Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations

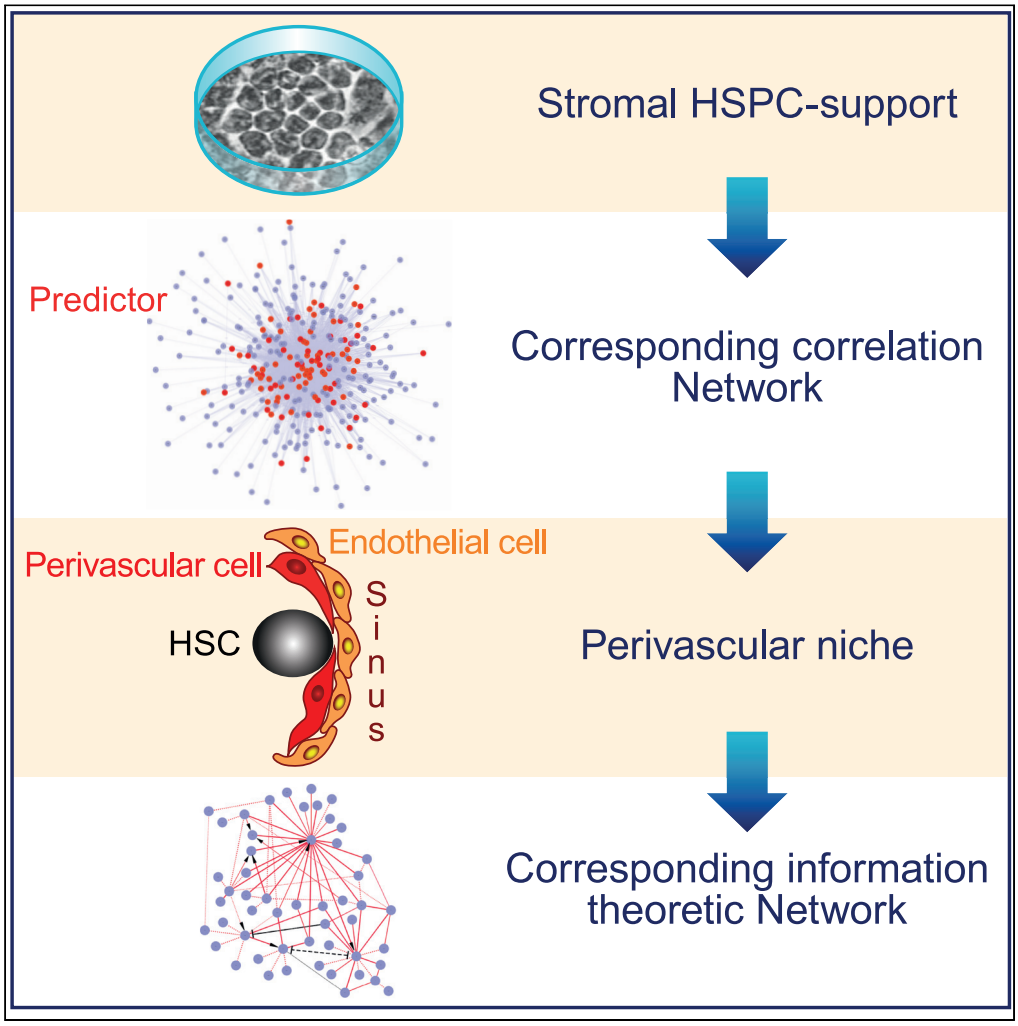
Christophe Desterke, Laurence Petit, Nadir Sella, Nathalie Chevallier, Vincent Cabeli, Laura Coquelin, Charles Durand, Robert A.J. Oostendorp, Hervé Isambert, Thierry Jaffredo, et al.

iScience

CellPress
OPEN ACCESS

Article

Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations



Christophe Desterke, Laurence Petit, Nadir Sella, ..., Hervé Isambert, Thierry Jaffredo, Pierre Charbord

pierre.charbord@inserm.fr

HIGHLIGHTS

A correlation network with predictor genes for the BM HSPC-supportive stromal niche

An information theoretic network for the supportive perivascular stromal niche

Wnt facilitator Rspo2 together with SCF to amplify *ex vivo* hematopoietic precursors

Resource combining bioinformatics algorithms to search for novel stromal mediators

Desterke et al., iScience *23*, 101222
June 26, 2020
https://doi.org/10.1016/j.isci.2020.101222

iScience

CellPress
OPEN ACCESS

Article

Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations

Christophe Desterke,[1] Laurence Petit,[2] Nadir Sella,[3] Nathalie Chevallier,[4] Vincent Cabeli,[3] Laura Coquelin,[4] Charles Durand,[2] Robert A.J. Oostendorp,[5] Hervé Isambert,[3] Thierry Jaffredo,[2] and Pierre Charbord[2,6,*]

SUMMARY

The cardinal property of bone marrow (BM) stromal cells is their capacity to contribute to hematopoietic stem cell (HSC) niches by providing mediators assisting HSC functions. In this study we first contrasted transcriptomes of stromal cells at different developmental stages and then included large number of HSC-supportive and non-supportive samples. Application of a combination of algorithms, comprising one identifying reliable paths and potential causative relationships in complex systems, revealed gene networks characteristic of the BM stromal HSC-supportive capacity and of defined niche populations of perivascular cells, osteoblasts, and mesenchymal stromal cells. Inclusion of single-cell transcriptomes enabled establishing for the perivascular cell subset a partially oriented graph of direct gene-to-gene interactions. As proof of concept we showed that R-spondin-2, expressed by the perivascular subset, synergized with Kit ligand to amplify *ex vivo* hematopoietic precursors. This study by identifying classifiers and hubs constitutes a resource to unravel candidate BM stromal mediators.

INTRODUCTION

It is now well established that stem cells reside in niches where they are associated with other cell types that control stem cell behavior. This is particularly true for hematopoietic stem cells (HSCs) whose niche include several cell types that express a large array of mediators that facilitate or impede the different stem cell attributes (self-renewal, lineage commitment, proliferation, and survival). During the last decade, *in vivo* studies, by means of transgenic mice and lineage tracing or transplantation of cells defined by phenotype, have started to unravel the complexity of the bone marrow (BM) HSC niches (Kfoury and Scadden, 2015; Mendelson and Frenette, 2014; Morrison and Scadden, 2014; Pinho and Frenette, 2019). Several types of microenvironmental cells are involved in HSC regulation, including neural, hematopoietic, and stromal cells. Stromal cell types essential for HSC maintenance include vascular endothelial cells and perivascular cells, peri-sinusoidal on the abluminal side of the endothelial lining of BM sinusoids, and peri-arteriolar in the media and adventitia of small BM arterioles (Acar et al., 2015; Asada et al., 2017; Chen et al., 2016; Ding and Morrison, 2013; Ding et al., 2012; Greenbaum et al., 2013; Kunisaki et al., 2013; Oguro et al., 2013; Sugiyama et al., 2006). Bone-forming osteoblasts also contribute to the HSC niche, in particular in irradiated animals transplanted with HSCs when perivascular and endothelial cells are impaired (Silberstein et al., 2016). Moreover, both perivascular and osteoblastic cell subsets are implicated in the regulation of distinctive, B-lymphoid and multipotent progenitors (Balzano et al., 2019; Cordeiro Gomes et al., 2016; Ding and Morrison, 2013; Greenbaum et al., 2013; Silberstein et al., 2016). Mesenchymal stromal cells (MSCs) may also be considered as niche components by virtue of their phenotype, perivascular location, pro-hematopoietic factor expression, and precursors of the osteoblastic lineage (Hu et al., 2016; Mendez-Ferrer et al., 2010; Morikawa et al., 2009; Pinho et al., 2013; Zhao et al., 2019; Zhou et al., 2014). MSCs with hematopoietic stem/progenitor cell (HSPC)-supportive ability have also been isolated from human fetal and adult BM by culture or by phenotype using different antigen membrane combinations, some of which are similar to those used for murine MSCs (Chan et al., 2018; Isern et al., 2013; Pinho et al., 2013; Sacchetti et al., 2007).

Recent transcriptomic and proteomic analyses at the single-cell level have helped define the gene sets characterizing the BM stromal populations under steady state, after stress, or in diseased conditions (Balzano et al., 2019; Baryawno et al., 2019; Tikhonova et al., 2019; Wolock et al., 2019). These works first

[1]Université Paris Sud, Hôpital Paul Brousse, Villejuif, France

[2]Sorbonne Université, UPMC Université Paris 06, IBPS, CNRS UMR7622, Inserm U 1156, Laboratoire de Biologie du Développement; Paris 75005, France

[3]Institut Curie, PSL Research University, CNRS UMR168, Paris, France

[4]IMRB U955-E10, INSERM, Unité d'Ingenierie et de Thérapie Cellulaire- EFS, Université Paris-EST, Créteil, France

[5]Clinic and Polyclinic for Internal Medicine III, Klinikum Rechts der Isar, Technical University Munich, Munich, Germany

[6]Lead Contact

*Correspondence: pierre.charbord@inserm.fr

https://doi.org/10.1016/j.isci.2020.101222

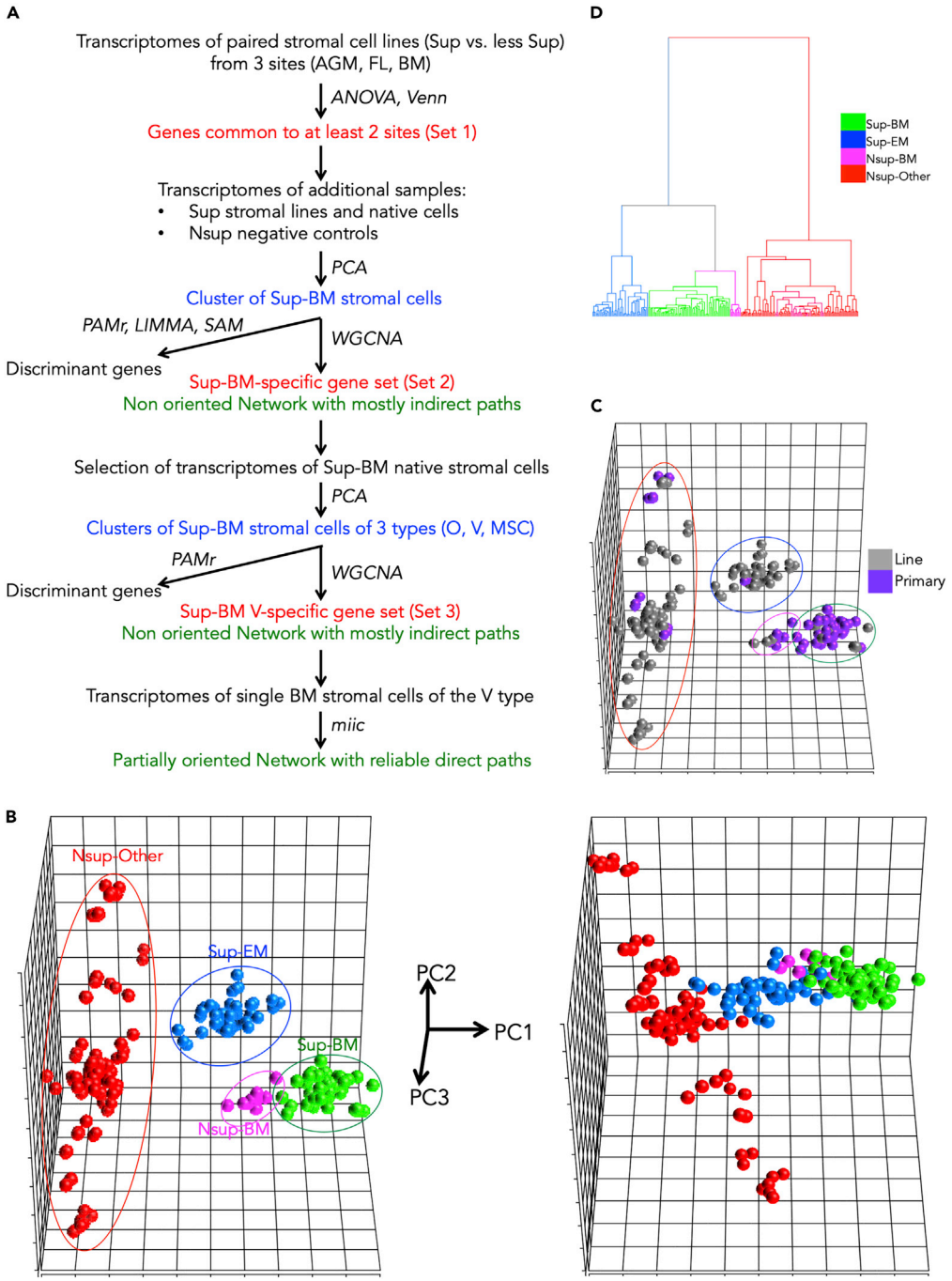


Figure 1. Study Flowchart and Cell Cluster Identification

(A) *Study flowchart.* Sup, supportive; Nsup, non-supportive; BM, bone marrow; O, osteoblast; V, perivascular cell; MSC, mesenchymal stromal cell; PCA, principal-component analysis; PAMr, Prediction Analysis for Microarrays; LIMMA, linear model for microarray analysis; SAM, Significance Analysis of Microarrays; WGCNA: Weighted Gene Correlation Network Analysis; miic, multivariate information-based inductive causation algorithm.

(B) *PCA highlighting the different clusters.* Left panel: The 3D score plot (PC1, PC1, PC3) using the entire sample set (224 samples) as observations and the Set1 (1,859 genes) as variables. Right panel: The 3D score plot using the entire sample set as observations and the entire set of protein-coding genes (15,056 genes) as variables. Sup-BM, supportive BM samples; Sup-EM, supportive extra-medullary samples; Nsup-BM, non-supportive BM samples; Nsup-Other, non-supportive samples of other origin than BM.

(C) PCA highlighting the cell sources (cell lines versus primary cells).

Figure 1. *Continued*
(D) *Hierarchical clustering highlighting the different clusters.* Euclidian distances and Ward's linkage were used.
see also Figure S1.

explored the cell heterogeneity and then inferred the possible capacity of support of the cell sets by unraveling their hematopoietic regulator profiles. On the contrary, in this work we aim at developing a strategy to first define the genes that orchestrate a functional phenotype and then investigate the cell populations contributing to that phenotype. In a previous work we have attempted to provide such information, defining several networks characterizing HSC-supportive stromal cells (Charbord et al., 2014). However, our previous study suffered from the following limitations: (1) it included too few observations for statistically sound correlation studies; (2) it did not indicate which genes in the characteristic gene sets were the best classifier for the corresponding cell population; (3) the described networks characterized a core gene set for stromal cells from different developmental origin, and consequently could not apply to defined BM stromal populations; and (4) the described networks were based on the analysis of gene-to-gene correlations; consequently most of the inferred edges included a number of indirect paths resulting from transitive correlations, which did not indicate causal relationships.

We undertook the present work to remedy those limitations. We enlarged the number of genes of interest by filtering out a minimal number of genes. We also considerably enlarged the number of observations by including transcriptomes selected from available databases according to well-defined criteria. The distribution of the samples in the 3-dimensional space defined by principal-component analysis (PCA) then enabled to identify a population of BM HSPC-supportive stromal cells, subsequently dissected in clusters corresponding to stromal cell subsets. We applied a discriminant analysis pipeline to identify the genes predominantly expressed by the different cell sets and analyzed the different cell populations to unravel the gene relatedness (Horvath, 2011; Langfelder and Horvath, 2008). Moreover, we took advantage of the numerous observations provided by recently published single-cell transcriptomes (Tikhonova et al., 2019) to disclose in the perivascular network the direct gene-to-gene interactions and reveal oriented links based on the signature of causality (Sella et al., 2018; Verny et al., 2017). Finally, as proof of concept, we studied the protein expression and activity of the top most discriminant and connected gene, the Wnt signaling pathway facilitator R-spondin 2 (*Rspo2*), and validated that it was expressed by a small set of perivascular cells, and acted in concert with the Stem Cell Factor (SCF) (also known as Kit Ligand) to amplify *ex vivo* hematopoietic precursors.

RESULTS

The major aim of our work was to unravel how genes expressed by the different stromal niche populations were related to one another, i.e., how genes were organized around central hubs serving as keystones of the network. To achieve this goal we devised a strategy of analysis consisting of a sequence of steps as indicated in the flowchart shown in Figure 1A.

Configuring a Dataset Representative of Stromal Cells of Mesenchymal Origin

Our starting hypothesis was the same as that published in 2014 (Charbord et al., 2014), considering that essential genes for HSC support would be those differentially expressed by at least two of the major developmental sites of definitive hematopoiesis. Thus, we used the same stromal cell lines as before, from the aorta-gonad-mesonephros (AGM), fetal liver (FL), and BM, either with supportive activity in *ex vivo* co-culture experiments with HSPCs or with reportedly less or non-supportive ability. To initiate the present study, we filtered out a minimal number of genes as the subsequent analyses (Weighted Gene Correlation Analysis [WGCNA], discriminant) constituted by themselves dimensional reduction methods. Thus, for comparing supportive with less supportive lines we now took into account all differentially expressed genes given by ANOVA (2-way with support, site and interaction as grouping variables) with a low-stringency p value < 0.06. The different genes sets were then intersected, which resulted in more than a 4-fold increase in the number of retained genes when compared with our previous study (2,327 and 481 genes in the present and previous work, respectively). The Venn diagram is shown in Figure S1A. The resulting gene set was designated as Set 1.

To account for the diversity of stromal cells, we selected in the publicly available databases datasets corresponding to articles wherein the reported supportive capacity had been evidenced by one or several of the following criteria (Table S1): (1) generation of hematopoietic colonies in co-culture of HSPCs with

stromal cells, (2) hematopoietic differentiation of embryonic stem cells co-cultured with the stromal cells, (3) *in vivo* co-localization of stromal cells expressing specific markers with HSCs, (4) lineage reconstitution *in vivo* by HSCs co-cultured with stromal cells for variable time spans, (5) reduction of the HSC pool *in vivo* after depletion of the stromal cell population, and (5) characteristic phenotype (e.g., expression of the major hematopoietic regulators *Kitl* and *Cxcl12*). The relative relevance of the six criteria in the ensemble of articles is shown in Figure S1B. The supportive datasets included mesenchymal cells from the AGM, spleen, FL, and BM, either derived as continuous cell lines or primary cells isolated from the different tissues by phenotype (and in one case by culture) from transgenic or wild-type mice. The non-supportive datasets included not only stromal cells unable to support HSPCs according to the aforementioned criteria but also pluripotent stem cells (induced or embryonic), either undifferentiated or differentiated into the neural-ectodermal or endodermal lineages, and adult cardiac cells. In all, by analyzing 30 datasets we extracted 224 samples. A normalization step was necessary to get rid of the batch effect (Figure S1C). When genes corresponding to Set 1 were mapped in all the datasets, a gene count of 1,869 genes remained. The final expression matrix was therefore 224 × 1,869 (observations × variables). Details on samples and the list of genes belonging to Set 1 are given in Tables S1 and S2, respectively.

PCA of the Dataset Allows Identifying Four Distinct Cell Clusters

Using the 224 samples and the 1,869 Set 1 genes PCA revealed four major clusters (Figure 1B, left panel). The p value of group discrimination evaluated by correlation on the first principal component (PC) axis was highly significant ($p = 2.45 \times 10^{-190}$). The BM stromal cells segregated in two homogeneous clusters (colored green and magenta) with 69 and 9 samples of supportive cells and non-supportive cells, respectively (Sup-BM and Nsup-BM). The blue cluster included 47 samples of supportive extra-medullary (Sup-EM) cells (26 AGM, 18 FL, and 3 spleen samples). Finally, the red cluster included 99 samples of non-supportive cells (Nsup-Other). The covariate cell source (primary cells or continuous cell lines) was not relevant in the context of this study because it did not generate sub-groups within the clusters (Figure 1C). As expected, the four clusters were not discriminated from one another when including all 15,056 genes, although a gradient from non-supportive to supportive samples was apparent (Figure 1B, right panel). Hierarchical clustering confirmed the four major clusters (Figure 1D).

One Gene Network Characterizes the BM Stromal Supportive Capacity

Analysis of the score plots in Figure 1B indicated that the PC1 representing 35% of the variance corresponded to the contrast between Sup-BM and Nsup-Other. To understand how Set 1 genes correlated to this trait were inter-related, we applied WGCNA (Horvath, 2011; Langfelder and Horvath, 2008), which enabled identifying one module (sub-network) named turquoise whose eigengene positively correlated to the trait (Figures 2A and S2). This module included 758 genes with significant ($p < 0.05$) gene significance and module membership. This gene set characteristic of the Sup-BM cell population was designated as Set 2 (Table S3). Gene Ontology analyses (Huang da et al., 2009a, b) indicated that Set 2 genes included molecules of the extracellular matrix (ECM) and implicated in the ECM-to-cell receptor interaction, and genes implicated in protein processing in the endoplasmic reticulum (glycosyl transferases and members of the glycosaminoglycan and N-glycan biosynthesis pathways) (Figure 2B). Importantly, genes belonging to the canonical Wnt signaling pathway (such as *Fzd1*, *Porcn*) were also enriched in the turquoise module. The most connected genes of the turquoise module are represented on the connectivity plots (Figure 2C). Of the two major hematopoietic regulators *Kitl* and *Cxcl12*, the former was included in the gene set, but not the latter due to its high expression in FL non-supportive lines and in some pluripotent stem cell lines. Among the other molecules some were soluble mediators (e.g., *Rspo2*, *Srpx2*, *Fam20a*, *Il7*), membrane receptors (e.g., *Lpl*), ECM components (e.g., *Col1a2*, *Efemp1*, *Tgfbi*), adhesion molecules (e.g., *Svep1*), and transcription factors (e.g., *Ebf3*, *Snai2*, *Prrx1*).

To further confirm that the turquoise module gene set clearly discriminated the Sup-BM samples from all others, we performed gene set enrichment analyses (GSEA) (Liberzon et al., 2015; Subramanian et al., 2005). Using the Sup-BM gene set as reference, the enrichment plot was clearly shifted to the left corresponding to up-regulated genes in the Sup-BM cell population (Figure 2D).

To uncover the best classifiers in the Sup-BM cluster we followed several steps through a customized pipeline. First, using the Prediction Analysis for Microarrays (PAM) algorithm (Tibshirani et al., 2002), supervised machine learning analysis was carried out on the four clusters observed in PCA. Leave-one-out with cross-validation indicated that the gene Set 1 enabled effective discrimination between each cluster because the

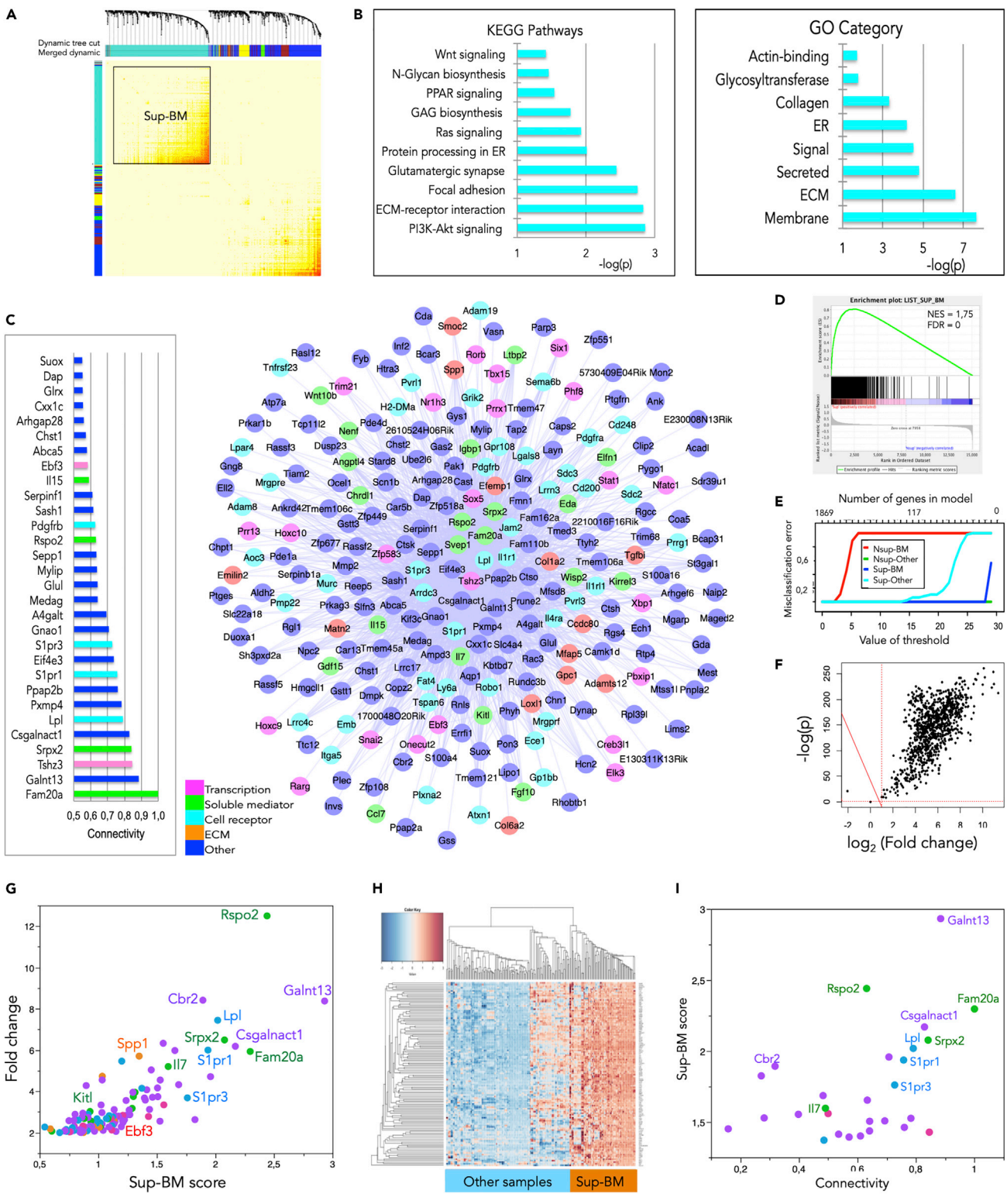


Figure 2. One Gene Network Characterizes the BM Stromal Supportive Capacity

(A) *Module detection by clustering using the Topological Overlap Matrix (TOM) and merged dynamic tree cut method.* The merged dynamic hybrid branch cutting method was used to define the module named by color. Parameters: power: 14, adjacency: signed; correlation: Pearson's; linkage: average; minimum module size: 30. The gene Set 2 corresponds to the set of 758 genes belonging to the turquoise module.

Figure 2. *Continued*

(B) *Major Gene Ontology (GO) categories corresponding to Set 2 genes.* ER, endoplasmic reticulum; ECM, extracellular matrix; GAG, glycosaminoglycan.
(C) *Connectivity plots of genes belonging to Set 2.* Left panel: Intra-modular connectivity bar plot of the top 30 genes. Right panel: Cytoscape connectivity plot of the 249 most connected genes; genes were selected for representation above a threshold weight $\geq$ 0.1228.
(D) *GSEA plot* using as reference the gene Set 2, as expression dataset the matrix of all genes and samples, and as phenotype labels Sup versus Nsup.
(E) *Misclassification error plot using the PAMr algorithm.*
(F) *Volcano plot obtained using the LIMMA algorithm with Rspo2 expression as quantitative linear predictor* (only genes linearly dependent on Rspo2 are represented).
(G) *Last step of the discriminant analysis* (*118 genes, "Discriminant 2" predictors*). Scatterplot of the Sup-BM PAMr score versus the fold change (FC) given by the SAM algorithm (FC of gene expression in Sup-BM versus all other populations).
(H) *Hierarchical clustering* (*Euclidian distance, average linkage*) *and heatmap of the 118 'Discriminant 2' genes.*
(I) *Gene intra-modular connectivity versus Sup-BM PAMr score scatterplot for the 28 most discriminant genes.*
See also Figures S2 and S3.

misclassification error was nil when a threshold value of 1,401 genes was reached (Figure 2E). The best 939 genes predictive of BM support were then selected and applied to a Random Forest algorithm (Breiman, 2001). After performing learning on training dataset (n = 168 samples randomly selected) the model was built on 100 trees with optimization of the number of features needed to perform the tree split (the "mtry" parameter) for a minimal error of tree aggregations (bagging). No misclassifications of samples were found using the validation set (n = 52 samples randomly selected) when testing the Sup-BM cluster versus other clusters pooled in one single class. ANOVA on the best 720 discriminating genes ("Discriminant1" gene set) indicated that *Rspo2* was the best predictive as it was up-regulated uniquely in the Sup-BM cluster. This gene was therefore used as quantitative predictor in linear models such as "linear model for microarray analysis" (LIMMA) (Figure 2F) (Ritchie et al., 2015). Significance Analysis for Microarray (SAM) (Tusher et al., 2001) was performed in parallel by supervised analysis between independent clusters (Sup-BM versus others). The intersection between SAM and LIMMA using *Rspo2* as predictor allowed obtaining a gene set of 118 genes called "Discriminant 2" (Figure 2G and Table S3). By unsupervised classification, this gene set achieved a perfect discrimination of Sup-BM when compared with all others (Figure 2H). Remarkably, the "Discriminant 2" gene set was included within the Set 2 (Table S3). The combination of the results given by the discriminant analysis on the one hand and WGCNA and GSEA on the other enabled identifying the most discriminative and most connected genes, foremost *Rspo2* (Figure 2I).

A Similar Gene Network Characterizes the BM Stromal Supportive Capacity after Inclusion of Human Samples

To investigate whether the most discriminative and connected genes found in mouse BM supportive stromal cells were also characteristic of human samples we included in the study five datasets (24 samples) reporting transcriptomes of supportive or non-supportive MSCs, either primary layers or cell "strains" pooling a few clones. Details are given in Table S1. The genes corresponding to Set 1 were selected in all databases, giving a final gene count of 1,619 (instead of 1,871 due to difference in species, some of the murine genes, such as *Ly6a*, having no corresponding human homolog). The final expression matrix was therefore 248 × 1,619 (observations × variables). In the PCA the human samples merged well with the murine ones, and the clusters were similar to those obtained with murine samples alone (Figure S3A). WGCNA allowed identifying the turquoise module whose eigengene was positively and most significantly positively correlated to PC1 ($p = 10^{-137}$). Most connected genes in the mouse turquoise network were among the most connected genes in the human + mouse blue module (e.g., *Kitl*, *Rspo2*, *Ebf3*) (Figure S3B). In conclusion, the inclusion of human BM samples had not introduced major modifications of the gene organization, suggesting its inter-species conservation.

The Supportive Capacity Is Implemented in Four BM Niche Cell Populations

As cultured stromal cell lines do not necessarily represent uncultured primary cells directly isolated from the *in vivo* BM niches, we selected the 51 samples corresponding to primary HSPC-supportive stromal cells isolated by phenotype (and in one case by culture) from transgenic or wild-type mice (Table S1). To investigate in which cell types the genes characteristic of supportive stromal cells were expressed and how they were organized, we analyzed the gene expression in these 51 primary cell samples, taking into account the 758 Set 2 genes up-regulated in the Sup-BM turquoise module described earlier. In the PCA the samples segregated into three clusters (Figure 3A, upper panel), which were confirmed by hierarchical clustering (Figure S4A). In the PCA the cluster on the right included 23 samples designated perivascular cells (V) as they were defined by the expression of the markers Nestin, Leptin receptor, Chondroitin sulfate proteoglycan, Glutamyl aminopeptidase, the cytokine

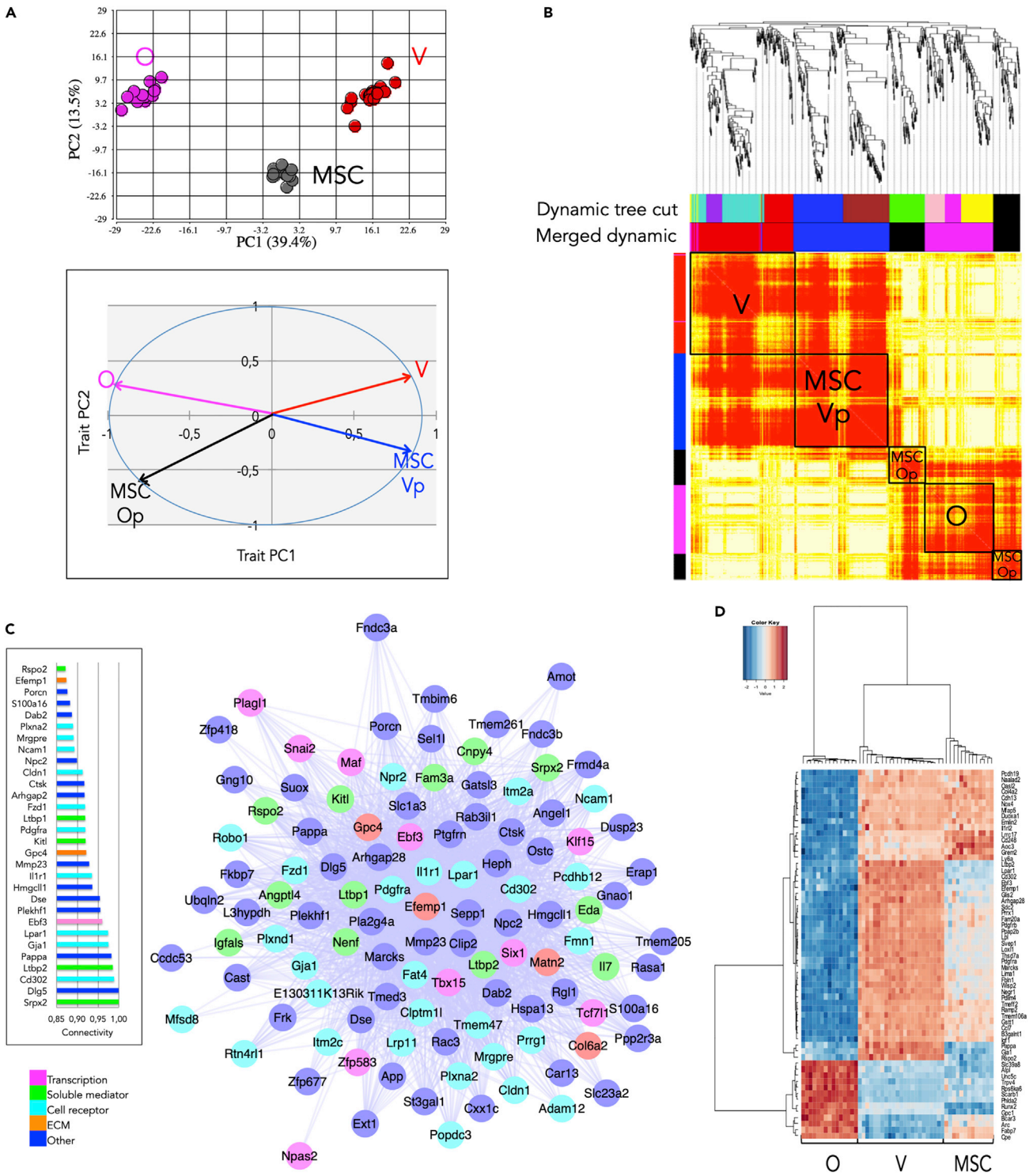


Figure 3. The Supportive Capacity Is Implemented in Four BM Niche Cell Populations

(A) *PCA and module eigenvectors corresponding to samples of osteoblasts (O), perivascular cells (V), and MSCs.* Upper panel: PC1 versus PC2 score plot using 51 samples and the 758 Set 2 genes. Lower panel: plot of the module eigenvectors versus PC1 and PC2 traits for the four modules identified by WGCNA.

(B) *Module detection by clustering using the Topological Overlap Matrix (TOM) and merged dynamic tree cut method.* See legend of Figure 2A. Parameters: power: 6; adjacency: signed; correlation: Pearson's; linkage: complete; minimum module size: 30.

Figure 3. *Continued*

(C) *Connectivity plots of genes belonging to the V module*. Right panel: Intra-modular connectivity bar plot of the top 30 genes. Left panel: Cytoscape connectivity plot of the 109 most connected genes; genes were selected for representation above a threshold weight $\geq$0.4623.

(D) *Discriminant analysis identifying the O-, V-, and MSC-specific genes*. Hierarchical clustering classification (Euclidean distance, Ward linkage) performed on the 61 best predictors.

See also Figure S4.

SCF, or the chemokine Stromal cell-derived factor 1 (Asada et al., 2017; Balzano et al., 2019; Chan et al., 2013; Decker et al., 2017; Ding et al., 2012; Greenbaum et al., 2013; Kunisaki et al., 2013; Mendez-Ferrer et al., 2010). The cluster on the left included 15 samples designated osteoblastic cells (O) as they were defined as Sca-1^{neg}, but CD166$^{pos/ne}$, Pdgfra or CD51^{pos} cells, or derived from Col2.3 mice (Eash et al., 2010; Nakamura et al., 2010; Schepers et al., 2013). The cluster in the middle included 11 samples designated MSCs as they were defined as Sca-1^{pos}, but Pdgfra^{pos} or CD166^{neg} cells, or as adherent cells from a clone cultured at high density (Greenbaum et al., 2013; Mead et al., 2017; Nakamura et al., 2010; Shoshani et al., 2014); it also included two samples of cells expressing Myosin-11 (Asada et al., 2017), which was probably related to their expression of the Sca-1 membrane antigen (Kunisaki et al., 2013) that also specifies PαS MSCs (Morikawa et al., 2009). PC1 (39.4% of the variance) corresponded to the contrast between the V and O clusters (PC1 trait), whereas PC2 (13.5% of the variance) corresponded to the contrast between the MSC cluster (including the two Myosin-11-positive cells) and the two others (PC2 trait).

In accord with the merged dynamic tree cut, WGCNA identified four modules named red, blue, magenta, and black (Figures 3B, S4B–S4D and Tables S4 and S5). The module eigengene correlations (Figure 3A, lower panel) to PC1 and PC2 traits and the analysis of the genes expressed in each module enabled characterizing the red and magenta modules as specific to V and O cells, respectively. Similarly, the black and blue modules were specified as MSC osteolineage-primed (MSC-Op) and perivascular-primed (MSC-Vp), respectively. The V module (Figure 3C) included the major hematopoietic regulator *Kitl* and the transcription factor *Ebf3* reported to maintain the supportive capacity and the architecture of the HSC niche (Seike et al., 2018). Its core of 109 highly correlated genes (out of 167) was designated as Set 3 (Table S4). The O module (Figure S4F) included the skeletal mineralization-inducing enzyme *Alpl* and the osteogenic transcription factor *Runx2*. The MSC-Op module (not shown) included the ECM components *Col1a1*, *Col1a2,* and *Sparc*, all implicated in ossification. Finally the MSC-Vp module (not shown) included the transcription factor *Prrx1* essential for the development of vascular and perivascular matrix (Bergwerff et al., 2000), and the ECM components *Emilin2* and *Thsd7a* implicated in vessel assembly and endothelial cell migration. KEGG pathway analysis confirmed the characterization of the V and MSC-Vp modules, canonical Wnt signaling and focal adhesion pathways being preeminent for genes in V and MSC-Vp modules, respectively (Figure S4E).

The proportion of the sum of genes in the niche cell-specific networks (V, MSC-Vp, O, MSC-Op) to the total number of genes in the Sup-BM network was 78.5%. However, the hubs in niche cell-specific networks were not necessarily hubs in the Sup-BM network. This result showed that the gene components found in the niche cell-specific networks were rewired in a different circuitry in the Sup-BM network.

To investigate which genes were predominantly expressed in each population we used the PAMr algorithm identifying 61 discriminant genes (designated "Discriminant 3"), including *Ebf3* and *Cd302* (also known as *Clec13a* or *Dcl1*) for V and *Alpl* and *Runx2* for O (Figure 3D and Tables S4 and S5). The combination of results given by discriminant analysis and WGCNA enabled identifying the most discriminative and most connected genes in V and MSC-Vp modules (Figure S4H).

Single-Cell-Level Analysis Identifies High-Confidence Paths in the Perivascular Cell Network

To refine the gene network characterizing the perivascular cell population, we analyzed a recent database (GSE108891) providing the transcriptome of the BM microenvironment at the single-cell level (Tikhonova et al., 2019). This publication compared three cell populations of vascular endothelial cells expressing VE-cadherin, perivascular cells expressing Leptin receptor, and osteoblasts expressing Col2.3. Different sub-populations were identified, some of which expressed the major pro-hematopoietic mediators implicated in cell-to-cell communication.

First, we confirmed that Set 3 including the 109 most connected genes of the V module specified the perivascular population. We randomly selected 100 cells from each of the three populations and applied PCA using either the whole transcriptome or the gene Set 3. The three cell populations were not segregated when including the entire transcriptome, whereas they were clearly resolved when using the Set3 (Figure 4A). The genes with the

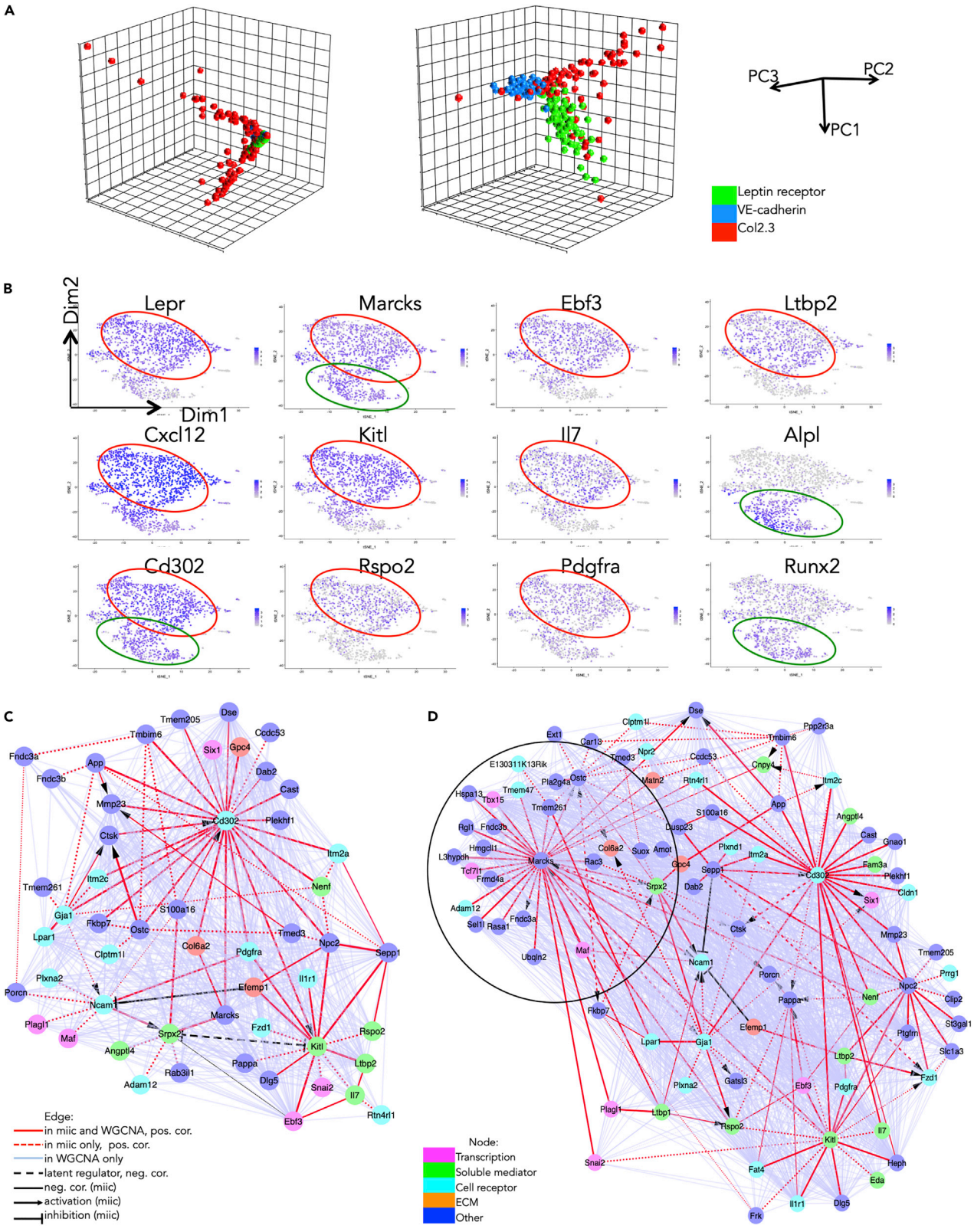

A
PC3
PC2
PC1
Leptin receptor
VE-cadherin
Col2.3
B
Dim2
Dim1
Lepr
Marcks
Ebf3
Ltbp2
Cxcl12
Kitl
Il7
Alpl
Cd302
Rspo2
Pdgfra
Runx2
C
Edge:
in miic and WGCNA, pos. cor.
in miic only, pos. cor.
in WGCNA only
latent regulator, neg. cor.
neg. cor. (miic)
activation (miic)
inhibition (miic)
D
Node:
Transcription
Soluble mediator
Cell receptor
ECM
Other

Figure 4. Single-Cell-Level Analysis Using a Recently Described Algorithm Identifies High-Confidence Paths in the Perivascular Cell Network

(A) PCA using either the whole transcriptome (left) or the Set 3 gene set (right) and 100 randomly selected genes from each of the three populations (VE-cadherin-positive, Leptin receptor-positive, and Col2.3-positive).

(B) *t-SNE of example genes expressed in the Leptin receptor-positive population at baseline*. All 1,712 cells were included. Negative cells are in gray, and scaled positivity for a given gene is in different hues of blue.

(C) *Miic plot of the Set3 genes expressed in the Leptin receptor-positive cell population at baseline*. The matrix consisting of 1,712 observations (cells of the Leptin receptor-positive population at baseline) and 109 variables (the Set 3 genes) was submitted to the miic algorithm. 52 genes (nodes) were retained. Edge width increases with degree of confidence in the link as given by miic. The edge line is continuous if the link was found both by miic and WGCNA, dotted if the link was found only by miic, and dashed if there is a common latent regulator found by miic. The edge color in red corresponds to a positive correlation by miic; in black, to a negative correlation by miic; and in light blue, to a positive correlation by WGCNA. The edge head pointed corresponds to an activation, and the edge head blunted corresponds to an inhibition as indicated by miic.

(D) *Miic plot of the Set3 genes expressed in the Leptin receptor-positive cell population after stress (treatment with 5-fluorouracil [5-FU])*. The matrix consisted of 3,467 observations (cells of the Leptin receptor-positive population after 5-FU treatment) and 109 variables. The new hub and linked targets appearing in condition of stress is encircled.

highest negative PC2 loadings were *Kitl*, *Cd302*, and *Ebf3*. Second, we applied t-distributed stochastic neighbor embedding (t-SNE) to the entire population of perivascular cells (1,712 cells) (Figure 4B). Some genes were expressed at high level in almost all cells (e.g., *Marcks*), whereas other genes were expressed in only a fraction of the cells. It was roughly possible to distinguish two regions, a large cluster A (circled red) including cells expressing genes of the V module (e.g., *Rspo2* and high amounts of *Kitl*) and a smaller cluster B (circled green) including cells expressing genes of the O module (e.g., *Alpl* and *Runx2*).

As the Set 3 genes specified the perivascular population corresponding to the major cell cluster A we took advantage of the large number of Leptin receptor-positive (LEPR) single cells to refine the V gene network. To identify direct paths between genes including causal relationships and inferring latent common regulators of expressed genes we applied the multivariate information-based inductive causation algorithm (miic) (Sella et al., 2018; Verny et al., 2017), submitting to it the matrix consisting of 1,712 observations and 109 variables (the Set 3 genes). As expected only a fraction of nodes (52) defining direct high-confidence paths were retained (Figure 4C). There were two major hubs, *Kitl* and *Cd302*. Remarkably, *Kitl* was linked to *Rspo2*, *Il7*, *Ebf3*, and *Snai2*. Moreover, the miic algorithm uncovered a number of "v-structures" (e.g., *Gja1*– > *Ctsk* < – *Ostc*) suggesting cause-and-effect interactions and the presence of latent regulators (e.g., a latent negative regulator upstream of *Srpx2* and *Kitl*).

Tikhonova et al. studied the transcriptome of LEPR cells not only under steady-state condition (baseline) but also after stress, i.e., after treatment of the mice with the drug 5-fluorouracil. We introduced therefore in our study this dataset including 3,467 cells. As shown by miic analysis the network was largely modified with a large increase in the number of nodes (89) and appearance of new hubs such as *Marcks* (Figure 4D). However, the hubs observed in baseline condition, in particular *Cd302* and *Kitl*, were still present and *Kitl* retained its direct links with *Rspo2*, *Il7*, *Ebf3*, and *Snai2*. There was also a large increase of new (not found in cells at baseline) "v-structures" (e.g., *Marcks*– > *Ncam* < – *Srpx2* or *Ltbp1*– > *Ncam* < – *Srpx2*). In short, by means of miic we were able to prune and remodel the gene network generated by WGCNA, which confirmed *Kitl* as major hub for the perivascular cell network, as well as identify possible causal interactions and show major network remodeling in cells responding to stress.

The Wnt Signaling Pathway Facilitator R-Spondin 2 Acts in Concert with the Kit Ligand to Amplify *Ex Vivo* Hematopoietic Precursors

In our analysis, *Rspo2* appeared as the best predictor of the Sup-BM capacity and was directly linked related to *Kitl* coding for the SCF widely recognized as a major stromal regulator of HSC attributes. R-spondin 2 (encoded by *Rspo2*) is known as a Wnt signaling pathway facilitator (Niehrs, 2012). Thus, to validate the aforementioned integrative analyses to define candidate stromal regulators of hematopoiesis, we studied the cell distribution of R-spondin 2 on BM biopsies and its effect on HSPC *in vitro*.

Histology of murine BM showed that R-spondin 2 was expressed in a subset of the LEPR peri-sinusoidal cell population, which confirmed the results of the single-cell transcriptome analysis described earlier (Figure 5A). It was also expressed in some LEPR-negative hematopoietic cells.

Over the whole range of our 224 samples, the expression level of *Rspo2* was positively correlated, in a highly significant way, to that of *Kitl*, being the seventh ranked gene among the top correlated ones (Figure 5B). To

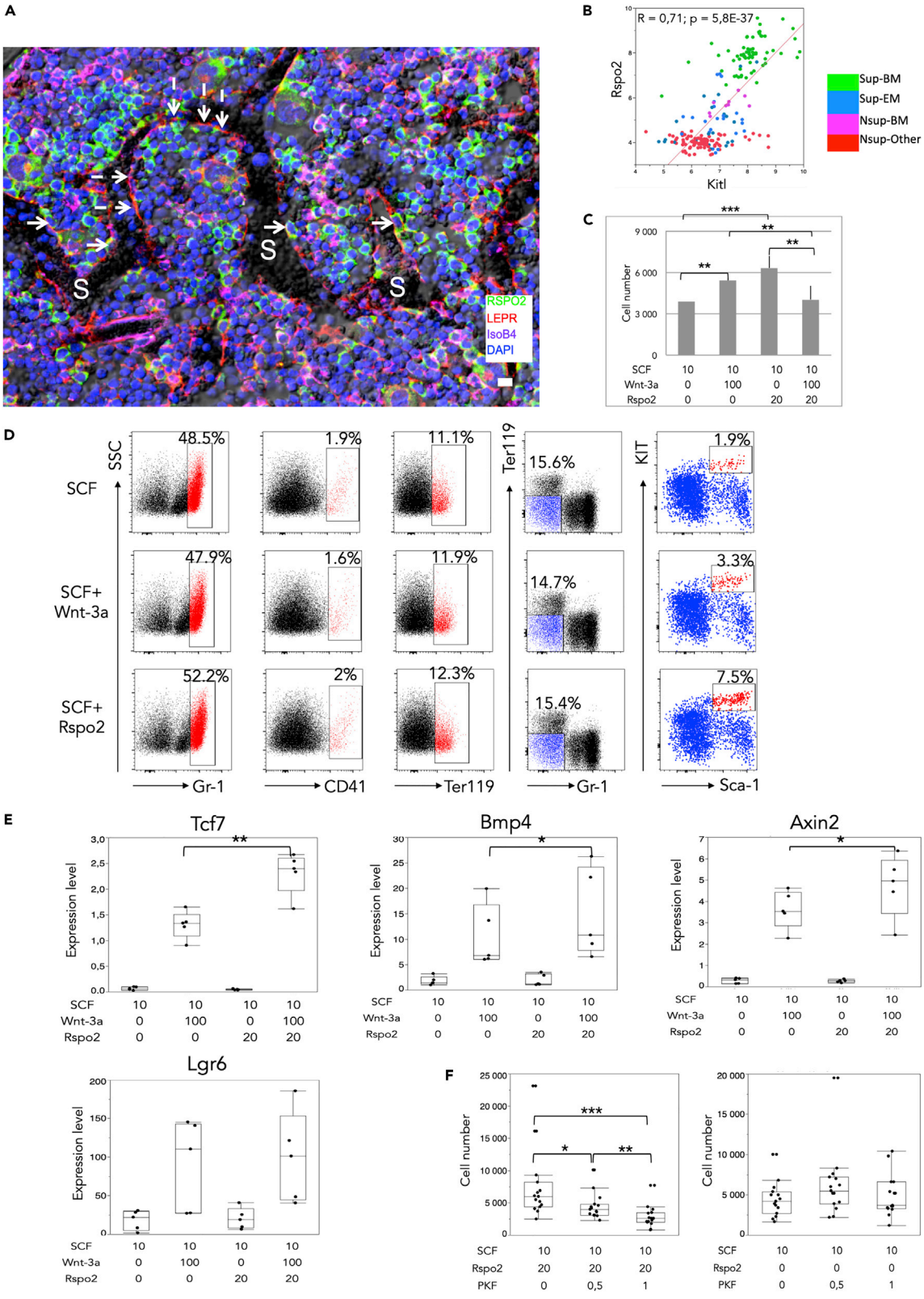

A
RSPO2
LEPR
IsoB4
DAPI
B
R = 0,71; p = 5,8E-37
Rspo2
Kitl
Sup-BM
Sup-EM
Nsup-BM
Nsup-Other
C
Cell number
SCF
Wnt-3a
Rspo2
D
SSC
Ter119
KIT
SCF
SCF+ Wnt-3a
SCF+ Rspo2
48.5%
1.9%
11.1%
15.6%
1.9%
47.9%
1.6%
11.9%
14.7%
3.3%
52.2%
2%
12.3%
15.4%
7.5%
Gr-1
CD41
Ter119
Gr-1
Sca-1
E
Tcf7
Bmp4
Axin2
Lgr6
Expression level
F
Cell number
PKF

Figure 5. The Wnt Signaling Pathway Facilitator R-Spondin 2 Acts in Concert with the Stem Cell Factor (SCF) to Amplify *Ex Vivo* Hematopoietic Precursors

(A) *Expression of* R-spondin 2 *in peri-sinusoidal cells*. R-spondin 2-positive/Leptin receptor-positive (LEPR) peri-sinusoidal cells: continuous arrows; R-spondin 2-negative/LEPR-positive peri-sinusoidal cells: discontinuous arrows; S, sinusoids; IsoB4, isolectin B4 labeling of endothelial cells and macrophages. Scale bar, 10 μm.

(B) Correlation of *Rspo2* to *Kitl* expression values in all studied samples.

(C) *Effect of* R-spondin 2 *on* $Lin^{neg}/Sca\text{-}1^{pos}/Kitl^{pos}$ *(LSK) cell number*. SCF (10 ng/mL) was present in all tested conditions, whereas Wnt-3a (100 ng/mL) and R-spondin 2 (20 ng/mL) were as indicated. N = 60 per condition (12 experiments with 5 wells per condition). Data are presented as mean ± SEM. Paired t test ***p < 0.0001, **p < 0.001 in paired comparisons.

(D) *Effect of* R-spondin 2 *on LSK cell phenotype (one representative experiment out of 8)*. Flow cytometry analysis was carried out on cells collected at day 12 and cultured in the presence of 10 ng/mL SCF alone (upper panels), 100 ng/mL SCF + Wnt-3a (middle), or 20 ng/mL SCF + R-spondin 2 (lower). To obtain a more precise evaluation of the $Sca\text{-}1^{pos}/Kit^{pos}$ cells we first selected the $Ter119^{neg}/Gr\text{-}1^{neg}$ cells before evaluating in this gate the % of $Sca\text{-}1^{pos}/Kit^{pos}$ cells.

(E) *Effect of* R-spondin 2 *and* Wnt-3a *on Wnt downstream targets Tcf7, Bmp4, and Axin2 and on the R-spondin receptor Lgr6*. Expression level as $2^{-\Delta Cq}$, ΔCq = Cq(gene of interest) - Cq(housekeeping gene), housekeeping gene being *Gapdh*. N = 5 per condition. Box limits are the 25th and 75th percentages, center lines are mean, and the whiskers are the mean SD values. Data comparison by paired t test. **p < 0.01, *p < 0.05 in paired comparisons.

(F) *Effect of the ß-catenin signaling inhibitor increasing concentrations of PKF118-310 (ng/mL) in the presence (left) or absence (right) of R-spondin 2 on LSK cell number*. N = 14 per condition. Legend as in (E). *p < 0.05, **p < 0.01, ***p < 0.001 in paired comparisons.

validate the functional consequences of *Rspo2* expression, we studied the effect of the recombinant R-spondin 2 on hematopoietic precursors in the presence or not of recombinant SCF and compared the effect with that of the canonical Wnt pathway effector Wnt-3a. $Lin^{neg}/Sca\text{-}1^{pos}/Kit^{pos}$ (LSK) cells were sorted from the BM of C57BL/6 mice and seeded in methylcellulose for colony-forming unit (CFU) evaluation in the presence of a range of R-spondin 2 concentrations (10, 20, and 50 ng/mL) with and without hematopoietic cytokines and Wnt-3a (100 ng/mL). CFUs were observed only in the presence of cytokines, without additional effect of Wnt-3a or R-spondin 2 (not shown). In a second set of experiments sorted LSK cells were seeded in 96-well plates (200 cells per well) in the presence of 10 ng/mL SCF, 20 ng/mL R-spondin 2, or 100 ng/mL Wnt-3a or a combination of these. After 12 days in culture in the presence of SCF alone the number of cells increased by 19-fold (Figure 5C) and cells differentiated into a majority of $Gr\text{-}1^{pos}$ granulocytes and monocytes (≈49% of the events), $CD41^{pos}$ megakaryocytes (≈2%), and $Ter119^{pos}$ erythroblasts (≈12%), whereas the percentage of $Sca\text{-}1^{pos}/Kit^{pos}$ dropped to ≈1.9% (Figure 5D). No growth was observed in the presence of R-spondin 2 or Wnt-3a without SCF (not shown). When R-spondin 2 and Wnt-3a were added to SCF the number of cells increased by 32- and 27-fold, respectively (Figure 5C) and the pattern of myeloid differentiation was unchanged (Figure 5D). On the contrary, there was no amplification in cultures supplemented with the three factors (SCF + R-spondin 2 + Wnt-3a) when compared with cultures with SCF alone. When compared with cultures with SCF alone, increases in amplification factor were highly significant in cultures with SCF + R-spondin 2 or SCF + Wnt-3a (Figure 5C). The percentage of $Sca\text{-}1^{pos}/Kit^{pos}$ increased by 3- to 4-fold in cultures with SCF + R-spondin 2, but only in five of eight cases (Figure 5D).

To verify that the proliferative effect of R-spondin 2 was mediated through the Wnt canonical pathway we carried out two sets of experiments. First, using qRT-PCR, the levels of the major Wnt downstream targets *Axin2*, *Tcf7*, and *Bmp4* were measured after 24 h of treatment with R-spondin 2, Wnt-3a, and their combination (Figure 5E). These experiments showed that R-spondin 2 had no detectable effect by itself, contrarily to Wnt-3a that induced, as expected, a significant increase. Importantly, the combination of Wnt-3a and R-spondin 2 significantly increased the effect even more when compared with Wnt-3a alone, confirming the facilitating role of R-spondin 2. Interestingly, in our culture system Wnt-3a significantly increased the expression of the R-spondin receptor *Lgr6* without additional effect of R-spondin 2 (Figure 5E), whereas the levels of the other receptors *Lgr4* and *Lgr5* remained low to nil (not shown). Second, sorted LSK cells were cultured in the presence of R-spondin 2 and with the β-catenin signaling chemical inhibitor PKF118-310 at increasing concentrations (Figure 5F). The inhibitor in the presence of R-spondin 2 induced a significant dose-response inhibition of LSK proliferation, which was not observed in the absence of R-spondin 2.

Taken together, these data indicate that, in the presence of SCF, R-spondin 2 has a direct proliferative effect on LSK cells, but does not modify the myeloid pattern of differentiation. This effect is mediated by the Wnt canonical pathway, which confirms the facilitating role of R-spondin 2 in our experimental setting.

DISCUSSION

In this study on the BM microenvironment using a combination of algorithms, one of which recently applied to identify reliable paths and network v-structures in complex systems (Sella et al., 2018; Verny et al., 2017),

we have identified a gene network characteristic of the HSC-supportive capacity of BM stromal cells and then defined gene networks for the different supportive niche cell populations contributing to that phenotype, i.e., perivascular cells (V), osteoblasts (O), and MSCs including perivascular-primed (MSC-Vp) or osteoblast-primed (MSC-Op) cells. Comparison of the gene networks specific to the niche cell populations with the network characterizing the HSC-supportive capacity as a whole made clear that the supportive phenotype resulted from the coordinated effect of several niche cell populations. The major hematopoietic regulator *Kitl* was a major hub in the network characterizing the supportive perivascular stromal cells, and the facilitator of the Wnt signaling pathway *Rspo2* appeared as the best predictor of the supportive capacity and was directly linked to *Kitl* in the perivascular cell network. We have therefore validated the biological relevance of R-spondin 2 by showing that this protein enhanced the proliferation of hematopoietic precursors in the presence of SCF. This study by identifying many classifiers and hubs beside *Rspo2* and *Kitl* constitutes a resource to unravel candidate BM stromal mediators.

In recently published analyses of transcriptomes where a large number of stromal cells were studied at the single-cell level the primary purpose has been to identify stromal cell sets, and the secondary purpose to infer possible capacity of support of the cell sets by unraveling their hematopoietic regulator profiles. Bearing in mind the opposite difference of approaches as here we first defined the genes that organize a functional phenotype and then investigate the cell populations contributing to that phenotype, one may attempt to compare how subsets resulting from our analysis relate to subpopulations resulting from single-cell analysis. Comparison of our gene sets with data reported by Tikhonova et al. indicates that the V and O clusters in our analysis broadly relate to adipocyte-biased P1 and P2 and osteo-primed P3 and P4 clusters of the LEPR population, respectively (Tikhonova et al., 2019). Further comparisons indicate that our V cluster is related to the MSC cluster 1 expressing *Lepr* in Baryawno et al. (Baryawno et al., 2019), to the MSC P3 cluster in Wolock et al. (Wolock et al., 2019), and to the cluster of peri-sinusoidal stromal cells in Balzano et al. (Balzano et al., 2019). In our work, MSCs, prospectively specified using a distinct phenotype or selected by growth in culture, were found to constitute a heterogeneous population including cells expressing genes implicated either in ossification (MSC-Op) or in vascularization (MSC-Vp). Taken together our work and the works using single-cell analysis identify a perivascular cell population with HSC-supportive capacity, in part due to the expression of known hematopoietic regulators. Moreover, whereas single-cell analyses help establish subsets within a mixed population and unravel differentiation trajectories between these subsets, network analysis gives important insight in the organization of gene expression within the population as a whole and indicates how networks in different cell sub-populations are coordinated to orchestrate the global functional phenotype.

As analyses, such as WGCNA, are optimally suited to identify correlated networks, but do not identify direct gene-to-gene interactions or potential causal links, we used miic analysis (Sella et al., 2018; Verny et al., 2017), recently extended to analyze continuous or mixed-type (e.g., continuous-categorical) data (Cabeli et al., 2020). The miic algorithm identifies direct paths with high confidence, and, as such, reveals potential cause-and-effect links as well as latent regulators, which may not show up when looking only at gene expression data. However, to yield significant results the miic algorithm requires a large number of observations. Consequently, we took advantage of the recently published transcriptomes of single LEPR cells (Tikhonova et al., 2019) to apply the algorithm using a subset of genes characterizing the perivascular V population. This approach enabled pruning and remodeling the network defined by WGCNA, thus providing a clearer picture by clearly individualizing the major hubs in different conditions.

The combined WGCNA and miic analysis confirmed that *Kitl* was a major hub. Kit ligand is a well-established major hematopoietic regulator; its *in vivo* deletion in perivascular stromal cells leads to dramatic depletion of the BM HSC pool (Asada et al., 2017; Ding et al., 2012). Moreover, in our miic analysis, we found that *Kitl* was directly linked to *Rspo2*, *Il7*, *Ebf3,* and *Snai2*, all genes known as hematopoietic regulators (this study and Balzano et al., 2019; Charbord et al., 2014; Seike et al., 2018). Furthermore, we also identified the C-type lectin *Cd302* as another major hub. CD302 has been reported as implicated in dendritic cell migration (Lo et al., 2016), but it remains to be studied how this lectin is involved in the stromal niche regulation of HSCs.

In cells exposed to stress, we observed remodeling of the gene networks with a large increase in the number of direct paths and the appearance of new hubs, such as the cytoskeletal Ca^{2+} regulator *Marcks*, suggesting that the cells had modified their actin cytoskeleton pattern with possible effect on migration and

adhesion. However, the hubs observed in baseline condition, in particular *Cd302* and *Kitl*, were still present with maintenance of their major connections, suggesting that the gene organization observed at baseline is not replaced, but augmented under condition of stress.

In our analysis *Rspo2* appeared as the best predictor of the Sup-BM capacity and was directly linked to *Kitl* in the perivascular cell network. R-spondin 2 is a facilitator of the Wnt signaling pathway. By binding to one of its Leucine-rich repeat-containing G-protein-coupled Receptors (LGR4-6) it forms a complex that hampers the degradation of the Wnt frizzled receptors through the RNF43 and ZNFR3 transmembrane ubiquitin ligases, thereby boosting Wnt signaling strength (de Lau et al., 2014; Niehrs, 2012). Both Wnt and R-spondin ligands are essential to impede the differentiation of Lgr5-positive intestinal stem cells and maintain their self-renewal capacity (Yan et al., 2017). The reported effects of Wnt signaling and stromal Wnt inhibitors have underscored the dose and environmental context dependency of the different mediators on HSC cycling and their regenerative capacity (Fleming et al., 2008; Renstrom et al., 2009; Ruf et al., 2016; Staal et al., 2016). We found that R-spondin 2 enhanced the proliferation of hematopoietic precursors in the presence of SCF. This effect was Wnt dependent, which is in agreement with similar studies on R-spondin in other cell types (Kazanskaya et al., 2004; van Andel et al., 2017). The effect might be mediated by the binding of R-spondin 2 to LGR6, which, we found, is the only receptor whose transcript is expressed at a significant level by the hematopoietic precursors. Alternatively, R-spondin 2 may regulate frizzled receptor stability through the RNF43/ZNFR3 ligases independently of the LGRs, through an as yet undefined co-receptor (Lebensohn and Rohatgi, 2018; Szenker-Ravi et al., 2018). By analyzing the recently published transcriptomes of single LEPR cells we found that the R-spondin 2-expressing cell population was a subset of LEPR perivascular cells, which was confirmed by the examination of BM biopsies showing a subset of R-spondin 2-positive cells on the abluminal side of sinusoidal endothelial cells. R-spondin 2 was also expressed in some Leptin receptor-negative hematopoietic cells, which is in agreement with the expression, mostly in the erythroblastic series, reported in Gene Expression Commons (https://gexc.riken.jp). Taken together, our analysis identifies R-spondin 2 as a yet unreported Wnt signaling pathway-associated stromal regulator of hematopoiesis.

Second to *Rspo2*, the genes coding for the soluble mediators *Fam20a* and *Srpx2* were highly connected best predictors of the Sup-BM capacity. *Fam20a* has been isolated from a hematopoietic cell line treated with SCF and is expressed in myeloid progenitors (Nalbant et al., 2005). This gene whose mutations are responsible for amelogenesis imperfecta and nephrocalcinosis codes for a secreted pseudokinase implicated in biomineralization (Ohyama et al., 2016). *Srpx2*, an X-linked gene acting as a ligand for the urokinase plasminogen activator surface receptor, plays a role in angiogenesis by inducing endothelial cell migration and the formation of vascular network (Miljkovic-Licina et al., 2009). *Srpx2* is upregulated in certain types of telocytes that are interstitial cells probably of mesenchymal origin found in multiple tissues (Zhu et al., 2015). Of note, another gene belonging to the Sup-BM module, *Svep1*, a ligand for integrin $\alpha 9\beta 1$ that has been isolated from the BM stromal line MS-5 (Gilges et al., 2000), is also upregulated in certain types of telocytes (Song et al., 2019).

In conclusion, this study by combining different bioinformatics algorithms constitutes a resource to identify candidate BM stromal mediators whose effect on the different HSC attributes should be demonstrated in detail by upcoming experimental work.

Limitations of the Study

The present analysis of transcriptomes relies on published datasets, and, as such, unpublished transcriptomes of relevant cell types could not be taken into account. In addition, this work indicates that R-spondin 2 amplifies in the presence of SCF the proliferation of murine hematopoietic precursors *in vitro*. However, *in vitro* studies do not faithfully recapitulate HSC self-renewal as studied in *in vivo* transplantation assays. Finally, although Wnt signaling is mostly preserved across species, and the identified gene network included human samples, the effect of R-spondin 2 on murine cells may not entirely reproduce the response of human precursors to this Wnt mediator.

Resource Availability

Lead Contact

Further information should be directed to and will be fulfilled by the Lead Contact, Pierre Charbord (pierre.charbord@inserm.fr).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

This study did not generate new dataset, but analyzed datasets in public repositories. Accession numbers for all datasets analyzed are given in Table S1.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101222.

ACKNOWLEDGMENTS

Work supported by grants of the Fondation pour la Recherche Médicale (DEQ20100318258), Institut national de la santé et de la recherche médicale (Inserm), Centre national de la recherche scientifique (CNRS), and German Research Foundation (RO: DFG OO8/16). We thank Sophie Gournet for revising the illustrations.

AUTHOR CONTRIBUTIONS

Conceptualization: P.C. and C. Desterke; Methodology: P.C., H.I., and C. Desterke; Software: C. Desterke, N.S., and V.C.; Formal analysis: P.C., H.I., and C. Desterke; Investigation: L.P., L.C., and N.C.; Resources: T.J.; Writing -Original Draft: P.C.; Writing -Review and Editing: P.C., R.A.J.O., T.J., C. Durand; Funding Acquisition: T.J.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 25, 2019
Revised: March 19, 2020
Accepted: May 27, 2020
Published: June 26, 2020

REFERENCES

Acar, M., Kocherlakota, K.S., Murphy, M.M., Peyer, J.G., Oguro, H., Inra, C.N., Jaiyeola, C., Zhao, Z., Luby-Phelps, K., and Morrison, S.J. (2015). Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. Nature *526*, 126–130.

Asada, N., Kunisaki, Y., Pierce, H., Wang, Z., Fernandez, N.F., Birbrair, A., Ma'ayan, A., and Frenette, P.S. (2017). Differential cytokine contributions of perivascular haematopoietic stem cell niches. Nat. Cell Biol. *19*, 214–223.

Balzano, M., De Grandis, M., Vu Manh, T.P., Chasson, L., Bardin, F., Farina, A., Serge, A., Bidaut, G., Charbord, P., Herault, L., et al. (2019). Nidogen-1 contributes to the interaction network involved in pro-B cell retention in the peri-sinusoidal hematopoietic stem cell niche. Cell Rep. *26*, 3257–3271.e8.

Baryawno, N., Przybylski, D., Kowalczyk, M.S., Kfoury, Y., Severe, N., Gustafsson, K., Kokkaliaris, K.D., Mercier, F., Tabaka, M., Hofree, M., et al. (2019). A cellular taxonomy of the bone marrow stroma in homeostasis and leukemia. Cell *177*, 1915–1932.e6.

Bergwerff, M., Gittenberger-de Groot, A.C., Wisse, L.J., DeRuiter, M.C., Wessels, A., Martin, J.F., Olson, E.N., and Kern, M.J. (2000). Loss of function of the Prx1 and Prx2 homeobox genes alters architecture of the great elastic arteries and ductus arteriosus. Virchows Arch. *436*, 12–19.

Breiman, L. (2001). Random forests. Mach. Learn. *45*, 5–32.

Cabeli, V., Verny, L., Sella, N., Uguzzoni, U., Verny, M., and Isambert, H. (2020). Learning clinical networks from medical records based on information estimates in mixed-type data. PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1007866.

Chan, C.K., Lindau, P., Jiang, W., Chen, J.Y., Zhang, L.F., Chen, C.C., Seita, J., Sahoo, D., Kim, J.B., Lee, A., et al. (2013). Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. Proc. Natl. Acad. Sci. U S A *110*, 12643–12648.

Chan, C.K.F., Gulati, G.S., Sinha, R., Tompkins, J.V., Lopez, M., Carter, A.C., Ransom, R.C., Reinisch, A., Wearda, T., Murphy, M., et al. (2018). Identification of the human skeletal stem cell. Cell *175*, 43–56.e21.

Charbord, P., Pouget, C., Binder, H., Dumont, F., Stik, G., Levy, P., Allain, F., Marchal, C., Richter, J., Uzan, B., et al. (2014). A systems biology approach for defining the molecular framework of the hematopoietic stem cell niche. Cell Stem Cell *15*, 376–391.

Chen, J.Y., Miyanishi, M., Wang, S.K., Yamazaki, S., Sinha, R., Kao, K.S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I.L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature *530*, 223–227.

Cordeiro Gomes, A., Hara, T., Lim, V.Y., Herndler-Brandstetter, D., Nevius, E., Sugiyama, T., Tani-Ichi, S., Schlenner, S., Richie, E., Rodewald, H.R., et al. (2016). Hematopoietic stem cell niches produce lineage-instructive signals to control multipotent progenitor differentiation. Immunity *45*, 1219–1231.

de Lau, W., Peng, W.C., Gros, P., and Clevers, H. (2014). The R-spondin/Lgr5/Rnf43 module:

regulator of Wnt signal strength. Genes Dev. *28*, 305–316.

Decker, M., Martinez-Morentin, L., Wang, G., Lee, Y., Liu, Q., Leslie, J., and Ding, L. (2017). Leptin-receptor-expressing bone marrow stromal cells are myofibroblasts in primary myelofibrosis. Nat. Cell Biol. *19*, 677–688.

Ding, L., and Morrison, S.J. (2013). Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature *495*, 231–235.

Ding, L., Saunders, T.L., Enikolopov, G., and Morrison, S.J. (2012). Endothelial and perivascular cells maintain haematopoietic stem cells. Nature *481*, 457–462.

Eash, K.J., Greenbaum, A.M., Gopalan, P.K., and Link, D.C. (2010). CXCR2 and CXCR4 antagonistically regulate neutrophil trafficking from murine bone marrow. J. Clin. Invest. *120*, 2423–2431.

Fleming, H.E., Janzen, V., Lo Celso, C., Guo, J., Leahy, K.M., Kronenberg, H.M., and Scadden, D.T. (2008). Wnt signaling in the niche enforces hematopoietic stem cell quiescence and is necessary to preserve self-renewal in vivo. Cell Stem Cell *2*, 274–283.

Gilges, D., Vinit, M.A., Callebaut, I., Coulombel, L., Cacheux, V., Romeo, P.H., and Vigon, I. (2000). Polydom: a secreted protein with pentraxin, complement control protein, epidermal growth factor and von Willebrand factor A domains. Biochem. J. *352* (Pt 1), 49–59.

Greenbaum, A., Hsu, Y.M., Day, R.B., Schuettpelz, L.G., Christopher, M.J., Borgerding, J.N., Nagasawa, T., and Link, D.C. (2013). CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature *495*, 227–230.

Horvath, S. (2011). Weighted Network Analysis. Applications in Genomics and Systems Biology (Springer).

Hu, X., Garcia, M., Weng, L., Jung, X., Murakami, J.L., Kumar, B., Warden, C.D., Todorov, I., and Chen, C.C. (2016). Identification of a common mesenchymal stromal progenitor for the adult haematopoietic niche. Nat. Commun. *7*, 13095.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Isern, J., Martin-Antonio, B., Ghazanfari, R., Martin, A.M., Lopez, J.A., del Toro, R., Sanchez-Aguilera, A., Arranz, L., Martin-Perez, D., Suarez-Lledo, M., et al. (2013). Self-renewing human bone marrow mesenspheres promote hematopoietic stem cell expansion. Cell Rep. *3*, 1714–1724.

Kazanskaya, O., Glinka, A., del Barco Barrantes, I., Stannek, P., Niehrs, C., and Wu, W. (2004). R-Spondin2 is a secreted activator of Wnt/beta-catenin signaling and is required for Xenopus myogenesis. Dev. Cell *7*, 525–534.

Kfoury, Y., and Scadden, D.T. (2015). Mesenchymal cell contributions to the stem cell niche. Cell Stem Cell *16*, 239–253.

Kunisaki, Y., Bruns, I., Scheiermann, C., Ahmed, J., Pinho, S., Zhang, D., Mizoguchi, T., Wei, Q., Lucas, D., Ito, K., et al. (2013). Arteriolar niches maintain haematopoietic stem cell quiescence. Nature *502*, 637–643.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lebensohn, A.M., and Rohatgi, R. (2018). R-spondins can potentiate WNT signaling without LGRs. Elife *7*, e33126.

Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lo, T.H., Silveira, P.A., Fromm, P.D., Verma, N.D., Vu, P.A., Kupresanin, F., Adam, R., Kato, M., Cogger, V.C., Clark, G.J., et al. (2016). Characterization of the expression and function of the C-type lectin receptor CD302 in mice and humans reveals a role in dendritic cell migration. J. Immunol. *197*, 885–898.

Mead, A.J., Neo, W.H., Barkas, N., Matsuoka, S., Giustacchini, A., Facchini, R., Thongjuea, S., Jamieson, L., Booth, C.A.G., Fordham, N., et al. (2017). Niche-mediated depletion of the normal hematopoietic stem cell reservoir by Flt3-ITD-induced myeloproliferation. J. Exp. Med. *214*, 2005–2021.

Mendelson, A., and Frenette, P.S. (2014). Hematopoietic stem cell niche maintenance during homeostasis and regeneration. Nat. Med. *20*, 833–846.

Mendez-Ferrer, S., Michurina, T.V., Ferraro, F., Mazloom, A.R., Macarthur, B.D., Lira, S.A., Scadden, D.T., Ma'ayan, A., Enikolopov, G.N., and Frenette, P.S. (2010). Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature *466*, 829–834.

Miljkovic-Licina, M., Hammel, P., Garrido-Urbani, S., Bradfield, P.F., Szepetowski, P., and Imhof, B.A. (2009). Sushi repeat protein X-linked 2, a novel mediator of angiogenesis. FASEB J. *23*, 4105–4116.

Morikawa, S., Mabuchi, Y., Kubota, Y., Nagai, Y., Niibe, K., Hiratsu, E., Suzuki, S., Miyauchi-Hara, C., Nagoshi, N., Sunabori, T., et al. (2009). Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. J. Exp. Med. *206*, 2483–2496.

Morrison, S.J., and Scadden, D.T. (2014). The bone marrow niche for haematopoietic stem cells. Nature *505*, 327–334.

Nakamura, Y., Arai, F., Iwasaki, H., Hosokawa, K., Kobayashi, I., Gomei, Y., Matsumoto, Y., Yoshihara, H., and Suda, T. (2010). Isolation and characterization of endosteal niche cell populations that regulate hematopoietic stem cells. Blood *116*, 1422–1432.

Nalbant, D., Youn, H., Nalbant, S.I., Sharma, S., Cobos, E., Beale, E.G., Du, Y., and Williams, S.C. (2005). FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. BMC Genomics *6*, 11.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nat. Rev. Mol. Cell Biol. *13*, 767–779.

Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell *13*, 102–116.

Ohyama, Y., Lin, J.H., Govitvattana, N., Lin, I.P., Venkitapathi, S., Alamoudi, A., Husein, D., An, C., Hotta, H., Kaku, M., et al. (2016). FAM20A binds to and regulates FAM20C localization. Sci. Rep. *6*, 27784.

Pinho, S., and Frenette, P.S. (2019). Haematopoietic stem cell activity and interactions with the niche. Nat. Rev. Mol. Cell Biol. *20*, 303–320.

Pinho, S., Lacombe, J., Hanoun, M., Mizoguchi, T., Bruns, I., Kunisaki, Y., and Frenette, P.S. (2013). PDGFRalpha and CD51 mark human Nestin+ sphere-forming mesenchymal stem cells capable of hematopoietic progenitor cell expansion. J. Exp. Med. *210*, 1351–1367.

Renstrom, J., Istvanffy, R., Gauthier, K., Shimono, A., Mages, J., Jardon-Alvarez, A., Kroger, M., Schiemann, M., Busch, D.H., Esposito, I., et al. (2009). Secreted frizzled-related protein 1 extrinsically regulates cycling activity and maintenance of hematopoietic stem cells. Cell Stem Cell *5*, 157–167.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Ruf, F., Schreck, C., Wagner, A., Grziwok, S., Pagel, C., Romero, S., Kieslinger, M., Shimono, A., Peschel, C., Gotze, K.S., et al. (2016). Loss of Sfrp2 in the niche amplifies stress-induced cellular responses, and impairs the in vivo regeneration of the hematopoietic stem cell pool. Stem Cells *34*, 2381–2392.

Sacchetti, B., Funari, A., Michienzi, S., Di Cesare, S., Piersanti, S., Saggio, I., Tagliafico, E., Ferrari, S., Robey, P.G., Riminucci, M., et al. (2007). Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. Cell *131*, 324–336.

Schepers, K., Pietras, E.M., Reynaud, D., Flach, J., Binnewies, M., Garg, T., Wagers, A.J., Hsiao, E.C., and Passegue, E. (2013). Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. Cell Stem Cell *13*, 285–299.

Seike, M., Omatsu, Y., Watanabe, H., Kondoh, G., and Nagasawa, T. (2018). Stem cell niche-specific Ebf3 maintains the bone marrow cavity. Genes Dev. *32*, 359–372.

Sella, N., Verny, L., Uguzzoni, G., Affeldt, S., and Isambert, H. (2018). MIIC online: a web server to reconstruct causal or non-causal networks from non-perturbative data. Bioinformatics *34*, 2311–2313.

Shoshani, O., Ravid, O., Massalha, H., Aharonov, A., Ovadya, Y., Pevsner-Fischer, M., Leshkowitz, D., and Zipori, D. (2014). Cell isolation induces fate changes of bone marrow mesenchymal cells leading to loss or alternatively to acquisition of new differentiation potentials. Stem Cells *32*, 2008–2020.

Silberstein, L., Goncalves, K.A., Kharchenko, P.V., Turcotte, R., Kfoury, Y., Mercier, F., Baryawno, N., Severe, N., Bachand, J., Spencer, J.A., et al. (2016). Proximity-based differential single-cell analysis of the niche to identify stem/progenitor cell regulators. Cell Stem Cell *19*, 530–543.

Song, D., Xu, M., Qi, R., Ma, R., Zhou, Y., Wu, D., Fang, H., and Wang, X. (2019). Influence of gene modification in biological behaviors and responses of mouse lung telocytes to inflammation. J. Transl Med. *17*, 158.

Staal, F.J., Chhatta, A., and Mikkers, H. (2016). Caught in a Wnt storm: complexities of Wnt signaling in hematopoiesis. Exp. Hematol. *44*, 451–457.

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.

Sugiyama, T., Kohara, H., Noda, M., and Nagasawa, T. (2006). Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. Immunity *25*, 977–988.

Szenker-Ravi, E., Altunoglu, U., Leushacke, M., Bosso-Lefevre, C., Khatoo, M., Thi Tran, H., Naert, T., Noelanders, R., Hajamohideen, A., Beneteau, C., et al. (2018). RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6. Nature *557*, 564–569.

Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc. Natl. Acad. Sci. U S A *99*, 6567–6572.

Tikhonova, A.N., Dolgalev, I., Hu, H., Sivaraj, K.K., Hoxha, E., Cuesta-Dominguez, A., Pinho, S., Akhmetzyanova, I., Gao, J., Witkowski, M., et al. (2019). The bone marrow microenvironment at single-cell resolution. Nature *569*, 222–228.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U S A *98*, 5116–5121.

van Andel, H., Ren, Z., Koopmans, I., Joosten, S.P., Kocemba, K.A., de Lau, W., Kersten, M.J., de Bruin, A.M., Guikema, J.E., Clevers, H., et al. (2017). Aberrantly expressed LGR4 empowers Wnt signaling in multiple myeloma by hijacking osteoblast-derived R-spondins. Proc. Natl. Acad. Sci. U S A *114*, 376–381.

Verny, L., Sella, N., Affeldt, S., Singh, P.P., and Isambert, H. (2017). Learning causal networks with latent variables from multivariate information in genomic data. PLoS Comput. Biol. *13*, e1005662.

Wolock, S.L., Krishnan, I., Tenen, D.E., Matkins, V., Camacho, V., Patel, S., Agarwal, P., Bhatia, R., Tenen, D.G., Klein, A.M., et al. (2019). Mapping distinct bone marrow niche populations and their differentiation paths. Cell Rep. *28*, 302–311.e5.

Yan, K.S., Janda, C.Y., Chang, J., Zheng, G.X.Y., Larkin, K.A., Luca, V.C., Chia, L.A., Mah, A.T., Han, A., Terry, J.M., et al. (2017). Non-equivalence of Wnt and R-spondin ligands during Lgr5(+) intestinal stem-cell self-renewal. Nature *545*, 238–242.

Zhao, M., Tao, F., Venkatraman, A., Li, Z., Smith, S.E., Unruh, J., Chen, S., Ward, C., Qian, P., Perry, J.M., et al. (2019). N-Cadherin-Expressing bone and marrow stromal progenitor cells maintain reserve hematopoietic stem cells. Cell Rep. *26*, 652–669.e6.

Zhou, B.O., Yue, R., Murphy, M.M., Peyer, J.G., and Morrison, S.J. (2014). Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. Cell Stem Cell *15*, 154–168.

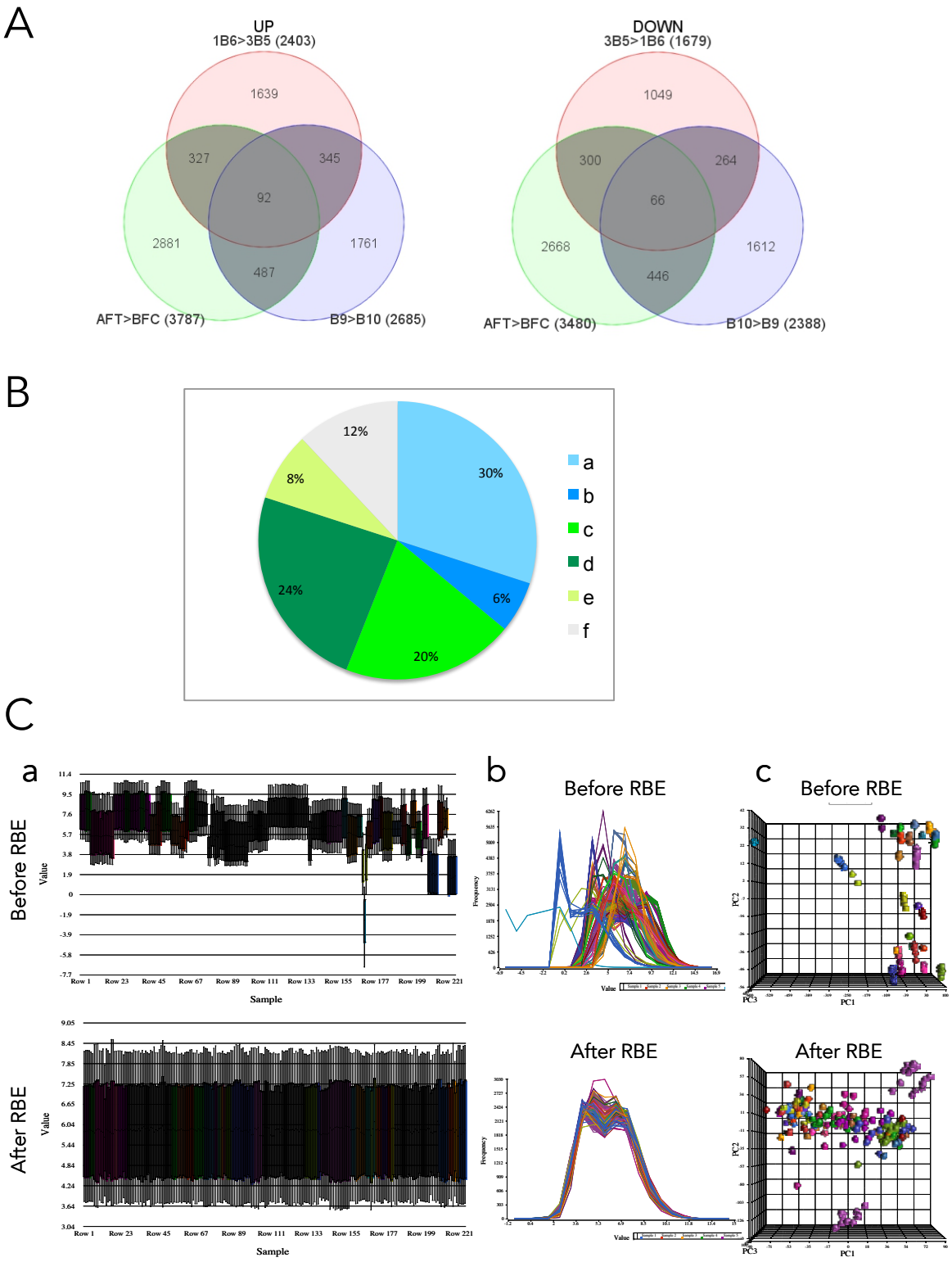
Zhu, Y., Zheng, M., Song, D., Ye, L., and Wang, X. (2015). Global comparison of chromosome X genes of pulmonary telocytes with mesenchymal stem cells, fibroblasts, alveolar type II cells, airway epithelial cells, and lymphocytes. J. Transl Med. *13*, 318.

iScience, Volume *23*

Supplemental Information

Inferring Gene Networks in Bone Marrow Hematopoietic Stem Cell-Supporting Stromal Niche Populations

Christophe Desterke, Laurence Petit, Nadir Sella, Nathalie Chevallier, Vincent Cabeli, Laura Coquelin, Charles Durand, Robert A.J. Oostendorp, Hervé Isambert, Thierry Jaffredo, and Pierre Charbord



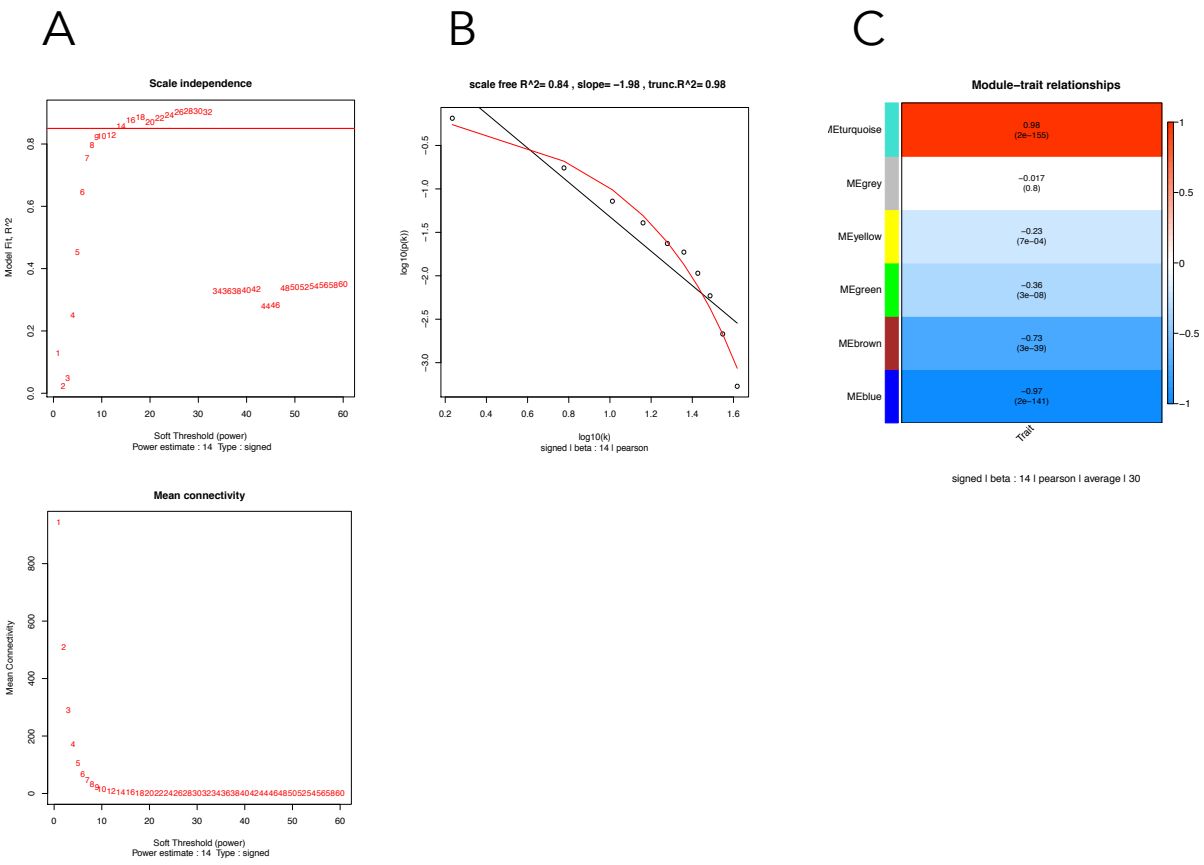
Supplemental Figure 1 (related to Figure 1): Study flowchart and cell cluster identification

A: Venn diagrams leading to the identification of the Set1 gene set. The gene sets obtained after ANOVA comparing transcriptomes of supportive vs. less or non supportive stromal lines from each developmental site were intersected. The resulting Venn diagrams for the genes UP- and DOWN-regulated in the supportive lines are shown on the left and right panels, respectively. Set 1 corresponds to the shaded areas. Abbreviations for the cell lines: 1B6: supportive AGM UG26.1B6;

3B5: less supportive AGM UG26.3B5; AFT: supportive FL AFT024; BFC: non-supportive FL BFC012; B9: supportive BM BMC9; B10: less supportive BM BMC10.

B: The relative relevance of the criteria evidencing the supportive capacity of stromal cells in the ensemble of articles corresponding to the datasets analyzed in this work. a) generation of hematopoietic colonies in co-culture of HSPCs with stromal cells, b) hematopoietic differentiation of embryonic stem cells co-cultured with stromal cells, c) *in vivo* co-localization of stromal cells expressing specific markers with HSCs, d) lineage reconstitution *in vivo* by HSCs co-cultured with stromal cells for variable time spans, e) reduction of the HSC pool *in vivo* after depletion of the stromal cell population, and f) characteristic phenotype.

C: Effect of Removal Batch Effect on Quality Control parameters using the matrix consisting in the 224 observations (samples) and the 15056 variables (entire transcriptome after dataset merging). Mean value and dispersion of the expression level of all genes in a sample (a), distribution of the gene expression levels (b), and representation of the samples in the PCA space (c).

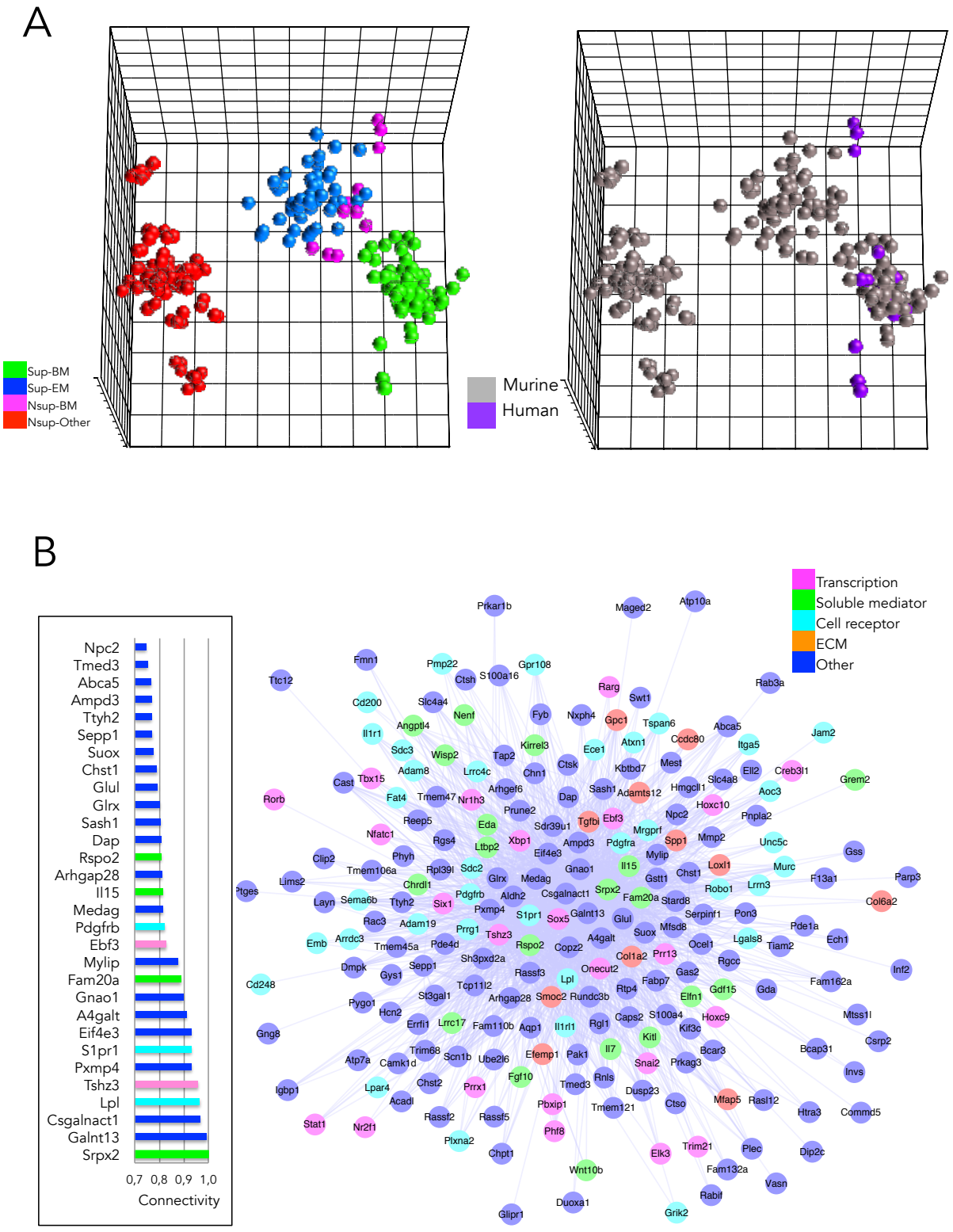


Supplemental Figure 2 (related to Figure 2): WGCNA of the 1869 Set 1 genes

*A: Choice of the power ß for the weighted adjacencies***.** Upper panel: scale free index fit (R^2) vs. ß power; ß=14 is chosen as the value where saturation is reached and above 0,8. Lower panel: connectivity distribution according to ß power.

B: Verification of the scale-free topology index for the chosen ß power value. Model fit for a power law (black) and an exponentially truncated power law (red).

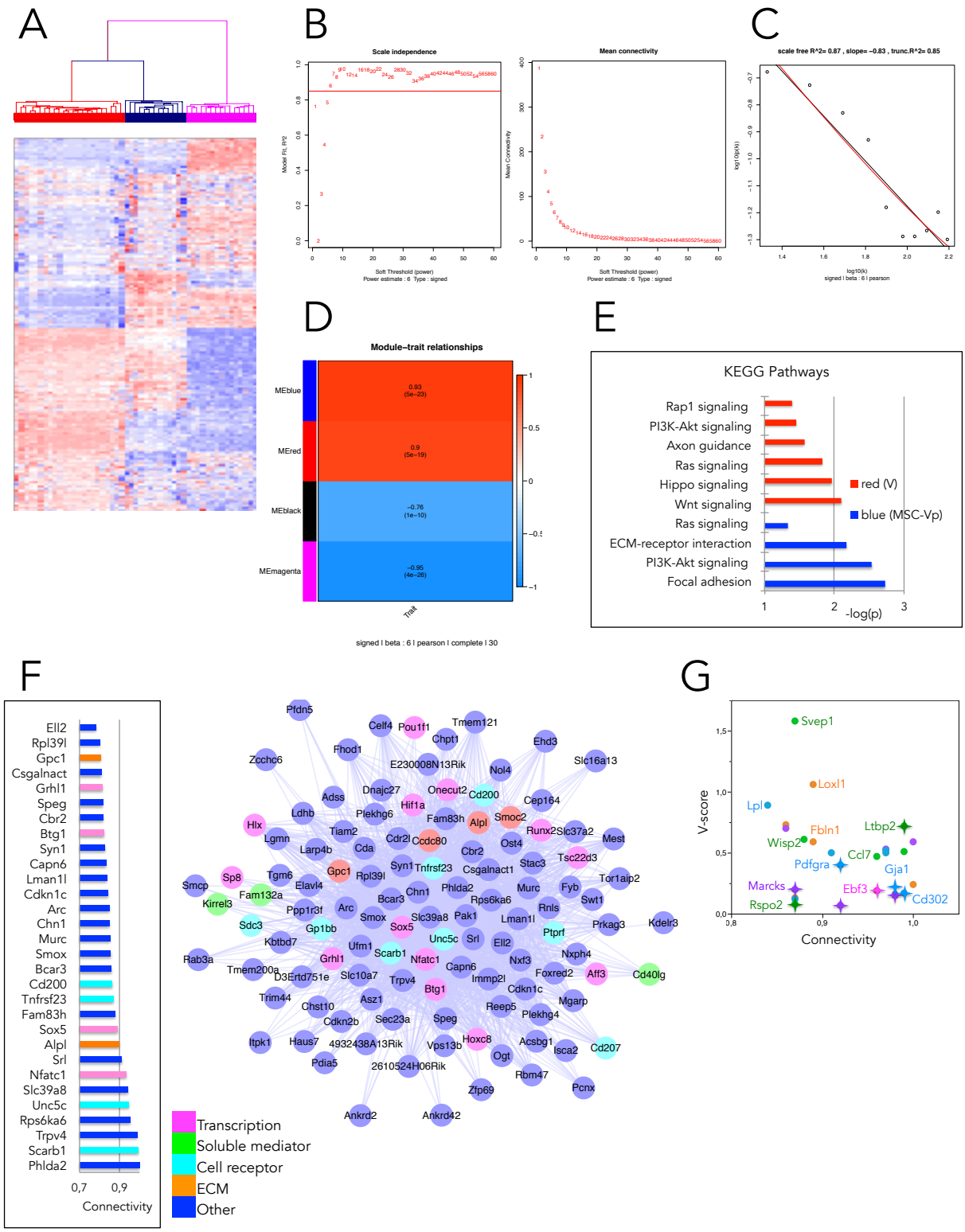
C: Module trait relationships. Pearson's correlation coefficient and corresponding p-value are given for each of the identified modules.



Supplemental Figure 3 (related to Figure 2): A similar gene network characterizes the BM stromal supportive capacity after inclusion of human samples

A: PCA using 248 samples as observations and 1619 genes as variables. The 3D score plot (PC1, PC1, PC3) is shown on the left panel (same sample colors as in Figure 1B). The right panel highlights the samples according to species (human vs. murine samples).

B: Cytoscape connectivity plot (right) and intra-modular connectivity bar plot of the Top-30 genes (left) for the turquoise module corresponding to the Sup-BM cluster.



Supplemental Figure 4 (related to Figure 3): The supportive capacity is implemented in four BM niche cell populations

A: Hierarchical clustering corresponding to O, V and MSC samples. Euclidian distances and Ward's linkage were used.

B: Choice of the power ß for the weighted adjacencies. Left panel: scale free index fit (R^2) vs. ß power; ß=6 is chosen as the value where saturation is reached and above 0,8. Right panel:

connectivity distribution according to ß power.

C: Verification of the scale-free topology index for the chosen ß power value. Model fit for a power law (black) and an exponentially truncated power law (red).

D: Module trait relationships. Pearson's correlation coefficient and corresponding p-value are given for each of the identified modules.

E: Significant KEGG pathways corresponding to genes in V (red) and MSC-Vp (blue) populations.

F: Connectivity plots of genes belonging the O module. Right panel: Intra-modular connectivity bar plot of the Top-30 genes. Left panel: Cytoscape connectivity plot of the 115 most connected genes; genes were selected for representation above a threshold weight ≥ 0,3110.

G: Gene connectivity vs. PAM score scatter plot for the V and MSC-Vp modules. The genes belonging to the V module are shown as stars.

Transparent Methods

Bioinformatics

Data sets: The different samples used in this study are described in Supplemental Table 1. The retrieved transcriptomes were studied using Affymetrix microarrays (different platforms: Mouse 430 2.0, Mouse Gene 1.0 ST, Mouse Gene 2.0 ST, Mouse Exon 1.0 ST) or Illumina RNAseq. The data sets were merged in a single matrix. A normalization step was necessary to get rid of the covariate series (the different experiments), using the Partek 'Remove batch effect' (RBE) algorithm. Using the matrix consisting in the 224 observations (samples) and the 15056 variables (entire transcriptome after datasets merging) the results of this RBE are shown on three QC parameters: the mean value and dispersion of the expression level of all genes in a sample (A), the distribution of the gene expression levels (B), and the representation of the samples in the PCA space is shown on Supplemental Fig 1A. The RBE resulted in the alignment of mean value and dispersion of the expression level of all genes in the different samples, the normalization of the gene distributions, and the splitting of the different samples belonging to one dataset to the whole space.

PCA: It was performed on matrices indicated in the text using Partek and factominer R-package. P-value of group discrimination was evaluated by correlation on the first principal component axis (Le et al., 2008).

Discriminant analysis: To discriminate the best classifiers for the Sup-BM cluster, the learning machine algorithm 'Prediction Analysis for Microarrays' (PAMr) was applied using the 4 cell clusters as defined (Tibshirani et al., 2002). Positive score probes obtained by leave-one-out learning with cross validation for the Sup-BM cluster were retained and applied to a random forest algorithm (Breiman, 2001). Before starting the Random Forest analysis the dataset was split randomly in training set and validation set comprising 75% and 25% of the samples, respectively. The Random Forest "mtry" parameter was tuned with Caret R package on the training set in order to achieve the learning step with a minimum error of bagging. The Random Forest model was built on training, and supervised by implementing the Sup-BM cluster against all others pooled in one class. Subsequently, validation of the Random Forest model was performed with accurate prediction on the validation set. *Rspo2* expression probe, found highly predictive for the Sup-BM cluster, was used as quantitative predictor in LIMMA algorithm (Ritchie et al., 2015) in order to find correlated gene expression profile. Conjointly, the supervised Significance Analysis for Microarray (SAM) algorithm (Tusher et al., 2001) was applied, comparing the Sup-BM vs. other clusters: the threshold was fixed for a positive value of fold change superior to 2 and a minimal false discovery rate of 0,05. To perform unsupervised classification with Euclidean distances we retained the genes in the intersection of two gene sets, one corresponding to genes whose expression was correlated to that of *Rspo2*, and the other corresponding to genes with positive fold change in Sup-BM cluster according to SAM. Microarray expression heat-plot was drawn with made4 R-package (Culhane et al., 2005). Statistical significance tested for each gene was performed with Fisher one-way ANOVA followed by post-Hoc Tukey test. The PAMr algorithm was also applied to identify the genes that were predominantly expressed in each population.

Weighted Gene Correlation Network Analysis (WGCNA): It was performed according to the algorithm developed by Horvath and co-workers (Horvath, 2011; Langfelder and Horvath, 2008), using different gene sets as indicated in the text. The chosen values for essential input parameters for each gene set were chosen as giving in each case the simplest module definition defined by clustering using the Topological Overlap Matrix (TOM) and merged dynamic tree cut method. They were: 1) the power ß obtained from the scale-free fit model curve ($r^2 > 0,8$); 2) the correlation, in all cases Pearson's; 3) the adjacency, in all cases signed; 4) the linkage, average or complete; 5) the minimum module size of 30.

For the Cytoscape connectivity plot edges were selected above a given weight (intra-modular topological overlap measure) threshold. Graph layout was always force-directed.

Gene Set Enrichment Analysis (GSEA): GSEA aims at identifying the significant genes within a gene expression set contrasting two phenotypes. The expression set is compared to pre-established datasets corresponding to divers biological process. Datasets are retained if the distribution of their genes is shifted toward either phenotype. The statistical relevance of the shift is established by random permutations of the phenotypes (Liberzon et al., 2015; Subramanian et al., 2005). In this work we used the Sup-BM gene set as reference dataset.

Gene Ontology (GO): Selected genes from the gene sets were processed using DAVID (database for annotation, visualization and integrated discovery) (Huang da et al., 2009a, b).

Multivariate information-based inductive causation (miic) analysis
We used the miic algorithm which combines constraint-based and information-theoretic frameworks to reconstruct causal and non-causal networks from large scale datasets (Verny et al., 2017; Sella et al., 2018). Miic network predictions belong to the broad class of 'ancestral graphs' that include undirected, directed as well as bi-directed edges originating from latent common causes unobserved in the available dataset and represented by dashed edges in the networks of Figure 4C and Figure 4D. In brief, starting from a fully connected graph, miic iteratively removes dispensable edges, by uncovering significant information contributions from indirect paths. This amounts to progressively uncover the best supported conditional independencies (*i.e. I(X;Y/(Ai)) = 0* implying no *XY* link in the underlying network) by iteratively 'taking off' the most significant indirect contributions of positive conditional 3-point information, $I(X;Y;A_k|_{k-1}) > 0$, from every 2-point (mutual) information, *I(X;Y)*, as,

$$I(X;Y|\{A_i\}_n) = I(X;Y) - I(X;Y;A_1) - I(X;Y;A_2|A_1) - \ldots - I(X;Y;A_n|\{A_i\}_{n-1})$$

Miic also provides an edge-specific confidence assessment of retained edges, which are oriented based on the signature of causality in observational data. This has long been known to be associated to 'v-structures', $X \rightarrow Z \leftarrow Y$, between two mutually (or conditionally) independent variables, *X* and *Y*, connected to a third variable *Z*. Indeed, this entails the orientations of the v-structure as the edges *XZ* and *YZ* cannot be undirected, nor *Z* be a cause of *X* and *Y*, since these alternative graphical models would imply correlations in contradiction with the independence between *X* and *Y*.

Single cell analysis
Single cell analyses were performed in R software environment version 3.5.3 with package Seurat_2.3.4 that uses dependencies of packages Matrix_1.2-17, cowplot_0.9.4 and ggplot2_3.1.1. Digital matrix was built with LEPR principal cluster from GSE108891 stromal data comprising 1712 cells. Data were scaled and normalized on Unique Molecular Identifiers and gene counts before performing dimensional reduction with t-SNE analysis on the first forty principal axes. Expression of selected markers was plotted on the dimension 1 vs. dimension 2 t-SNE map.

Hematopoietic cell isolation and culture
Mice: Adult C57Bl/6 female mice (10-16 weeks old) were provided from Janvier Labs and maintained in the animal facility of the Laboratory of Developmental Biology according to institutional guidelines (experiment approved by the 'Charles. Darwin' ethical committee of the Sorbonne University). Femurs and tibias were flushed with DMEM (Gibco) supplemented with 10% FCS (Eurobio).

Cell sorting: BM lineage-negative cells were first isolated by depletion of hematopoietic lineage marker (Ter119, Gr1, B220, CD4, CD8, CD11b)-expressing cells using MACS columns (Miltenyi Biotec). Cells were then stained with PE-conjugated anti-Sca-1 and APC-conjugated anti-CD117 (c-kit). LSK cells were isolated using a FACS AriaIII (BD) cell sorter.

Colony-forming units: 250 LSK sorted cells were plated in 1 mL methyl-cellulose in IMDM (Gibco) supplemented with 30% FCS (Eurobio), 1% crystallized BSA (Sigma), and 10^{-4}M ß-mercaptoethanol (Sigma), in the presence of 20 ng/mL SCF, 10 ng/mL IL-3, 2 U/mL EPO and 20 ng/mL TPO (all from PromoKine). Tested RSPO2 (R&D) concentrations and/or 100 ng/mL WNT3A (Peprotech) were added. Samples were plated in duplicate 35-mm Petri dishes and incubated for 7 days at 37°C.

Culture of LSK cells in liquid medium: 200 LSK sorted cells were plated in U-bottom 96 well-plates in 100 µL X-Vivo 15 medium (Lonza) supplemented with 10% FCS, 5.10^{-5} M ß-mercaptoethanol (Sigma) and 10^{-4} M Methyl-ß-cyclodextrin (Sigma), in the presence of 20 ng/mL SCF, 20 ng/mL RSPO2and/or 100 ng/mL WNT3A. Five wells were seeded per condition and incubated for 12 days at 37°C, cells in each well were then counted using a Mallassez hemocytometer.

qRT-PCR: RNA from 40000 LSK sorted cells or incubated for 24h in conditions described above was isolated using RNeasy Micro Plus Kit (Qiagen). Reverse transcription was performed using iScript cDNA synthesis kit (Biorad). Relative quantitative PCR was performed using the SsoAdvanced Universal SYBR Green Supermix (Biorad) with gene-specific PCR primers on a PikoReal instrument (Thermo Scientific). After one step at 9°C for 3 minutes the samples were cycled 50 times (denaturation at 95°C for 25 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 20 seconds). Cq values were measured using the PikoReal software. Expression level as $2^{-\Delta Cq}$, ΔCq = Cq(gene of interest) - Cq(housekeeping gene), housekeeping gene being *Gapdh*. The following primer sequences (5' to 3') were used:
Gapdh fw : ATGGTGAAGGTCGGTGTGAA
Gapdh rv : AATGAAGGGGTCGTTGATGG
Axin2 fw: TGACTCTCCTTCCAGATCCCA
Axin2 rv: TGCCCACACTAGGCTGACA
Bmp4 fw: TTCCTGGTAACCGAATGCTGA
Bmp4 rv: CCTGAATCTCGGCGACTTTTT
Lgr6 fw: GAGGACGGCATCATGCTGTC
Lgr6 rv: GCTCCGTGAGGTTGTTCATACT
Tcf7 fw: AGCTTTCTCCACTCTACGAACA
Tcf7 rv: AATCCAGAGAGATCGGGGGTC

Immunohistochemistry
Femurs were incubated for 5h in formol before incubation in sucrose 30% at 4°C overnight. Bones embedded in OCT were cryosectioned on Superfrost+ slides and immediately fixed in acetone for 5min at -20°C. After drying, section were washed in distilled water (dH_2O) then antigen retrieval was performed by incubating sections in boiling 10 mM citrate buffer pH6 (Dako). After cooling down for 30 min at room temperature (RT), sections were washed in PBS then in PBS-0.1%Tween and blocked for 1h at RT in PBS-5%. Sections were then incubated with primary antibody (rabbit anti-RSPO2 (1/100, NBP2-38688 Novus); goat anti-LEPR (1/40, AF497 R&D systems); biotinylated Isolectin-b4 (1/200, B-1205 Vector)) in a humidified chamber overnight at 4°C. After 3 washes in PBS, slides were incubated with the anti-goat AF555 (1/500, Thermofisher A21432) for 30 min at RT then washed 3 times before adding the anti-rabbit AF488 (1/500, Jackson 111-545-144) and the streptavidin 647 (1/1000, Thermofisher S32357) for 30 min at RT. After washing, slides were incubated for 10min with DAPI (1/5000, Sigma), washed in dH_2O and mounted in Fluoromount (Thermofisher). Images were acquired using LSM800 confocal microscope (Zeiss) and processed with Fiji (version 1.47 v; National Institutes of Health, USA, https://imagej-nih-gov.gate2.inist.fr/ij/).

Statistical analysis

For experiments comparing one condition to another (e.g. LSK cells cultured in presence of Rspo2 + SCF vs. cultures with SCF only) results were analyzed using paired t-tests and expressed as means ± SEM.

Supplemental References

Culhane, A.C., Thioulouse, J., Perriere, G., and Higgins, D.G. (2005). MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics *21*, 2789-2790.

Le, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software *25*, 1-18.