T7-(dT)24 primer containing a T7 RNA polymerase promoter sequence (Genset Oligos, La Jolla, CA) was employed to prepare ds cDNA from unamplified RNA samples, using 8- to 10-μg aliquots of total RNA as the template for first strand cDNA synthesis. The PicoPure RNA isolation kit (Arcturus, Inc., Mountain View, CA) was employed for extraction of RNA from LCM-dissected single-cell MPC samples. The RiboAmp RNA amplification kit (Arcturus, Inc.) [48] was used to amplify RNA from the LCM-dissected single cell samples by performing two rounds of amplification, terminating the reaction after completion of ds cDNA synthesis. The entire amplified RNA sample was used as the template for cDNA synthesis. The following steps were identical for both unamplified and amplified RNA samples. In vitro transcription (IVT) was performed in the presence of biotinylated UTP and CTP to produce biotin-labeled cRNA (Bioarray High Yield RNA Transcription labeling kit, Enzo Diagnostics, Inc., Farmingdale, NY), followed by cleaning of the reaction products with RNeasy mini kit columns (Qiagen, Valencia, CA). The purified, biotin-labeled cRNA samples were then submitted to the Microarray Core Facility at the University of Florida, Gainesville, where the following steps were performed: (a) controlled fragmentation of target cRNA in the presence of heat and Mg2+, (b) hybridization of fragmented cRNA (15 μg) for 16 h at 45°C to a microarray of known gene probes (U95Av2 oligonucleotide microarray, containing ~12,500 gene probes), (c) washing and staining of the probe array with SAPE (streptavidin phycoerythrin) (Molecular Probes, Inc., Eugene, OR), (d) array scanning with an Agilent argon-ion laser equipped with 488-nm emission and 570-nm detection wavelengths (GeneArray Scanner), and (e) background subtraction/signal quantification and first-tier analysis using the newest-version of the microarray quantification software Microarray Analysis Suite (MAS v.5, Affymetrix, Inc.). That a gene within a given a sample was absent (A), marginal (M), or present/positive (P) was determined at a detection P value of < 0.05, according to the statistical expression analysis algorithm implemented in MAS v.5. First-tier analysis also included per chip (or per sample) normalization in MAS v.5 by scaling the trimmed mean signal of a probe-array to a constant target signal value of 2,500 to facilitate comparison of the results from different samples.

**Second-tier data-analysis/data mining**

The microarray data outputted by MAS v.5 (in the form of tab delimited text files) were imported into GeneSpring software version 4.2.1 (Silicon Genetics, Redwood City, CA). Following instructions accompanying GeneSpring, each gene was normalized to itself (per-gene normalization) by using the median of the gene’s expression values over all the samples of an experimental group (or groups) and dividing each measurement for that gene by the corresponding median value, assuming that it was at least 0.01. The ratios were then log transformed to base e. No per-sample normalization was performed in GeneSpring because it was already done as part of MAS v.5 analysis. The purpose of the above data transformations, including scaling and normalization, was to remove systematic error within and across conditions or experimental groups. GeneSpring was used to achieve the following data-analysis objectives.
a. Filtering for reliably present genes by eliminating the genes showing weak expressions statistically close to the background estimate. As per the instructions accompanying GeneSpring, random error was estimated from control strength or median measurement level using the two-component global error model of Rocke–Lorenzato that assumes variability between replicates as being similar for all genes showing similar measurement levels [49]. The formula for the error model of normalized expression levels may be written as follows:

\[ S_{\text{norm}}^2 = a^2/C + b^2, \]

where \( S \) is the standard error of normalized expression data, \( a \) and \( b \) are the two error components (\( a \) is an absolute or fixed error component impacting at lower measurement values and \( b \) is a relative or proportional error component impacting at higher measurement levels), and \( C \) is the control strength. According to the manufacturer, a curve is fitted for each group of replicates, with standard error of normalized data on \( y \) axis vs control strength on \( x \) axis. At lower end of control strength, the normalized standard error would be high and as the control strength increases, the standard error would decrease, reaching a point where the curve flattens and data become more reliable. Control strength for each condition or sample group, where \( C = ab \), at which the two error components contribute equally, was calculated as follows, for collective MPCs, \( C = 128.68 \); collective USCs, \( C = 253.52 \); and single-cell MPCs, \( C = 348.32 \). Each condition was filtered for genes expressing signals greater than the respective control strength, thus eliminating the genes with weak expressions from each group. Of 12,625 gene probes tested, 7,196 genes passed the restriction in case of cMPCs, 7,287 genes in case of cUSCs, and 5,937 in case of sMPCs. Corresponding gene lists were created.

b. Filtering for genes “present” across replicates in each sample group. GeneSpring’s “Add data file restriction” feature was used to prepare the respective lists of genes that were present (or expressed or active) in at least 7 of 8 cUSC samples, 4 of 5 cMPC samples, and 9 of 10 sMPC samples. Of 12,625 genes tested, 5,204 genes passed the restriction in case of cMPCs, 7,287 genes in case of cUSCs, and 5,937 in case of sMPCs. Corresponding gene lists were created.

c. Exclusion of genes with weak expressions from genes “present” in each sample group. Respective gene lists for each sample group from steps a and b, were intersected via Venn diagrams. As a result, 5,204 genes passed the restriction in case of cMPCs, 4,761 genes in case of cUSCs, and 3,124 genes in case of sMPCs, which are almost identical to the numbers obtained as under step b, except for a difference of two genes in case of cUSCs, thus providing no significant improvement in restricting the data beyond under step b. This is a reflection of the high stringency of the
criterion used under step b. The two genes in case of cUSCs that passed the restriction under step b but failed the combined restriction under step c did show weak expressions (raw signals ranging, 142–331). Corresponding gene lists were created.

d. Preparation of the master list of stromal cell genes. Respective gene lists for the three sample groups from step c were intersected via Venn diagrams, resulting in identification of a list of 2,755 genes that are uniformly present or expressed in at least 20 of 23 stromal cell samples investigated. The stromal cell gene list thus arrived at contained genes that are representative of diverse mesenchymal lineages. Parenthetically, intersecting of gene lists corresponding to the three sample groups from step b resulted in a stromal cell gene list consisting of 2,756 genes, thus differing by 1 gene from the master list of stromal cell genes.

e. Two-way hierarchical clustering of 2,755 stromal cell genes based on expression profiles in 23 stromal cell samples. Only the data that were “cleaned up” of genes with weak expressions as outlined under step a were used for hierarchical clustering. This necessitated further processing of data in Microsoft Access before analysis using GeneSpring. The data for each individual sample as outputted by MAS v.5 contained probe IDs, quantitative and qualitative data, as well as other information such as annotations and are readily recognizable by GeneSpring. In contrast, the gene list, resulting from step a, contained only probe IDs and could not contain the data associated with each individual sample and was not recognizable by GeneSpring for inputting as part of an experiment. Therefore, the microarray data for each group of individual samples (in Excel format) as well as the corresponding gene list for that group from step a (also in Excel format) were imported into an Access database. The genes that did not pass the test under step a were deleted from the microarray data for each individual sample by querying and intersecting with the appropriate post-clean-up gene list. The resulting data files were saved first as Excel files, then resaved as tab delimited text files and then imported into GeneSpring as modified experiments. Per-gene normalization and log transformation were applied as described above. “Gene Tree” and “Experiment Tree” were constructed by applying a method similar to that of Eisen et al. [50] as implemented in GeneSpring and by using the stromal cell gene list and the following parameters: standard correlation as similarity measure, a minimum distance of 0.001, and a separation ratio of 0.5 in case of Gene Tree and 1.0 in case of Experiment Tree.

f. Preparation of stromal cell gene lists as relevant to different cellular phenotypes and/or functions. The gene lists associated with distinct mesenchymal cell lineages or phenotypes, i.e., osteoblasts, fibroblasts, muscle cells, and adipocytes, were prepared using a combination of methods. These include (1) visually inspecting the entire stromal-cell gene list for relevant key words; (2) directly searching the master list of stromal cell genes by using key words of interest via “Advanced Find Genes” feature under Edit menu in GeneSpring and by selecting “Search Only Current Gene List”; and (3) intersecting the stromal cell gene list with gene lists of interest from Gene Ontology lists, e.g., list of oncogenes, via Venn diagrams.

g. Visualization of gene-expression plots. The expression pattern of a gene across a given group (or groups) of samples of interest was pictured via Gene Inspector window, utilizing desired display options.

h. Statistical analysis of random variation in expression of the master list of stromal cell genes vs the complete list of genes tested. The master list of stromal cell genes with probe IDs from step d was imported into Microsoft Access and intersected with the table containing complete Affymetrix primary data sets (Table S16). The resulting file was exported as Excel file consisting of the master list of stromal cell genes with the associated Affy data (Table S15). The Affy data as outputted by MAS v.5 (in the form of Excel Tables S15 and S16) were then imported into ArrayStat software, Version 1.0, Rev. 2.0 (Imaging Research Inc., St. Catharines, ON, Canada). The data for each group of samples were log transformed to base 10, which allowed the software to construct the scatter plots, standard deviation vs mean (Fig. 5). To see the computer-generated labels clearly, view the figure at 100% or higher or have it printed.

i. Calculation of basic statistics for different sample groups. The mean and SD values presented as part of Tables S14, S15, and S16 were calculated using MCG ArrayStat Program (Richard A. McIndoe, URL: http://www.genomics.mcg.edu/niddk/btc/Software.htm). The accuracy of the reported mean and SD values was checked using the Excel program.

Online supplemental documents

Supplemental documents consist of the expanded tables of stromal cell gene lists and the complete Affymetrix primary data sets in Excel format (Tables S1–S16) and are accessible from the online article at ScienceDirect.

Table S1. Affymetrix (hybridization and housekeeping) positive control genes
Table S2. Stromal-derived factors (SDFs)
Table S3. Osteoblast
Table S4. Muscle
Table S5. Fibroblast
Table S6. Adipocyte
Table S7. Epithelial cell/carcinoma
Table S8. Endothelial cell/angiogenesis/vasculogenesis
Table S9. Neural cell
 Stromal genes contained at least 186 of the TGFβ/H9252 family proteins, ion channels, KIAA/cDNAs identified by Bottinger’s group), heat-shock genes/molecular chaperones, hypothetical pro-blast, G proteins, growth factors/cytokines/chemokines, macrophage, extracellular matrix, ferritin/iron storage, Drosophila genes identified related, cell cycle, human homologs of Drosophila and/or homeotic genes, endothelial cell/angiogenic/vasculogenic genes, epithelial cell/carcinoma, epithelial-mesenchymal transition (EMT)/TGFβ target genes (the master list of stromal genes contained at least 186 of the TGFβ target genes identified by Bottinger’s group [51]), erythrocyte/macrophase, extracellular matrix, ferritin/iron storage, Fibroblast, G proteins, growth factors/cytokines/chemosokines, heat-shock genes/molecular chaperones, hypothetical proteins, ion channels, KIAA/cDNAs identified in Kazusa cDNA project, kinases, lectins/galectins, lipoproteins, liver cell/liver cell disorders, major histocompatibility complex, metallocproteinases/disintegrins, mesenchymal genes/generic, morphogenesis/differentiation-related, muscle cell/muscular disorders, myeloid cell/myeloid leukemia, nerve cell/neuroendocrine cell/neurologic disorder, osteoblast/bone cell/bone disorders, phosphatases, protooncogenes, ribosomal proteins, stromal-derived factors (SDFs), T-cell/NK cell, transcription factors, tumor suppressor genes, and ubiquitin/proteasome pathway. Some of these gene lists are not entirely relevant to the central theme of the present report. Nonetheless these gene lists are important for future research in stromal cell biology and pathology. Instead of presenting the entire series of these subsections, we present the lineage-associated sublists, and the master list of stromal cell genes in the form of Excel file entitled “Master list of stromal cell genes with Affy data & group statistics” (Table S15). This will allow the readers to individualize the stromal cell gene lists to suit their own interest. We also publish the complete Affymetrix primary data sets (Table S16) so that other investigators can substantiate or check our claims. The signal or expression values shown represent the data as outputted by MAS v.5 after background subtraction, signal quantification and first-tier analysis, which is described at the end of the section “Microarray sample preparation and testing” under Detailed materials and methods. In addition to the primary data, Tables S14, S15, and S16 contain the mean and SD values for the three sample types to facilitate “eyeballing” the data.

References