Stem Cell Reports, Volume *10*

Supplemental Information

Transcriptionally and Functionally Distinct Mesenchymal Subpopulations Are Generated from Human Pluripotent Stem Cells

Chee Jia Chin, Suwen Li, Mirko Corselli, David Casero, Yuhua Zhu, Chong Bin He, Reef Hardy, Bruno Péault, and Gay M. Crooks

Figure S1

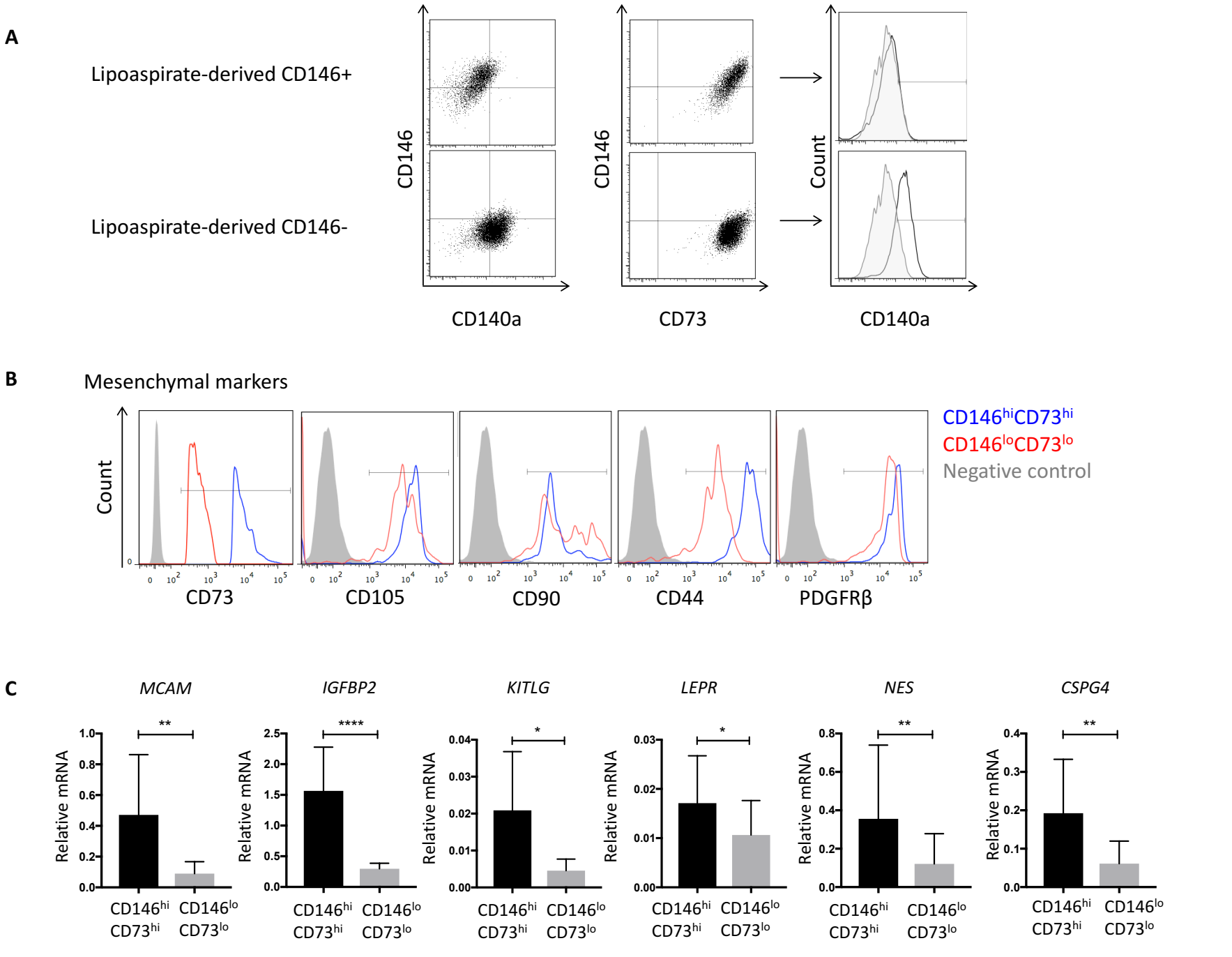
A
Lipoaspirate-derived CD146+
Lipoaspirate-derived CD146-
CD146
CD140a
CD146
CD73
Count
CD140a
B
Mesenchymal markers
Count
CD73
CD105
CD90
CD44
PDGFRβ
CD146hiCD73hi
CD146loCD73lo
Negative control
C
MCAM
IGFBP2
KITLG
LEPR
NES
CSPG4
Relative mRNA
**
****
*
*
**
**
CD146hi CD73hi
CD146lo CD73lo

Fig S1. Immunophenotypic and molecular characterization of primary and hPSC-derived mesenchymal cells. Related to Fig 1. (A) Inverse relationship of CD146 and CD73 to CD140a on mesenchymal cells from primary human tissue are similar to hPSC derived mesenchyme. Flow cytometry analysis of lipoaspirate-derived $CD146^{+}$ and $CD146^{-}$ cells, purified and cultured (passage 5-13). Unstained control shown shaded in grey. Representative data from 3 biological replicates. **(B)** $CD146^{hi}CD73^{hi}$ and $CD146^{lo}CD73^{lo}$ cells express conventional mesenchymal markers. H1 hEMP-derived mesenchymal cultures were analyzed at week 2 by FACS. All cells were gated as $CD31^{neg}$ and then either $CD146^{hi}CD73^{hi}$ (blue) or $CD146^{lo}CD73^{lo}$ (red). Data are representative of 4 independent analyses. **(C)** Expression of genes associated with the HSPC niche is enriched in the $CD146^{hi}CD73^{hi}$ population. Quantitative RT-PCR of RNA from $CD146^{hi}CD73^{hi}$ and $CD146^{lo}CD73^{lo}$ cells (means and SEM shown, n= 4-6 independent experiments. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$; paired Student's t test.

Figure S2

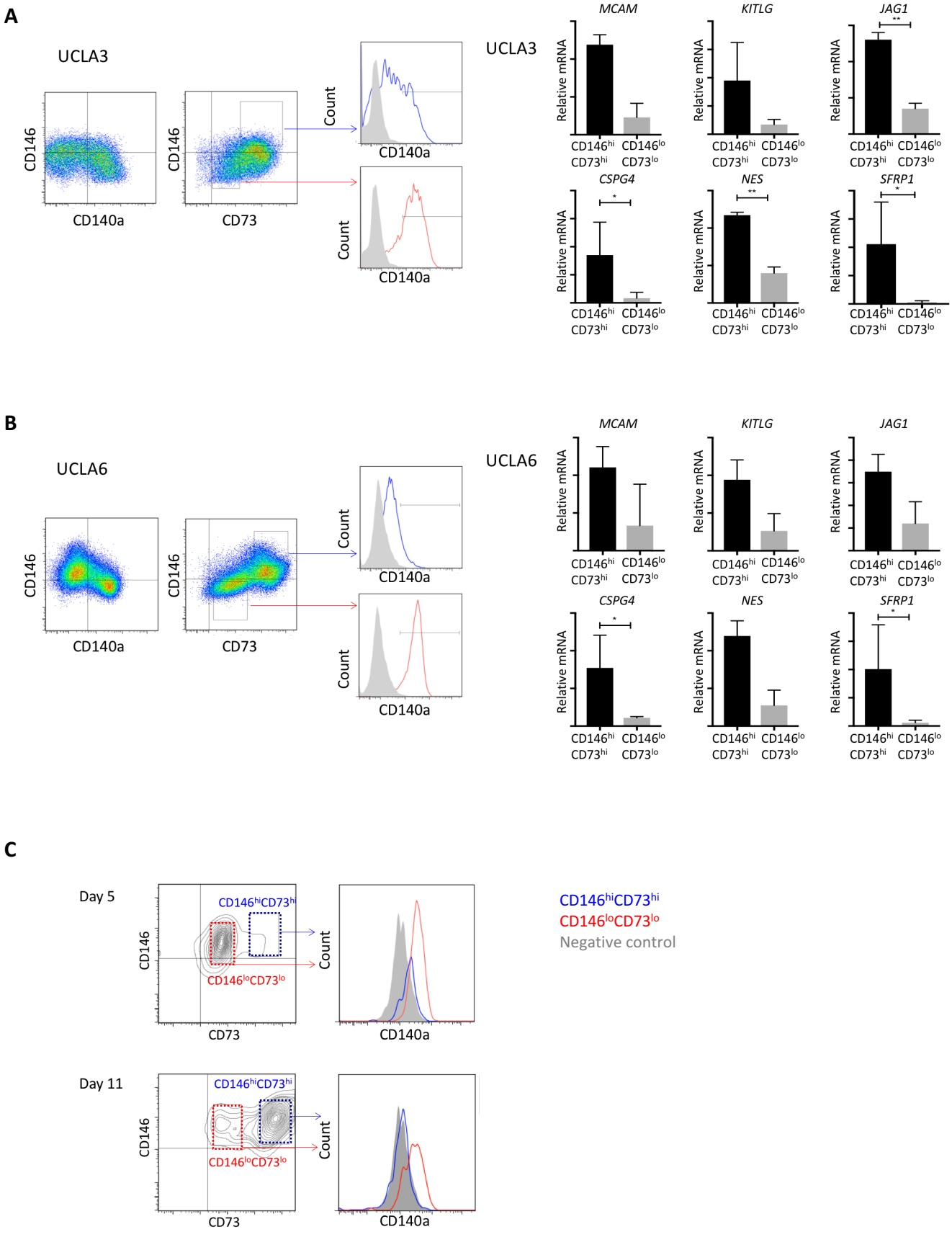
A
UCLA3
CD146
CD140a
CD146
CD73
Count
CD140a
Count
CD140a
UCLA3
MCAM
KITLG
JAG1
CSPG4
NES
SFRP1
Relative mRNA
CD146hi CD73hi
CD146lo CD73lo
B
UCLA6
CD146
CD140a
CD146
CD73
Count
CD140a
Count
CD140a
UCLA6
MCAM
KITLG
JAG1
CSPG4
NES
SFRP1
Relative mRNA
CD146hi CD73hi
CD146lo CD73lo
C
Day 5
CD146hiCD73hi
CD146loCD73lo
CD146
CD73
Count
CD140a
Day 11
CD146hiCD73hi
CD146loCD73lo
CD146
CD73
Count
CD140a
CD146hiCD73hi
CD146loCD73lo
Negative control

Fig S2. Consistency of immunophenotype and gene expression in mesenchymal subsets from additional hPSC lines and at different times during differentiation. Related to Fig 1. (A) and **(B)** Mesenchymal populations generated from two additional hESC lines demonstrate similar immunophenotype and gene expression to H1. Mesenchymal differentiation was initiated from hEMP generated from the hESC lines **(A)** UCLA3 and **(B)** UCLA6, and then analyzed after 2 weeks by Flow cytometry (representative of n=2 biological replicates per cell line) and quantitative RT-PCR (mean and SEM shown, n= 2-3 independent experiments) * $p<0.05$, ** $p<0.01$; paired two-tailed Student's t test. **(C)** Distinct mesenchymal populations with an inverse relationship of CD73 and CD140a are detectable early on during the differentiation from hEMP. Flow cytometry analysis of differentiation cultures from hEMP showed that two mesenchymal populations distinguished by CD73 expression were detected as early as day 5 of differentiation. As with later stages of differentiation, the $CD73^{++}$ population was $CD140a^{-}$ while the $CD73^{+}$ population was $CD140a^{+}$. Fluorescent minus one (FMO) for CD140a-PE was used to set gates. All cells were first gated as $CD31^{neg}$. Representative data from 3 biological replicates.

Figure S3

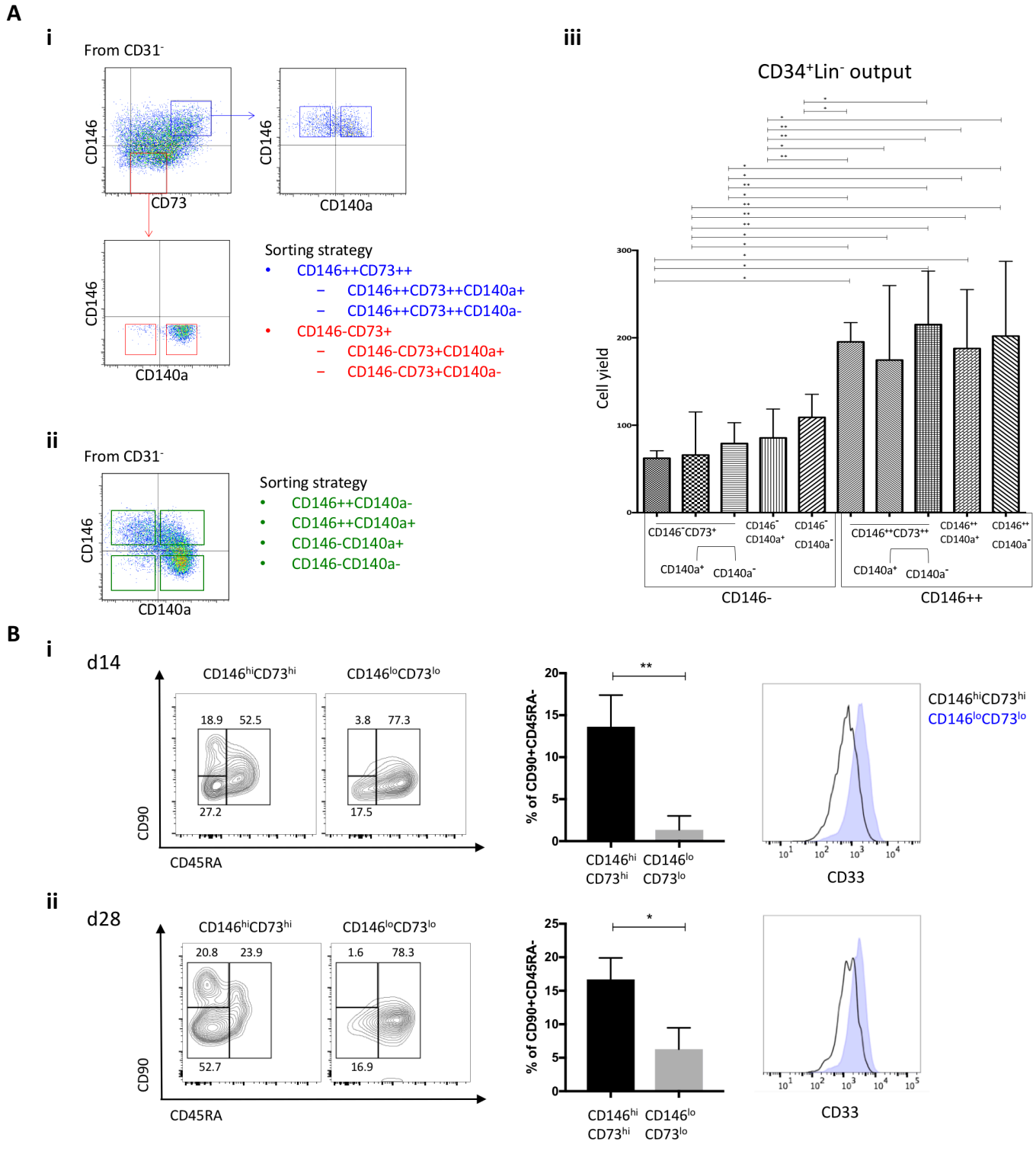
A
i
From CD31-
CD146
CD73
CD140a
Sorting strategy
CD146++CD73++
CD146++CD73++CD140a+
CD146++CD73++CD140a-
CD146-CD73+
CD146-CD73+CD140a+
CD146-CD73+CD140a-
ii
From CD31-
Sorting strategy
CD146++CD140a-
CD146++CD140a+
CD146-CD140a+
CD146-CD140a-
iii
CD34+Lin- output
Cell yield
CD146-
CD146++
B
i
d14
CD146hiCD73hi
CD146loCD73lo
18.9
52.5
27.2
3.8
77.3
17.5
CD90
CD45RA
% of CD90+CD45RA-
CD33
ii
d28
20.8
23.9
52.7
1.6
78.3
16.9

Fig S3. Analyses of alternative mesenchymal and hematopoietic populations. Related to Fig 2. (A) CD146 expression of hPSC-derived mesenchyme defines HSPC supportive capacity. Multiple different CD31- subpopulations were purified after two-week mesenchymal differentiation from H1 hEMP and tested for their ability to support HSPC *ex vivo*. **(i)** Initial gating on CD146 v CD73, then subdivided based on CD140a expression. **(ii)** Alternative gating strategy on CD146 v CD140a without CD73. **(iii)** High CD146 expression best served as a distinguishing cell surface marker to identify HSPC-supporting stroma based on CD34+lin- yield per 1000 initially seeded CD34+ cells (means and SEM shown, n=3 biological replicates, * $p<0.05$, ** $p<0.01$; unpaired two-tailed Student's t test). **(B) $CD146^{hi}CD73^{hi}$ mesenchyme supports ex *vivo* maintenance of HSC.** CD34+CD38-Lin- CB cells were isolated by flow cytometry and co-cultured on either $CD146^{hi}CD73^{hi}$ or $CD146^{lo}CD73^{lo}$ mesenchymal cells generated from H1-derived hEMP in 5% serum with the addition of Fms-related tyrosine kinase 3 ligand (50ng/ml) and thrombopoietin (50ng/ml). After **(i)** 14 or **(ii)** 28 days, hematopoietic cells were harvested and analyzed by flow cytometry. FACS plots shown were first gated on live (DAPI-) Lin-CD34+ cells. Bar graphs show the percentage of live Lin-CD34+ cells that were CD90+CD45RA- by flow cytometry at each time point (mean and SEM shown, n=4 biological replicates, * $p<0.05$, ** $p<0.01$; paired two-tailed Student's t test). On the right is the expression of CD33 of the Lin-CD34+ cells from co-cultures. $CD146^{hi}CD73^{hi}$ shown as solid black line and $CD146^{lo}CD73^{lo}$ shown as shaded. **(C) $CD146^{hi}CD73^{hi}$ mesenchyme supports engraftable HSPCs.** Bone marrow engraftment (% human cells) of mice 6 weeks after primary and secondary transplants. CB $CD34^{+}$ cells were co-cultured with $CD146^{hi}CD73^{hi}$ or $CD146^{lo}CD73^{lo}$ monolayers for 2 weeks. An equal number of $CD45^{+}$ cells (1×10^{5}) harvested from each co-culture was intratibially injected into each sublethally irradiated (250 cGy), 6- to 8-week-old NSG mouse (The Jackson Laboratory). Control animals received 1×10^{5} cells fresh, enriched CD34+ cells from the same CB (CB w/o culture). (means and SEM shown, n = 5 independent experiments, total 9-14 mice per group; ***P <0.0001).

Figure S4

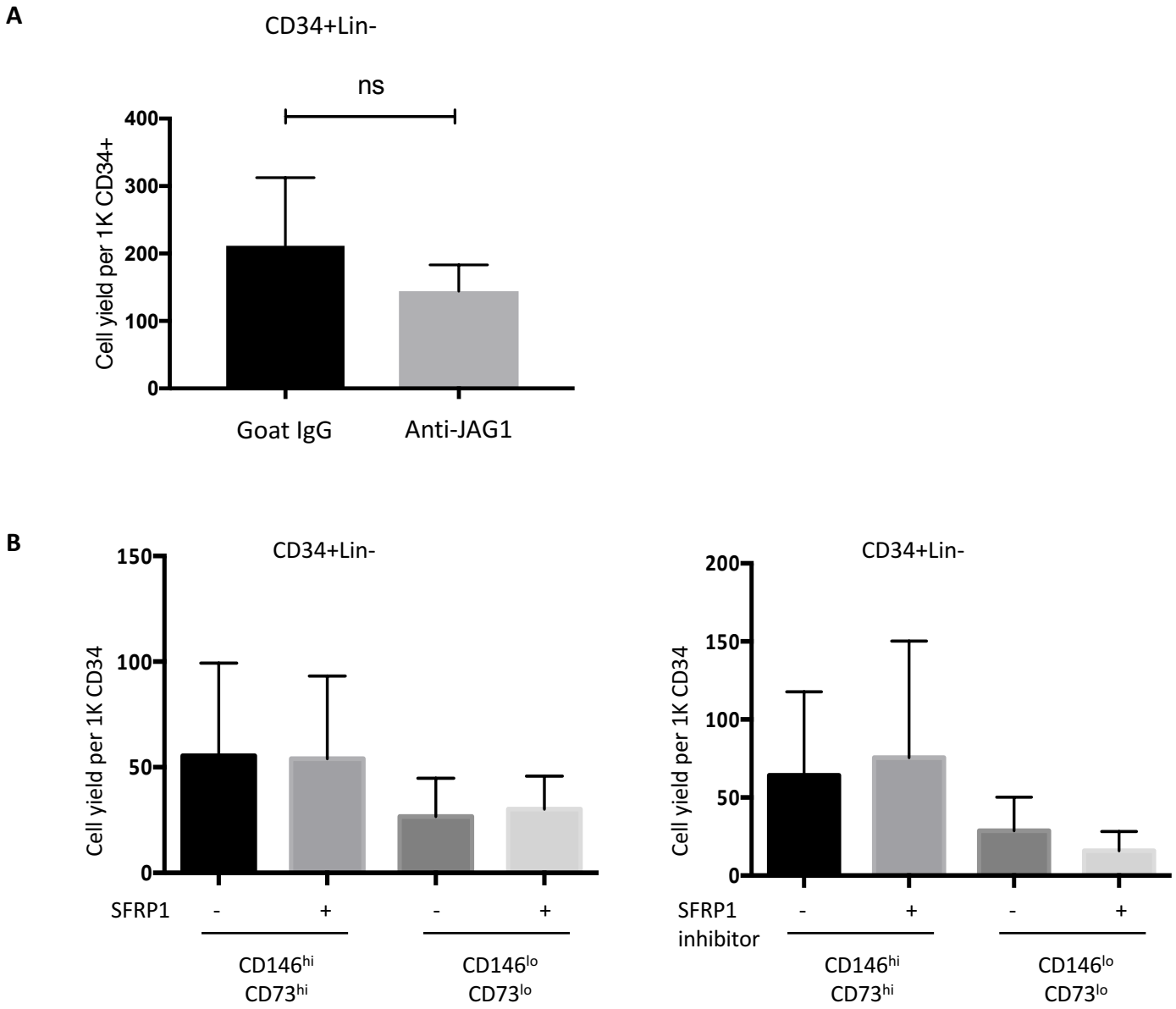
A
CD34+Lin-
ns
Cell yield per 1K CD34+
400
300
200
100
0
Goat IgG
Anti-JAG1
B
CD34+Lin-
Cell yield per 1K CD34
150
100
50
0
SFRP1
-
+
-
+
CD146hi
CD73hi
CD146lo
CD73lo
CD34+Lin-
Cell yield per 1K CD34
200
150
100
50
0
SFRP1
inhibitor
-
+
-
+
CD146hi
CD73hi
CD146lo
CD73lo

Fig S4. Mechanisms of HSPC support of hPSC-derived mesenchymes. Related to Fig 3. (A) HSPC support on OP9 stroma is not significantly affected by JAG1 blockade. Co-culture was performed as in Fig 3D using OP9 cells instead of hPSC-derived mesenchyme. Shown is CD34+Lin− cell number with addition of Goat IgG or human JAG1 blocking antibody to hPSC-mesenchyme co-culture (mean and SEM shown, n=4 independent experiments). **(B)** Neither over-expression nor inhibition of the WNT inhibitor SFRP1 in hPSC-mesenchyme co-cultures had a significant effect on HSPC *ex vivo* maintenance. Co-culture of CB-derived CD34+ with hPSC-derived $CD146^{hi}CD73^{hi}$ and $CD146^{lo}CD73^{lo}$ mesenchymal cells was performed as described previously. Shown is number of CD34+Lin− cells after addition of soluble SFRP1 or SFPR1 inhibitor (see supplemental methods) (mean and SEM are shown, n=3 independent experiments).

Table S1. Highly regulated genes shared by primary/hPSC derived cells. Related to Fig 4. (see excel file Table S1) Gene set enrichment analysis (GSEA) identifies a core set of genes differentially regulated between CD146+ vs. CD146- cells. A gene signature was defined from RNA-Seq of hPSC-derived mesenchymal cells (at least 2-fold up-, down-regulated between $CD146^{hi}CD73^{hi}$ vs. $CD146^{lo}CD73^{lo}$ cells, FDR<10%). The gene signature was compared to RNA-Seq fold changes from lipoaspirate-derived CD146+ and CD146- cells. GSEA was used to identify genes similarly regulated between hPSC-derived and A) freshly purified AND/OR B) cultured lipoaspirate-derived cells. Genes in the table represent the intersection between A) AND B).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

hESC culture and mesoderm induction conditions

The hESC lines H1 (WiCell, Madison, WI), UCLA3 and UCLA6 (UCLA BSCRC hESC core) were maintained and expanded on irradiated primary mouse embryonic fibroblasts (EMD Millipore, Billerica, MA). To induce mesoderm differentiation, hESC colonies were cut into uniform-sized pieces using the StemProEZPassage tool (Invitrogen, Thermo Fisher Scientific, Waltham, MA), transferred into 6-well plates pre-coated for 1 hour with Matrigel (growth factor reduced, no phenol red; BD Biosciences, San Jose, CA), and cultured initially in mTESR medium (Stem Cell Technologies, Vancouver, BC) until 50–60% confluent (typically 3 days). To initiate mesoderm differentiation, mTESR medium was replaced with basal induction medium X-Vivo 15 (Lonza, Walkersville, MD). Basal induction medium was supplemented with human growth factors BMP4, VEGF and bFGF for 3.5 days (all at 10 ng/mL; R&D Systems, Minneapolis, MN), with the inclusion of activin A only on day 1 ("A-BVF" condition (Evseenko et al., 2010)). After 3.5 days of mesoderm induction, $CD326^{-}CD56^{+}$ human embryonic mesoderm progenitors (hEMP) were isolated by flow cytometry (typically 5-10% of total population, **Figure 1A**) and placed onto matrigel coated 6 well plates for mesenchymal differentiation over the next 14-18 days.

Flow cytometry and cell sorting

Flow cytometry analysis was performed on LSRII or LSRFortessa and cell isolation on a FACSAria II (all from Becton Dickinson). Cultured cells were dissociated into single cell suspension with Accutase (Innovative Cell Technologies, San Diego, CA) and immunofluorescence staining was performed with human-specific monoclonal antibodies. Nonspecific binding was blocked with intravenous immunoglobulin (1%) (CSL Behring, King of Prussia, PA) before staining with fluorochrome-conjugated antibodies. Unstained cells and fluorescence-minus-one (FMO) were used as controls for gating. Cell acquisition was performed using FACSDiva (Becton Dickinson) and the analysis was performed using FlowJo (Tree Star, Ashland, OR). Forward and side scatter (FSC/SSC) and 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen) were used to identify live cells.

Immunophenotype analysis of hEMP-derived mesenchyme

hEMP-derived mesenchyme at d14-18 of differentiation were harvested and analyzed on an LSR II flow cytometer (Becton Dickinson). Cells were stained with monoclonal antibodies: CD146-FITC (AbD Serotec, MCA2141FT) (for isolation) or CD146-BV711 (BD Biosciences, 563186) (for analysis), CD73-PE Cy7 (BD Biosciences, 561258), CD45-BV711 (Biolegend, 304050), CD31-APC (Biolegend, 303116), CD144-PE (BD Biosciences, 560410), CD34-PE Cy7 (BD Biosciences, 348791), CD105-PE (Biolegend, 323206), CD90-PE (Biolegend, 328110), CD44-PE (Biolegend, 338808), CD140a-PE (BD Biosciences, 556002), CD140b-PE (Biolegend, 323606), CSPG4-PE (BD Biosciences, 562415). Unstained cells and fluorescence-minus-one (FMO) were used as controls for gating.

To isolate the mesenchymal population by flow cytometry, the following gating sequence was used: CD31-APC (Biolegend, 303116) identified the endothelial population. From the non-endothelial compartment ($CD31^{-}$), the mesenchymal populations were identified based on CD73-PE-Cy7 (BD Biosciences, 561258) expression (i.e. $CD31^{-}CD73^{+}$). The CD146-FITC (AbD Serotec, MCA2141FT) bright cells were gated as $CD146^{hi}CD73^{hi}$ mesenchyme, and CD146-FITC negative cells were gated as $CD146^{lo}CD73^{lo}$ mesenchyme.

Flow cytometric analysis of cultured CB $CD34^{+}$ cells

After 2 weeks of co-culture, cells were harvested and stained with the following antibodies: CD45-BV711 (Biolegend, 304050), CD34-APC Cy7 (Biolegend, 343514), CD14-BV605 (Biolegend, 301834), CD10-PE Cy7 (Biolegend, 312214), CD33-PE (BD Biosciences, 347787), CD19-BV510 (Biolegend, 302242), FITC anti-human Lineage Cocktail (CD3, CD14, CD16, CD19, CD20, CD56) (Biolegend, 348801), CD90-APC (Biolegend, 328114), CD45RA-PerCP-Cy5.5 (Biolegend, 304122).

Colony forming unit assay

After 2 weeks of co-culture on mesenchymal monolayers, cells were harvested and plated in methylcellulose (Methocult 4434; Stem Cell Technologies) at 2.5×10^3 cells/well. Colonies were scored after 14 days and reported as the sum of the progeny of colony forming unit (CFU) granulocyte-macrophage (CFU-GM), burst-forming unit erythroid (BFU-E), and CFU mixed/CFU GEMM .

In vivo repopulation assay

CB CD34^{+} cells were co-cultured with CD146^{hi}CD73^{hi} or CD146^{lo}CD73^{lo} monolayers for 2 weeks. After flow cytometry analysis, an equal number of CD45^{+} cells (1 x 10^{5}) was obtained from the co-cultures and was intratibially injected into each sublethally irradiated (250 cGy), 6- to 8-week-old NSG mouse (The Jackson Laboratory). Control animals received 1 x 10^{5} non-cultured CD45+ cells from the same CB. Mice were sacrificed 6 weeks post-transplantation. Engraftment of human hematopoietic cells was evaluated by FACS analysis after staining with anti-human specific monoclonal antibodies: CD45-APC Cy7, HLA (A/B/C)-PE, CD34-PE Cy7, CD14-APC, CD19-FITC (all from BD Biosciences, 557833, 555553,348791, 340436, 555412). Engraftment was defined as presence of huCD45+HLA+ cells in at least 1 % of the bone marrow. For secondary transplantation, bone marrow from left and right tibias and femurs from each engrafted mouse was pooled and intra-tibially injected into a secondary host (one to one transplant). Engraftment was evaluated 6 weeks after secondary transplantation (12 weeks total). All animal studies were performed under a protocol approved by the UCLA animal care and use committee.

Transwell Experiments
Co-cultures were performed in HTS Transwell 96 Permeable Support Culture Plate System, 0.4µm Polycarbonate Membrane (Corning). Contact= mesenchyme and CB CD34+ both plated on the bottom well. No Contact= mesenchyme plated on the bottom well, CB CD34+ cells plated on the top well.

Notch and Wnt experiments
For the inhibition of JAG1, 1 µg/ml of anti-Jag1 N-17 (sc-34473, Santa Cruz Biotechnology) was added to each well every 48 hours from d0 to time of analysis at d13. An equal concentration of irrelevant Goat IgG (Sigma I9140) was added to wells as negative controls for anti-Jagged1 antibody.

For analysis of effects of sFRP-1 and Wnt signaling, 2 µg/ml recombinant human sFRP-1 protein (R&D Systems), 10uM WAY 316606 hydrochloride (sFRP-1 inhibitor) (R&D Systems) or 3uM CHIR99021 ("CHIR"=CHIR99021, Wnt agonist, GSK3 inhibitor) (Tocris Bioscience) were added to each well every 48 hours from d0 to time of analysis at d14. An equal volume of dimethylsulfoxide (DMSO; Sigma) was added to wells as negative control for WAY 316606 and CHIR99021 treatment.

Cell cycle analysis
At d14 of mesenchymal differentiation from hEMP, 1 mM BrdU (diluted in EGM-2 supplemented with SB-431542) was added to the culture for 40 minutes before harvesting cells for flow cytometry analysis. BrdU incorporation was measured using the FITC BrdU Flow Kit (BD Biosciences). The cell-cycle phases were defined as apoptotic (BrdU^{-} and DNA<2n), G0/G1 (BrdU^{-} and 2n DNA), S (BrdU+ and 2n≤DNA≤ 4n) and G2/M (BrdU^{-} and 4n DNA).

Quantitative RT-PCR
Freshly sorted hEMP-derived mesenchyme subsets (CD146^{hi}CD73^{hi} and CD146^{lo}CD73^{lo}) were processed for RNA extraction using a Qiagen micro kit. An Omniscript reverse transcriptase (RT) kit was used to make complementary DNA, which was subjected to quantitative polymerase chain reaction (qPCR) using Taqman probe-based gene expression assay (Applied Biosystems) and β2 microglobulin (β2M) as housekeeping gene. Best coverage primer/probe sets were selected for all target genes (*MCAM, IGFBP2, KITLG, LEPR, NES* and *CSPG4*). A 7500 real-time PCR system was used (ABI). Data were analyzed using the comparative ΔΔCT method.

For the Notch and Wnt target gene expression assays, co-culture of CB CD34+ cells with CD146^{hi}CD73^{hi} or CD146^{lo}CD73^{lo} was performed as described. 7 days after co-culture, cells were harvested and isolated as CD34+ cells and Lin+ cells by flow cytometry. RNA extraction and qPCR was performed as above to examine the expression of target genes *HES1*, *TCF7* and *LEF1*.

Taqman primers
For Quantitative RT-PCR, best coverage primer/probe sets were selected for all target genes as following:

	Assay ID
JAG1	Hs01070032_m1
HES1	Hs00172878_m1
TCF7	Hs01556515_m1

LEF1	Hs01547250_m1
MCAM	Hs00174838_m1
IGFBP2	Hs01040719_m1
KITLG	Hs00241497_m1
LEPR	Hs00174497_m1
NES	Hs04187831_g1
CSPG4	Hs00361541_g1
SFRP1	Hs00610060_m1
B2M	Hs00187842_m1

Isolation of human primary pericytes from lipoaspirate and RNA-Seq
Human pericytes were derived from human lipoaspirate specimens obtained as discarded anonymous waste tissue and thus deemed exempt from institutional review board approval. One hundred milliliters of lipoaspirate were incubated at 37°C for 30 minutes with digestion solution composed by RPMI-1640 (Cellgro), 3.5% bovine serum albumin (Sigma), and 1 mg/mL collagenase type II (Sigma). Adipocytes were discarded after centrifugation while the pellet was resuspended and incubated in red blood cell lysis (eBioscience) to obtain the stromal vascular fraction (SVF). SVF was processed for fluorescence-activated cell sorting (FACS).

Cells were incubated with the following antibodies: CD45-APC Cy7 (BD Biosciences), CD34-APC (BD Biosciences), and CD146-FITC (AbD Serotec). The viability dye 4,6 diamidino-2-phenylindole (DAPI; Sigma) was added before sorting on a FACS Aria III (BD Biosciences). $DAPI^{-}CD45^{-}CD34^{-}CD146^{+}$ were gated as pericytes and DAPI-CD31-CD45-CD34+CD146- were gated as adventitial cells.

For RNA-Seq, lipoaspirate-derived adventitial cells (CD31-CD45-CD34+CD146-) and pericytes (CD31-CD45-CD34-CD146+) were sorted from two donors. A fraction was pelleted for RNA extraction and the remaining cultured for 3 passages in DMEM containing 20% FBS. Following three passages in culture, the adventitial cells were pelleted for RNA extraction. Because of phenotype drifting, the pericytes were resorted for a consistent phenotype (CD31-CD45-CD34-CD146+) prior to collection for RNA extraction.

50ng of total RNA was extracted from a given population obtained from each of two donors and combined to yield a pooled sample of 100ng of total RNA. The samples were sequenced on an Illumina HiSeq 2000 using paired-end 100bp reads to a depth of coverage of ~ 10-15 million reads.

Isolation of human $CD34^{+}$ cells from CB
Umbilical cord blood (CB) was collected from normal deliveries without individually identifiable information, therefore no institutional review board approval was required. MNCs were isolated by density gradient centrifugation using Ficoll-Paque (GE Healthcare). Enrichment of $CD34^{+}$ cells was then performed using the magnetic-activated cell sorting system (Miltenyi Biotec) as per the manufacturer's instructions. Only samples with >80% purity were used in experiments.

RNA-Seq and analysis
hEMP ($CD326^{-}CD56^{+}$ cells) and freshly sorted day 14 hEMP-derived mesenchyme subsets ($CD146^{hi}CD73^{hi}$ and $CD146^{lo}CD73^{lo}$) were extracted with Trizol and purified using miRNeasy Mini Kit (Qiagen). Three biological replicates were obtained for each population. 500ng-2ug of total RNA was input to generate cDNA using Nugen Ovation RNA-Seq System v2 and the sequencing libraries were generated using prepX DNA library enzyme kit (IntegenX Inc.) per manufacturer's instructions. Paired-end 100bp sequencing was performed on Illumina HiSeq 2000 with six samples multiplexed per lane. Raw sequence files were obtained using Illumina's proprietary software and are available at NCBI's Gene Expression Omnibus (accession number GSE77879 and GSE83443).

RNA-Seq reads were aligned using STAR v2.3.0 (Dobin et al., 2013). The GRCh37 assembly (hg19) of the human genome and the corresponding junction database from Ensemble's gene annotation were used as reference for STAR. The count matrix for genes in the Ensembl genome annotation was generated with HTSeq-count v0.6.1p2 (Anders et al., 2015). DESeq v1.14.0 (Anders and Huber, 2010) was used for normalization (using the geometric mean across

samples), differential expression (to classify genes as differentially expressed, Benjamini-Hochberg adjusted p-value < 0.01) and to compute moderate expression estimates by means of variance-stabilized data. Heatmaps were built with GENE-E (http://www.broadinstitute.org/cancer/software/GENE-E/) applying relative min/max normalization to moderate expression estimates. Gene expression estimates of hPSC (H1 line), hEMP, hPSC-derived mesenchyme and lipoaspirate-derived mesenchyme were available in Table S1.

Gene Set Enrichment Analysis and STRING protein interaction analysis

Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) was used to perform the comparative analysis between the RNA-Seq expression signatures of hEMP-derived and lipoaspirate pericytes. The differentially expressed genes in hPSC-derived $CD146^{hi}CD73^{hi}$ vs $CD146^{lo}CD73^{lo}$ mesenchyme (fold change>2, FDR<0.01) defined the gene set of interest (*positive signature*: upregulated genes (n=540); *negative signature*: downregulated genes (n=351)). In both cases, the background for enrichment consisted of a ranked gene list where top ranking genes were more expressed in lipoaspirate-derived CD146+ pericytes as compared to CD146- cells. GSEA was run independently using both freshly sorted and cultured cells as background. For the positive (negative) signature, significant enrichment was only found for genes up- (down-) regulated in both freshly sorted and cultured lipoaspirate-derived CD146+ pericytes. The core gene sets provided by GSEA (highlighted in grey in Fig 4A, B) were further filtered for those genes that are consistently up or downregulated in $CD146^{++}$ from all sources. This final $CD146^{++}$ gene expression signature was then imported into the STRING database (Jensen et al., 2009). For protein interaction analyses, STRING (version 10) was used with default parameters. GO enrichment analysis was performed to identify the biological processes involved in the predicted STRING protein network.

SUPPLEMENTAL REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Evseenko, D., Zhu, Y., Schenke-Layland, K., Kuo, J., Latour, B., Ge, S., Scholes, J., Dravid, G., Li, X., MacLellan, W.*, et al.* (2010). Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *107*, 13742-13747.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M.*, et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res *37*, D412-416.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S.*, et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.