

# A Systems Biology Approach for Defining the Molecular Framework of the Hematopoietic Stem Cell Niche

Pierre Charbord,<sup>3,\*</sup> Claire Pouget,<sup>4</sup> Hans Binder,<sup>5</sup> Florent Dumont,<sup>6</sup> Grégoire Stik,<sup>1,2</sup> Pacifique Levy,<sup>1,2</sup> Fabrice Allain,<sup>1,2</sup> Céline Marchal,<sup>1,2</sup> Jenna Richter,<sup>4</sup> Benjamin Uzan,<sup>7</sup> Françoise Pflumio,<sup>7</sup> Franck Letourneur,<sup>6</sup> Henry Wirth,<sup>5</sup> Elaine Dzierzak,<sup>8</sup> David Traver,<sup>4</sup> Thierry Jaffredo,<sup>1,2,9</sup> and Charles Durand<sup>1,2,9,\*</sup>

<sup>1</sup>Sorbonne Universités, UPMC Paris 06, IBPS, UMR 7622, Laboratoire de Biologie du Développement, 75005 Paris

<sup>2</sup>CNRS, INSERM U1156, IBPS, UMR 7622, Laboratoire de Biologie du Développement, 75005 Paris, France

<sup>3</sup>INSERM U972, University Paris 11, Hôpital Paul Brousse, 94807 Villejuif, France

<sup>4</sup>Department of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92093-0380, USA

<sup>5</sup>Interdisciplinary Center for Bioinformatics, University of Leipzig, 04107 Leipzig, Germany

<sup>6</sup>Genomic Platform, Institut Cochin, INSERM U567, 75014 Paris, France

<sup>7</sup>UMR967 INSERM, LSHL/IRCM, CEA, University Paris 7, 92260 Fontenay-aux-Roses, France

<sup>8</sup>Department of Cell Biology, Erasmus Stem Cell Institute, Erasmus Medical Center, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands

<sup>9</sup>Co-senior author

\*Correspondence: [pierre.charbord@inserm.fr](mailto:pierre.charbord@inserm.fr) (P.C.), [charles.durand@upmc.fr](mailto:charles.durand@upmc.fr) (C.D.)

<http://dx.doi.org/10.1016/j.stem.2014.06.005>

## SUMMARY

Despite progress in identifying the cellular composition of hematopoietic stem/progenitor cell (HSPC) niches, little is known about the molecular requirements of HSPC support. To address this issue, we used a panel of six recognized HSPC-supportive stromal lines and less-supportive counterparts originating from embryonic and adult hematopoietic sites. Through comprehensive transcriptomic meta-analyses, we identified 481 mRNAs and 17 microRNAs organized in a modular network implicated in paracrine signaling. Further inclusion of 18 additional cell strains demonstrated that this mRNA subset was predictive of HSPC support. Our gene set contains most known HSPC regulators as well as a number of unexpected ones, such as *Pax9* and *Ccdc80*, as validated by functional studies in zebrafish embryos. In sum, our approach has identified the core molecular network required for HSPC support. These cues, along with a searchable web resource, will inform ongoing efforts to instruct HSPC ex vivo amplification and formation from pluripotent precursors.

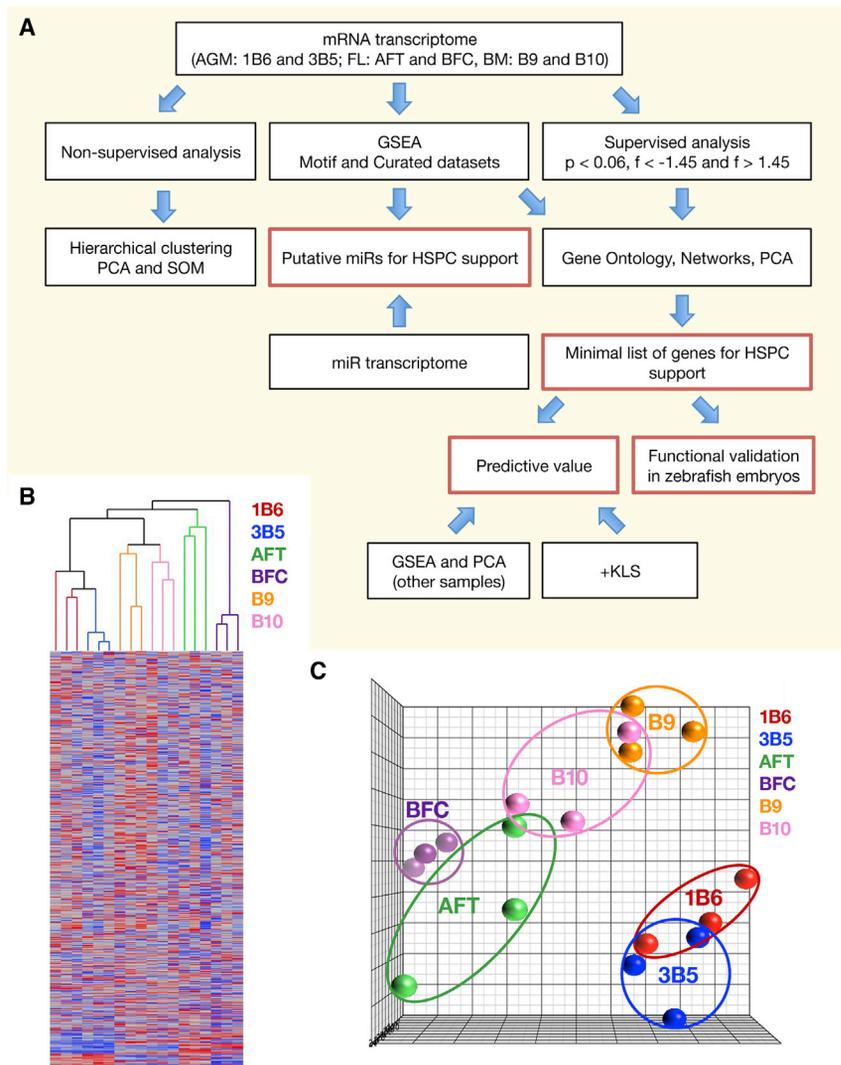
## INTRODUCTION

The molecular characterization of hematopoietic stem cell (HSC) microenvironments (also termed niches) is a fundamental goal in the field of stem cell biology and regenerative medicine (Wagers, 2012). Although signaling pathways and extracellular matrix (ECM) components critical for HSC regulation have been identified (Morrison and Spradling, 2008), the existence of a core genetic network responsible for HSC support remains elusive.

Despite significant progress in identifying cells that comprise the bone marrow (BM) niche and their molecular characterization (Morrison and Scadden, 2014), a molecular comprehensive understanding of the adult HSC niche has not yet been determined. This lack of information is largely due to the complexity of the BM microenvironment and the difficulty of obtaining and studying hematopoietic-supportive cells in living animals. Additionally, it is currently controversial as to what cell types actually comprise the murine BM niche because endosteal cells, endothelial cells, perivascular cells, and mesenchymal stem cells (MSCs) have all been implicated as key players (Calvi et al., 2003; Corselli et al., 2013; Ding et al., 2012; Greenbaum et al., 2013; Kiel et al., 2005; Kunisaki et al., 2013; Méndez-Ferrer et al., 2010; Zhang et al., 2003).

In addition to the BM niche, several additional supportive environments play key roles in HSC support during development. The aorta gonad mesonephros (AGM) region is responsible for HSC generation, whereas the fetal liver (FL) promotes HSC maturation and amplification (Dzierzak and Speck, 2008). To better characterize these different HSC niches, we utilized clonal murine stromal cell lines generated from embryonic day (E) 11 AGM, E14 FL, and adult BM (Chateauvieux et al., 2007; Moore et al., 1997; Oostendorp et al., 2002). These stromal lines immortalized with *TSV40* exhibit mesenchymal differentiation potential (Chateauvieux et al., 2007; Durand et al., 2006) and differentially support HSPCs in coculture experiments (Chateauvieux et al., 2007; Hackney et al., 2002; Moore et al., 1997; Oostendorp et al., 2002). Importantly, comparison of their gene expression profiles has been instrumental in the identification of additional HSC regulators (Durand et al., 2007; Hackney et al., 2002; Renström et al., 2009).

Here, we present in-depth analyses of the mRNA and microRNA (miR) transcriptomes expressed by these stromal lines. Using an ensemble of systems biology approaches, we have established shared molecular commonalities in HSPC niches from distinct temporal and spatial ontogenic locations. Our data were incorporated into an interactive, searchable website



**Figure 1. General Outline of the Study and Unsupervised Analysis of Transcripts**

(A) Flow chart outlining the analyses of stromal cell line transcriptomes.

(B) Heatmap and hierarchical clustering on the basic set of six cell lines (1B6, 3B5, AFT, BFC, B9, and B10, each in triplicate). Linkage average.

(C) PCA with the entire set of mRNAs as variables and the basic set of six cell lines as observations. PC1 versus PC2 score plot.

AFT024 (AFT, HSPC-supportive) and BFC012 (BFC, nonsupportive) FL lines, and BMC9 (B9, HSPC-supportive) and BMC10 (B10, less-supportive) BM lines.

### Identification of the Stromal Gene Network Essential for HSPC Support

To determine whether we could identify unique genetic signatures present in HSPC-supportive versus less-supportive stromal cell lines, we performed hierarchical clustering (Figure 1B) and principal component analysis (PCA; Figure 1C) of the entire set of mRNAs. Our findings indicated that there were no discernable differences between HSPC-supportive and less-supportive lines (Figure 1C). The first two components accounted for 29% of the variance of the results. This unsupervised analysis uncovered the presence of tissue-imprinted genes that clustered with the AGM, FL, and BM lines. To circumvent this signature, we subtracted, for each site, the gene set significantly expressed in the less-supportive line from

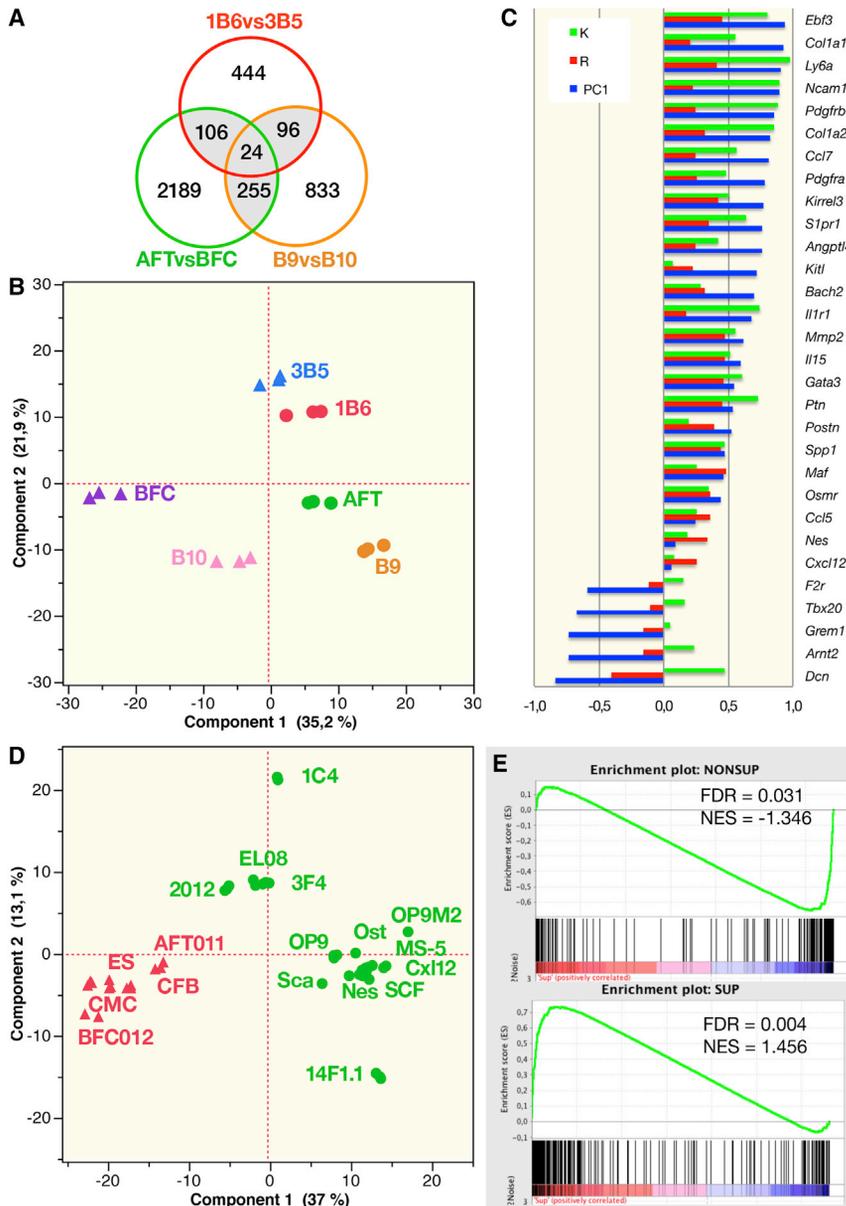
(<http://stemniche.snv.jussieu.fr/>) in order to interrogate mRNA and miR networks in the HSPC niches during development and adulthood. Altogether, information gleaned from these studies should help devise pharmaceutical drugs for the treatment of leukemias and determine methods for amplifying HSCs ex vivo and generating them from pluripotent stem cells, both of which are key issues in regenerative medicine approaches.

## RESULTS

### A Systems Biology Approach

To determine the core of genes characteristic of the stromal HSPC-supportive capacity, we developed a systems biology approach whereby high-throughput technology and bioinformatics analyses were combined (Figure 1A). We analyzed the mRNA and miR transcriptomes of stromal lines established from mouse AGM, FL, and BM. For each site, we chose two lines with differing capacity to maintain HSPCs ex vivo as revealed by repopulation assays and/or long-term cultures. The stromal cells were as follows: UG26.1B6 (1B6, HSPC-supportive) and UG26.3B5 (3B5, less-supportive) AGM lines,

the set significantly expressed in the supportive one. For this supervised analysis, gene expression levels were compared with two-way ANOVA with site and support as interacting factors. Data were filtered with  $p$  values  $< 0.06$  for transcripts regarded as statistically significant and fold changes of  $f \geq 1.45$  for upregulated genes and  $f \leq -1.45$  for downregulated genes in HSPC-supportive lines. Assuming that genes essential for HSPC support would be conserved in at least two out of the three tissues analyzed, we defined a group of 481 transcripts that we designated as set 1 (shaded area in Figure 2A, complete gene list and annotations in Table S1 available online) that corresponds to 481 unique genes either up- or downregulated in at least two HSPC-supportive cell lines. The lists of differentially expressed mRNAs between the HSPC-supportive and less-supportive stromal lines are accessible at <http://stemniche.snv.jussieu.fr/>. On the score plot (first two components) of PCA using set 1 genes as variables and 18 samples as observations (Figure 2B), the segregation between the three stromal lines that provide a potent support for HSPCs (on the right) and the three that do not or less efficiently support HSPCs (on the left) indicate that the first principal component (PC1) with the largest eigenvalue



**Figure 2. Identification of the Gene Set that Characterizes the HSPC-Supportive Capacity of Stromal Cells**

(A) Venn diagram of genes obtained by supervised analysis. Set 1 is shaded. (B) PCA with set 1 (481) genes as variables and the basic set of six cell lines as observations (18 samples). PC1 versus PC2 score plot. (C) Study of known HSPC regulators belonging to set 1. Blue bars represent correlations of gene expression to the first principal component (PC1 loadings). Red bars represent gene rank metric scores (R) given by GSEA. Green bars represent gene scaled connectivities K given by WGCNA. Value ranges: for PC1 and R [-1, 1], for K [0, 1]. (D) PCA with set 1 as variables and a set of 18 cell strains different from the basic set as observations (48 samples). PC1 versus PC2 score plot. Green circles correspond to HSPC-supportive cell strains, and red triangles correspond to non-supportive strains. (E) GSEA with set 1 as reference gene set and the same 48 samples utilized for PCA as expression data set. Each sample belongs to either one of the two phenotypes (supportive versus non-supportive). The top shows distribution of down-regulated genes in nonsupportive samples. The bottom shows distribution of upregulated genes in supportive stromal cells. NES, normalized enrichment score. See also Table S1.

corresponds to the factor support. The first two components accounted for 57% of the variance. The difference in PC1 coefficients was maximal when comparing AFT to BFC, which was expected, given that these two lines were the most distinct in terms of hematopoietic support (Hackney et al., 2002). These data indicate that we can effectively enrich for genes essential for HSPC support by removing site-specific genes identified in our original unsupervised analyses.

Several genes present in set 1 have been previously identified as critical factors involved in HSPC support, validating our experimental approach. In Figure 2C, we indicate for the 30 genes most reported in the literature (see Table S1) the value of their correlation to the PC1 loadings. Remarkably, positive regulators of HSPCs (such as *Col1a1*, *Ly6a*, *Ncam1*, *Pdgfrb*, *S1pr1*, *Spp1*, *Ptn*, *Kitl*, and *Kirrel3*) are positively correlated, whereas negative regulators (such as *Dcn* and *F2r*) are anticorre-

lated. Poor PC1 correlation value for some of the known regulators was due to significantly ( $p < 0.05$ ) divergent expression in two sites in comparison to the third (e.g., *Cxcl12* upregulated in B9 and 1B6 but downregulated in AFT).

To confirm that our subtractive strategy has provided an adequate classification of the known regulators, we utilized set 1 genes as reference gene set in gene-set enrichment analysis (GSEA). The expression data set consisted in the same 18 samples used for PCA. Each sample was labeled according to one of the two phenotypes and the genes were ranked on the basis of their level of expression. The values of the rank metric score (R) for the 30 most reported genes were correlated and anticorrelated for positive and negative regulators, respectively (Figure 2C). This result further validates the subtractive strategy with the use of a second line of analysis (GSEA) totally distinct from the first (PCA).

upon literature mining. It is apparent from the values of the connectivity  $K$  in Figure 2C that some of the known genes highly correlated to the “support” trait are highly connected to other genes of the network, defining hubs. Moreover, WGCNA allowed identification, within the network of 481 genes, of five subnetworks (modules) significantly ( $p \leq 0.006$ ) correlated to support, three positively (Figure 3A) and two negatively (data not shown). Although the three positively correlated networks differ in terms of mean connectivity (density), they all contain highly connected hubs (Figure 3B). For example, among *Tgfb1*, *Snai2*, *Wnt10b*, and *Ccdc80* that all belong to the same “blue” module, only *Ccdc80* and *Snai2* are highly connected to the other genes of the module. In contrast, the interconnected *Pax9* and *Kitl* genes are detected in the “turquoise” module in a marginal situation, indicating “fuzzy” membership characteristic of nodes intermediate to several modules. Interestingly, *Gata3* has a relatively high membership to the “black” module, supporting its recently observed role in HSPC regulation (Mirshakar-Syahkal et al., 2014). Altogether, these data identify a gene network characteristic of the HSPC supportive capacity of stromal cells using a limited number of stromal lines derived from distinct developmental sites and differing in their ability to maintain HSPCs ex vivo.

### Predictive Value of the Gene Network

Then, we investigated whether the list of set 1 genes harbors a predictive advantage of supportive ability. To test this hypothesis, we studied 18 additional cell strains. To attain this number, we generated additional transcriptomes from lines grown in our laboratory and also took advantage of the numerous data sets stored in the Gene Expression Omnibus (GEO) of the NCBI. The additional strains included stromal cell lines, AM14.1C4 and AM30.3F4 from AGM (Oostendorp et al., 2002), EL08.1D2 from E11 embryonic liver (Oostendorp et al., 2002; Ledran et al., 2008), 2012 and AFT011 from FL (Wineman et al., 1996), and 14F1.1 (Zipori et al., 1985), MS-5 (Issaad et al., 1993), OP-9 (Nakano et al., 1994), and OP9M2 (a subclone of OP9) (Magnusson et al., 2013) from BM. We also included primary stromal cells recently identified as critical components of the BM HSC niche; i.e., nestin-GFP<sup>+</sup> (Méndez-Ferrer et al., 2010), SCF-GFP<sup>+</sup> (Ding et al., 2012), Cxcl12-GFP<sup>+</sup> and Pdgfra<sup>+</sup>/Sca1<sup>+</sup> cells (Greenbaum et al., 2013), and osteoblasts sorted from Col2.3-GFP<sup>+</sup> transgenic mice (Eash et al., 2010). Finally, we included embryonic stem cells (Zhao et al., 2009), cardiomyocytes, and cardiac fibroblasts (Ieda et al., 2010) as negative control data sets. Gene expression levels were compared with three-way ANOVA with site of origin, support capacity, and experimental series as factors. All samples were appropriately normalized by removing the factor series to provide a unique  $66 \times 16,530$  matrix. Given that all stromal strains studied had been characterized in terms of HSPC support, the factor support has been used as a discriminating genetic trait in subsequent analyses. As shown in Figure 2D, the score plot (first two components) of PCA (using the 48 samples not included in our initial study as observations and the set 1 genes as variables) indicates that HSPC-supportive samples (green circles) segregate clearly from the nonsupportive ones (red triangles). Interestingly, supportive AGM and embryonic or fetal liver samples are distinct from supportive BM samples, which indicates persistent imprinting of the site of origin. PC1 alone accounted for 37% of

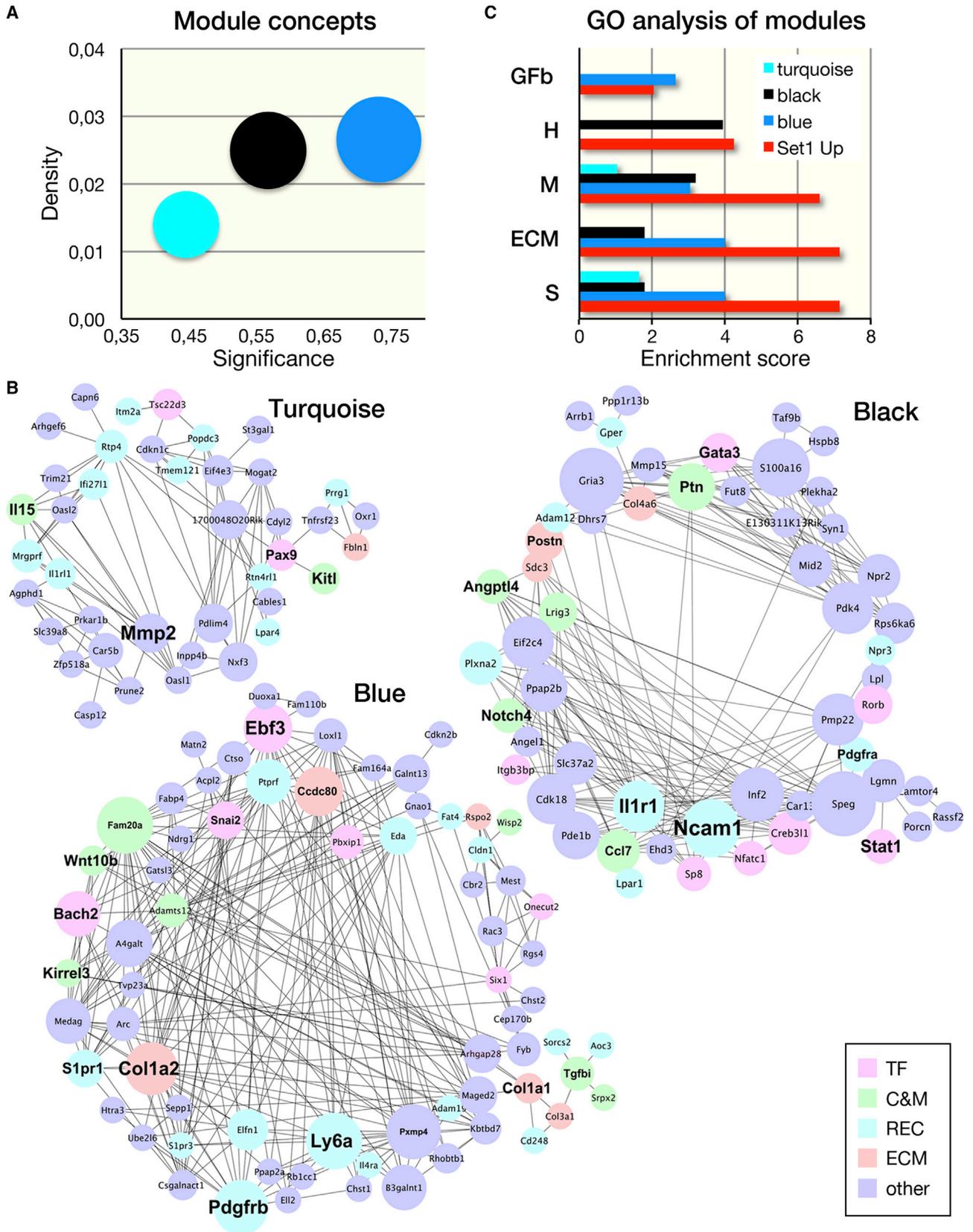
the variance, a value almost unchanged compared to that found in the PCA with our 18 initial samples (35%).

Then, we utilized set 1 genes partitioned in two gene sets of reference (up- and downregulated) in GSEA. The expression data set consisted in the same 48 samples as used for PCA, each sample being labeled according to one of the two phenotypes (supportive versus nonsupportive). Given that the transcriptomes of the samples had been carried out with different Affymetrix platforms (Mouse Gene 1.0 ST, Mouse Exon 1.0 ST, and Mouse Expression 430 array), we designed our own annotation file. As shown in Figure 2E, the enrichment score plot was shifted to the left in supportive samples, correlating well with the HSPC-supportive nature of these cells. Conversely, the enrichment score of nonsupportive cells was shifted to the right, correlating with their nonsupportive quality. The low values of the false discovery rate (FDR) indicated that the difference in profiles was highly significant ( $FDR \leq 0.03$ ). Altogether, these data strongly suggest that the list of genes in set 1 carries a strong predictive value of stromal cells to support HSPCs.

### Molecular Pathways Dedicated to HSPC Support

To identify the molecular pathways utilized by genes in the network characteristic of the stromal supportive function, we performed Gene Ontology (GO) analysis with our basic set of six lines; i.e., 18 samples. Using the filters described above, we analyzed the gene lists established for each HSPC-supportive line versus its less-supportive counterpart and used database for annotation, visualization, and integrated discovery (DAVID) to determine the GO categories with high enrichment score.

We found that 67 GO categories were differentially expressed in at least one HSPC-supportive in comparison to its corresponding less-supportive counterpart. Using these GO categories as variables and pairwise comparisons between lines as observations, we found that PC1 still corresponded to factor support (Figure 4A, score plot on the left). The first two components accounted for 58.5% of the variance. In the loading plot shown on the right in Figure 4A, variables are represented as vectors (red arrows). This indicated that many variables whose extremity lies close to the unit circle were positively correlated to the factor support (vectors in the left-to-right orientation). To find which of these variables were the most relevant, we compared for each category the means of DAVID enrichment scores in HSPC-supportive versus less-supportive cell lines. This comparison revealed that 14 categories were significantly ( $p \leq 0.05$ ) upregulated in the supportive lines (Table S2). The cellular components ECM and membrane (M); the biological processes cell adhesion (A), cell migration (Mig), and vessel development (V); and the molecular functions secreted (S), signal (Sig), heparin binding (H), growth factor binding (GFb), cytokine binding, peptidase, gpi anchor, Egf-like domain, and immunoglobulin domain were all upregulated in HSPC-supportive cell lines. We found the same categories for set 1 when comparing the genes upregulated to the downregulated ones (Figure 4A and Table S2). Remarkably, the same categories were found when analyzing the list of genes belonging to the WGCNA modules positively correlated to the factor support (Figure 3C). However, categories were not all represented in each module because of their small number gene contents. Some of the categories were predominant in a given module (e.g., GFb in the



(legend on next page)

blue module and H in the black module), reflecting their difference in the gene makeup.

Then, using Ingenuity Systems and Genomatix, we analyzed literature-based molecular networks characteristic of each supportive line in comparison to its less-supportive counterpart. Results from this analysis are shown in Figure 4B for the AGM HSPC-supportive line 1B6 and Figures S1A and S1B for the HSPC-supportive FL (AFT) and BM (B9) lines. A network consisting of genes implicated in cell communication is apparent, including molecules involved in ECM synthesis (yellow) and degradation (purple) and transcripts coding for soluble and/or transmembrane factors (gray). To validate the differential expression of candidate genes identified in our microarray analyses, we performed quantitative RT-PCR (qRT-PCR) with TaqMan Low-Density Arrays (TLDA) for 43 transcripts coding for cytokines, morphogens, receptors, cell adhesion molecules, and ECM proteins (Table S3). The regression between mRNA expression values obtained by microarray analyses and qRT-PCR was significant ( $p < 0.0001$ ), and the results for cytokines and morphogens are shown in Figure S1C. Importantly, we confirmed the higher expression level of numerous genes detected in the molecular network characteristic of HSPC support. An example is shown for the supportive AGM cell line 1B6 (Figure 4C). Altogether, these data indicate that the gene network characteristic of the HSPC supportive capacity of stromal cells includes well-conserved biological pathways implicated in cell-to-cell and cell-to-matrix communication. However, although the core network consists in identical pathways, pairwise analysis indicates that the precise gene makeup of each pathway is specific to each hematopoietic site.

### Analysis of Site-Specific Gene Signatures

To address the question of site-specific molecular pathways, we took advantage of the large panel of stromal cells studied and used GSEA and two different settings. First, the expression profiles of the HSPC-supportive AGM, FL, and BM stromal cells were independently compared to those of nonsupportive samples. Second, the expression profiles of supportive cells from one tissue were compared to supportive cells from the two other tissues. We retained the data sets with high normalized enrichment scores and low FDRs ( $\leq 0.05$ ), selected the genes belonging to the leading edge of each retained data set, and analyzed, using DAVID, the GO categories corresponding to the summing list of genes. Consistent with our previous observation, AGM, FL, and BM supportive stromal cells expressed genes involved in the well-conserved pathways described above. On top of these “canonical” pathways dedicated to HSPC support, we observed that AGM, FL, and BM stromal cells exhibited specific gene signatures corresponding to additional GO categories significantly ( $p \leq 0.05$ ) upregulated in supportive lines

of a given site (Table S4). BM-derived supportive cells harbored a mesenchymal phenotype, expressing genes specific to adipocytes (such as fatty acid or triglyceride metabolic process and regulation of lipase activity) and to osteoblasts (response to vitamin). In contrast, AGM-derived supportive cells expressed genes involved in hemostasis, vascular smooth muscle cell contraction, and nucleotide phosphodiesterase activity. FL-derived supportive cells expressed genes related to the cell cycle, such as cyclin, DNA repair, and DNA replication. PCA with 66 GO categories as variables and pairwise comparisons between cell strains as observations summarize these data (Figure S2). Notably, the orthogonal axes had to be rotated in order to disclose factors easy to interpret, canonical pathways being highly correlated to factor 1, whereas BM specific pathways were correlated to factor 2.

Thus, this study unravels three factors accounting for gene expression. The site-specific factor, apparent already at the entire transcriptome level, corresponds to tissue-imprinted genes. It is clearly disclosed by hierarchical clustering with the initial set of 18 samples (see Figure 1B). It is confirmed when using the whole set of 57 stromal samples of AGM, FL, and BM origin, as shown in Figure S3. The two other factors are discrete, requiring a subtractive strategy (core support factor) or a combination of strain comparisons (site/support-specific factor). The three factors are intricate as shown by PCA with the samples not included in our initial study as observations and set 1 genes as variables where AGM and FL samples still segregate from the BM ones (see Figure 2C).

### Analysis of Molecular Pathways after Contact with HSPCs

To evaluate whether gene pathways upregulated in HSPC-supportive lines defined a stromal cell state, we compared, for proof of concept, the most distinctive FL lines and analyzed the transcriptomes of AFT and BFC stromal cells cultured in the presence or absence of BM HSPCs. Given that these stromal lines do not exhibit contact inhibition because of the activity of the thermosensitive SV40 T antigen, we set up 4-day cocultures with BM c-Kit<sup>+</sup>Lin<sup>-</sup>Sca1<sup>+</sup> (KLS) cells in order to assess whether short time spans are sufficient for the development of hematopoietic colonies. Data shown in Figure S4 confirm that AFT and BFC stromal cells differentially support HSPCs.

AFT and BFC cells, exposed to KLS cells for 4 days or not, were sorted by flow cytometry on the basis of their lack of expression of the CD45 antigen (Figure S4). AFT and BFC cells cultured with KLS cells are referred to as AFT<sub>KLS</sub> and BFC<sub>KLS</sub>, respectively. Total RNA was extracted from CD45-negative sorted stromal cells and used for transcriptome analysis. AFT<sub>KLS</sub> cells retained the gene expressions corresponding to GO categories upregulated in supportive lines, whereas these

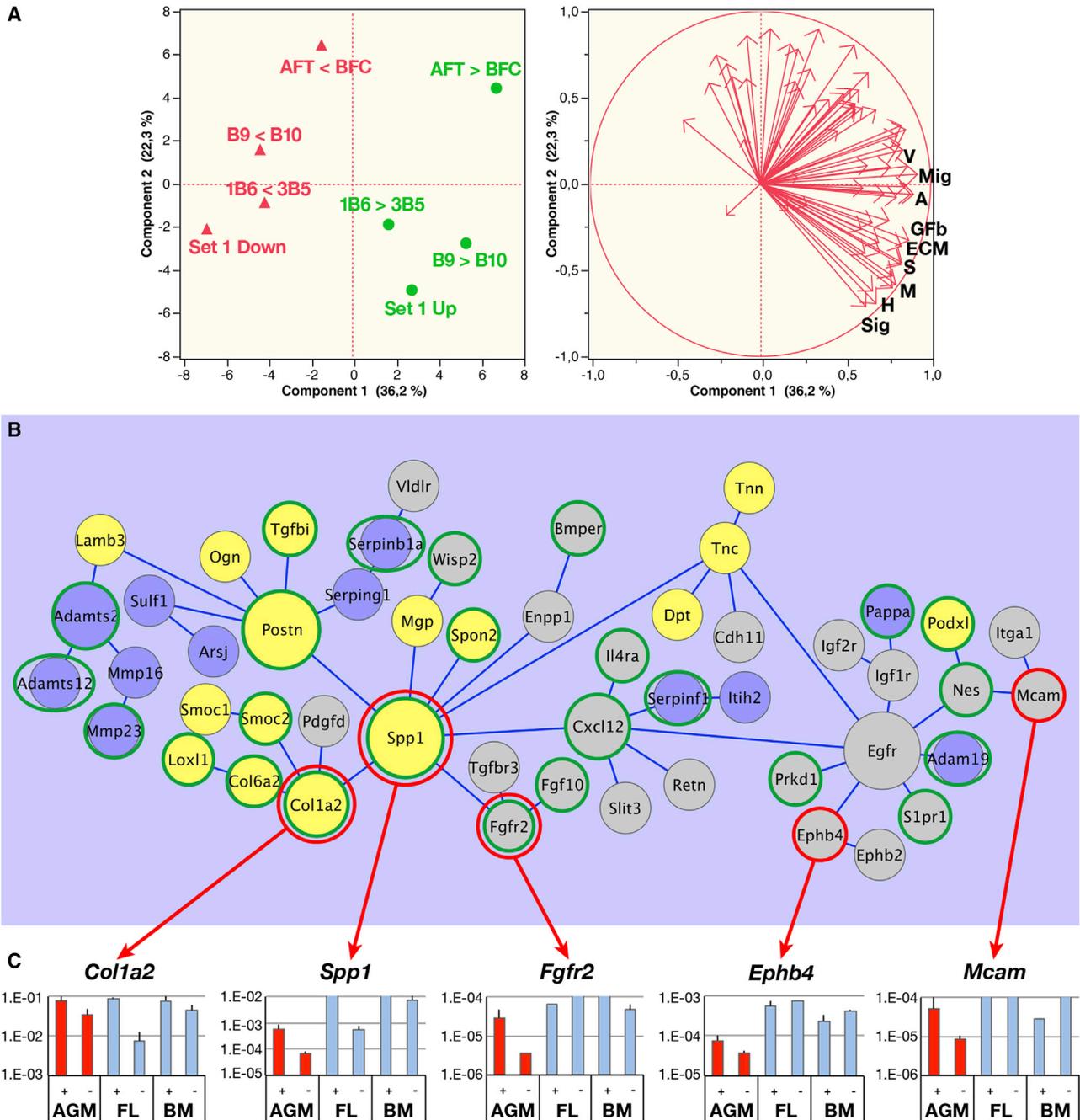
### Figure 3. Structure of the Gene Network Identifying Coordinately Expressed Genes in Set 1

(A) Major network concepts for the three subnetworks (modules) positively correlated to the factor support. Module significance is the mean of the correlations to the factor support of the genes belonging to the module. Module density approximates the mean connectivity of the genes belonging to the module. Bubble color corresponds to the arbitrary color of the module given by WGCNA.

(B) Organization of the three modules. The node size is proportional to the gene connectivity. Genes emphasized in the text or in other figures are in bold.

(C) Major GO categories given by DAVID for genes belonging to the three modules positively correlated to the factor support and, for comparison, for the upregulated genes in set 1. S, secreted; ECM, extracellular matrix; M, membrane; H, heparin binding; A, cell adhesion; GFb, growth factor binding.

See also Table S1.



**Figure 4. Molecular Pathways Representative of HSPC Supportive Activity**

(A) PCA (PC2 versus PC1) with 67 GO categories used as variables and eight pairwise comparisons of the six stromal lines (1B6 versus 3B5, AFT versus BFC, and B9 versus B10) and of the gene set 1 (genes up- versus downregulated) used as observations. The 67 categories were selected as differentially expressed in at least one supportive line in comparison to less-supportive counterpart. Left, score plot. Right, loading plot (variables as red arrows). The unit circle is shown in red. Most of the 14 categories significantly ( $p \leq 0.05$  using two-tailed Student's *t* test) upregulated in the supportive lines are indicated in bold. V, vessel development; Mig, migration. Other abbreviations are indicated in the legend of Figure 3C.

(B) Literature-based gene network upregulated in 1B6 cell line in comparison to 3B5. Gene categories: ECM synthesis (yellow nodes), ECM degradation (purple), soluble/transmembrane factors (gray). Hub genes are of larger size. Green circles indicate genes present in mRNA Set 1.

(C) qRT-PCR analysis of individual genes. Genes encircled in red in (B) were analyzed by qRT-PCR with TLDA. Values for AGM lines are shown as red bars. For comparison, values for FL and BM lines are shown as blue bars. +, HSPC-supportive line; -, less-supportive line. Mean of two independent experiments. Scale bars represent SD.

See also Figures S1 and S2 and Tables S2–S4.

GO categories did not appear in BFC<sub>KLS</sub> cells (Table S5). Moreover, a few additional GO categories were upregulated in AFT<sub>KLS</sub>, including growth factor activity, hexose metabolic process, phospholipid binding, ion transmembrane transport, and sphingolipid metabolic process. Using the whole set of GO categories as variables and pairwise comparisons as observations, PC1 still corresponded to the factor support (Figure 5A, score plot on the left). The loading plot (Figure 5A, right) revealed that the category growth factor activity was correlated to PC1 along with the other GO categories characteristic of supportive pathways. On the contrary, the other additional categories appearing in AFT<sub>KLS</sub> were correlated to PC2.

The literature-based gene network in AFT<sub>KLS</sub> cells was essentially the same as for AFT (Figure 5B), but with the inclusion of additional nodes, in particular those connected with *Vegfa* (entire list in Table S6), which is in agreement with the additional growth factor activity category (Figure 5A). Using GSEA with set 1 as a reference gene set, we found that the global transcriptomes of AFT<sub>KLS</sub> and BFC<sub>KLS</sub> correlated with the phenotype supportive versus nonsupportive, respectively (Figure 5C). These data indicate that the network of upregulated genes in the FL-supportive line exists prior contact with KLS cells but is extended after contact. Such gene pattern is not found in the nonsupportive line, even after contact, which indicates that the function support cannot be induced de novo after contact with KLS cells. Finally, the data further validate the set 1 gene list as strongly representative of the HSPC-supportive function.

### MicroRNA Expression in HSPC-Supportive Cell Lines

To determine whether HSPC-supportive cell lines also expressed specific miR signatures, we examined the relative expression of more than 600 characterized mouse miRs in the 18 samples corresponding to our initial set of stromal cell lines. Similar to what we found for mRNAs, we could not discriminate by PCA HSPC-supportive from less-supportive lines with entire gene sets (Figure 6A). The first two components accounted for 31.4% of the variance. Then, the supervised analysis was carried out with the same strategy adopted for mRNA transcriptomes but with the additional step of ignoring genes with significantly divergent expression in two cell lines in comparison to the third. Seventeen miRs (five of which were represented by two probes) were up- or downregulated in the HSPC-supportive cell lines. This list, designated as miR set 1, corresponded to approximately 3% of entire studied miR transcriptome (Figure 6B and Table S7). The lists of differentially expressed miRs between the HSPC-supportive and less-supportive stromal lines and their integration with the lists of differentially expressed mRNAs are accessible at <http://stemniche.snv.jussieu.fr/>. On the score plot (first two components) of PCA with miR set 1 genes as variables and 18 samples as observations, the segregation between the three stromal lines that provide a potent support for HSPCs (on the right) and the three that do not or less efficiently support HSPCs (on the left) indicated that PC1 corresponds to the factor support (Figure 6C, left). The first two components accounted for 65% of the variance. Analysis of PC1 loading plots showed that most miRs were positively correlated to the factor support (e.g., *miR-143*, *miR-214\**, and *miR-9\**), whereas a few were negatively correlated (e.g., *miR-155*; Figure 6C, right). Interestingly, the location of some miRs in the 2D loading plot of PC2 versus

PC1 coincided with the location of cell lines in the score plot (e.g., *miR-155* and the three 3B5 samples on Figure 6C), which gave a prediction on the relevant role of these miRs in the corresponding lines. Using qRT-PCR, we have confirmed the differential expression of *miR-9\** that was the only miR upregulated in the three supportive lines (Figure S5A).

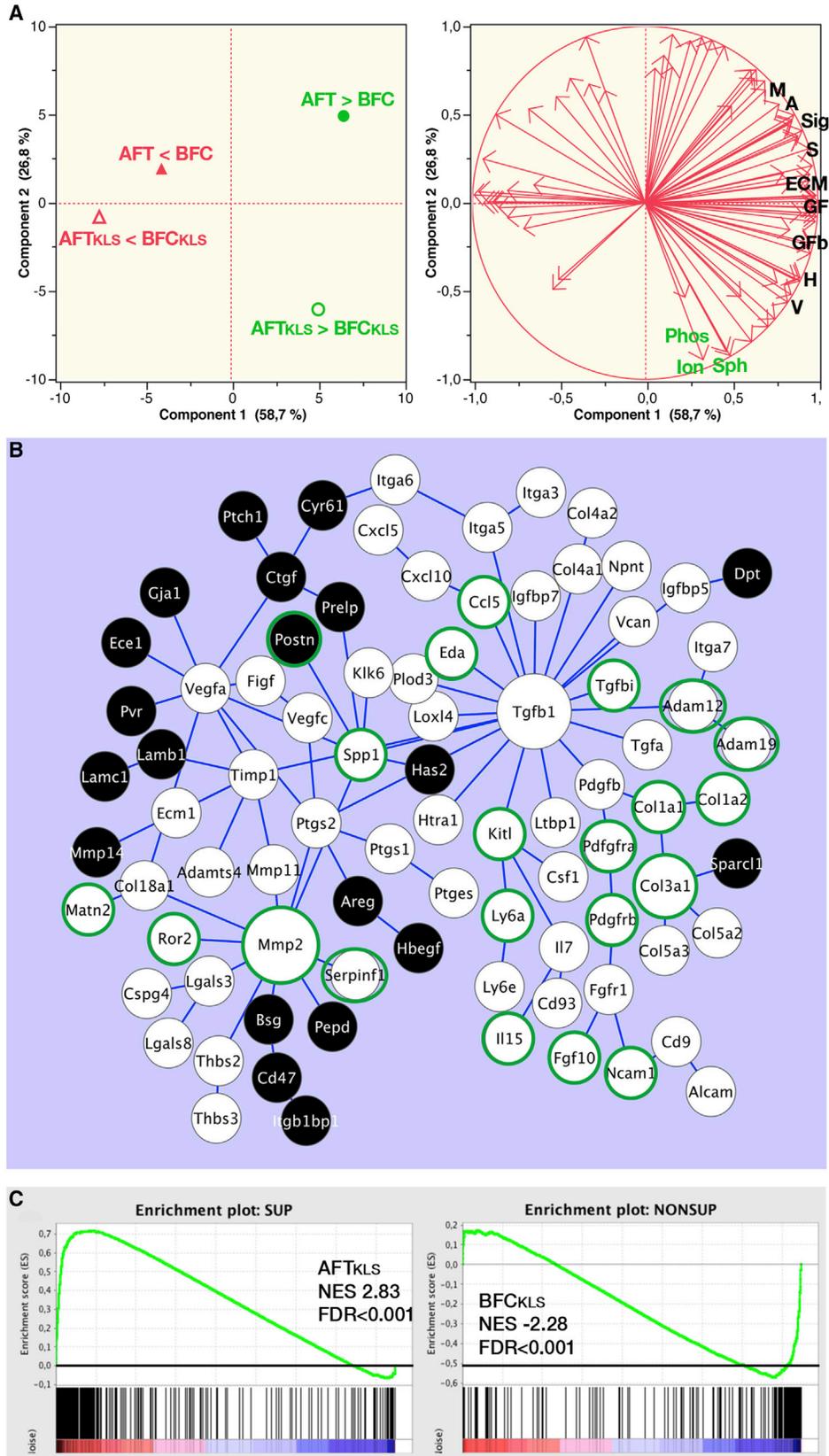
Transcripts of chemokine *Cxcl12* and cytokine receptor antagonist *Il13ra1*, significantly ( $p \leq 0.01\%$ ) upregulated in the supportive lines 1B6 and B9 (see Table S1), are validated targets of *miR-155* that negatively correlated to the factor support (PC1 loading =  $-0.42$ ). Therefore, one explanation for the lack of supportive activity by the 3B5 and B10 lines may be the upregulation of *miR-155* leading to downregulation of its targets in these lines. Another set of data concerned mRNAs upregulated in less-supportive lines. Several of these, significantly ( $p \leq 0.01\%$ ) upregulated in 3B5 or BFC, were common targets of miRs that positively correlated to the factor support. Therefore, the corresponding miRs may constitute essential nodes in gene networks activated in supportive lines. Some of their targets (*Krt19*, *Cdh1*, *Foxa2*, and *Isl1*) are markers for, or transcription factors affiliated with, endodermal differentiation.

We used GSEA to analyze motifs present in 3' untranslated regions of up- or downregulated genes, which include the binding sequence of the seed regions of miRs (Figure S5B). We hypothesized that miRs whose mRNA targets were found increased in HSPC-supportive lines would be operative in less-supportive lines. Several miRs predicted by GSEA as active in one of the less-supportive lines were found significantly ( $p < 0.06$ ) increased in the line (Figure S5B).

Altogether, these results strongly suggest that miRs may be critical cell-autonomous regulators of the stromal function affecting the expression of positive regulators in less-supportive lines or that of negative regulators in supportive ones.

### Global View of Gene Expression with Self-Organizing Map Analysis

To confirm our results with an independent method as well as to obtain more detailed insights into the transcriptional expression of HSPC-supportive cell lines, we performed self-organizing map (SOM) analysis. In essence, SOM analysis portrays the individual expression landscape of each sample in terms of mosaic images of metagenes, each representing a minicluster of coregulated genes (Wirth et al., 2011). The portraits generated for each of the replicated samples were very similar, reflecting high homogeneity of gene expression within each cell line (data not shown). Next, we generated the mean portraits of each cell line as maps where red and blue spots indicate meta genes that are over- or underexpressed in each line in comparison to expression in all samples (Figure 6D, top). Then, to determine the meta gene pattern characteristic of HSPC support, difference portraits were obtained by subtracting the mean image of supportive lines from that of less-supportive ones (Figure 6D, middle). Comparison of portraits indicated differential expression of metagenes in several regions (red dotted lines encircle upregulated metagenes, and blue dashed lines encircle downregulated metagenes). Interestingly, the difference portraits for the FL cell lines showed the most distinctive spot patterns, confirming the results of PCA (see Figure 2B). Moreover, there was strong similarity of the difference maps of FL cell lines



(legend on next page)

exposed or not to KLS, confirming that the molecular pathways are not substantially modified after contact with HSPCs. That most of the support-related metagenes formed nonoverlapping entities was made apparent when genes from set 1 were superimposed to the SOM images (Figure 6D, bottom). The subsets of differentially up- or downregulated genes of set 1 mostly accumulate in regions where metagenes are differentially expressed (encircled areas). These data indicate consistency between results from independent analyses, and illustrate that set 1 gene set covers the full spectrum of the gene expression landscape. Remarkably, most target genes of the miR set 1 accumulated in the region of metagene strong differential down-regulation, indicating their repressive effect on gene expression. Finally, we included in the SOM analysis gene sets corresponding to GO categories already identified in this study (Figure S5C). These sets mainly accumulated in the regions where metagenes were upregulated, matching well with support-related patterns.

In summary, comprehensive SOM analysis identifies and disentangles regulatory modes of gene expression due to HSPC support with a simple-to-interpret visualization frame.

### In Vivo Validation of Developmental HSPC Regulators in the Zebrafish Embryo

To functionally validate our computational analysis, we performed in vivo knockdown (KD) of gene function using antisense morpholinos (MOs) in zebrafish embryos. The zebrafish is an excellent system for rapidly and robustly testing candidate gene function in this manner (McKinney-Freeman et al., 2012; Tijssen et al., 2011). We selected *Tgfb1*, *Snai2*, *Wnt10b*, *Pax9*, and *Ccdc80* genes for KD experiments on the basis of their relative expression levels in our transcriptome analyses (Table S1). *Tgfb1* plays an important role in cell-collagen interactions that affect mesenchymal differentiation; *Tgfb1* was significantly ( $p < 0.004$ ) upregulated in the three HSPC-supportive cell lines with PC1 loading of 0.88. The zinc-finger protein *Snai2* is a transcription factor involved in epithelial to mesenchyme transition; *Snai2* was significantly ( $p \leq 0.03$ ) upregulated in 1B6 and B9 cell lines with PC1 loading of 0.74. *Wnt10b* plays a role in the homeostasis of stem cells, including HSCs and MSCs; *Wnt10b* was significantly ( $p < 0.004$ ) upregulated in AFT and B9 lines with PC1 loading of 0.87. *Pax9* and *Ccdc80* are significantly upregulated in AFT and B9 with PC1 loadings of 0.72 and 0.89, respectively. *Pax9* encodes a paired box transcription factor involved in the development of the thymus, parathyroid glands, teeth, and skeletal elements of skull and larynx. *Ccdc80* encodes an extracellular protein implicated in cell adhesion and matrix assembly.

Embryos were injected at the one-cell-stage with MOs directed against zebrafish homologs of these genes and subsequently analyzed by whole-mount in situ hybridization (WISH)

for the HSPC marker *cmyb* and the T cell marker *rag1*. *Tgfb1*, *snai2*, *pax9*, and *ccdc80*, but not *wnt10b*, morphants exhibited strong hematopoietic defects in the 36 hr postfertilization dorsal aorta at a time when HSPCs emerge (Bertrand et al., 2010; Kissa and Herbomel, 2010) (Figure 7A). Importantly, the *fli1*:eGFP<sup>+</sup> vasculature was intact, indicating that the reduction in HSPCs in *tgfb1*, *snai2*, *pax9*, and *ccdc80* morphants was not resulting from vascular abnormalities (Figure S6A). To analyze HSCs, we utilized *cmyb*:eGFP; *kdrl*:memCherry animals, whereby double-positive cells in the dorsal aorta are definitive HSCs (Bertrand et al., 2010). Using confocal microscopy, we were able to precisely quantify the numbers of HSCs in the dorsal aorta of KD animals (Figure S6B). Figure 7B shows significant ( $p < 0.05$ ) reduction of *cmyb*<sup>+</sup>/*kdrl*<sup>+</sup> cells in *tgfb1*<sup>-</sup>, *snai2*<sup>-</sup>, *pax9*<sup>-</sup>, and *ccdc80*-deficient embryos in comparison to controls.

To examine later stages of definitive hematopoiesis, we examined HSPCs in the caudal hematopoietic tissue (CHT, equivalent of mammalian FL) and intrathymic T cell development. HSPC numbers were severely impaired in *tgfb1* and *snai2* morphants, as demonstrated by respective decreases in *cmyb* and *rag1* expression (Figure 7C, top, rows 1–3). A reduction in *cmyb* and *rag1* expression was also observed in *pax9* and *ccdc80* morphants (Figure 7C, bottom, rows 1–3). In the developing kidney, which is the future site of adult hematopoiesis, *snai2*, *pax9*, and *ccdc80* morphants showed a decrease in *cmyb* expression, whereas no *cmyb*<sup>+</sup> cells were detected in *tgfb1* and *wnt10b* morphants (Figure 7C).

In addition, to examine whether our gene set is also predictive for negative HSPC regulators, we selected *Grem1* encoding a BMP antagonist, because its expression is negatively correlated to the factor support with PC1 loading of  $-0.74$  (Figure 2C and Table S1). Gain-of-function for *grem1* reduced the expression of *cmyb* in the dorsal aorta in comparison to control embryos (Figure S6C). This observation is in agreement with findings in the mouse showing an inhibition of HSPC activity and a reduction in the numbers of highly enriched HSCs when AGM explants are cultured in the presence of recombinant gremlin proteins (Boisset et al., 2013; Durand et al., 2007).

These data confirm that our systems biology approach for identifying critical regulators of HSPC specification and survival can be functionally assessed with the zebrafish model system. Importantly, these results identify *tgfb1*, *snai2*, *pax9*, and *ccdc80* as regulators of HSC specification and confirm the role of *wnt10b* for HSC maintenance.

## DISCUSSION

The primary objective of the present work has been to provide a list of core genes operative in several sites of hematopoiesis and

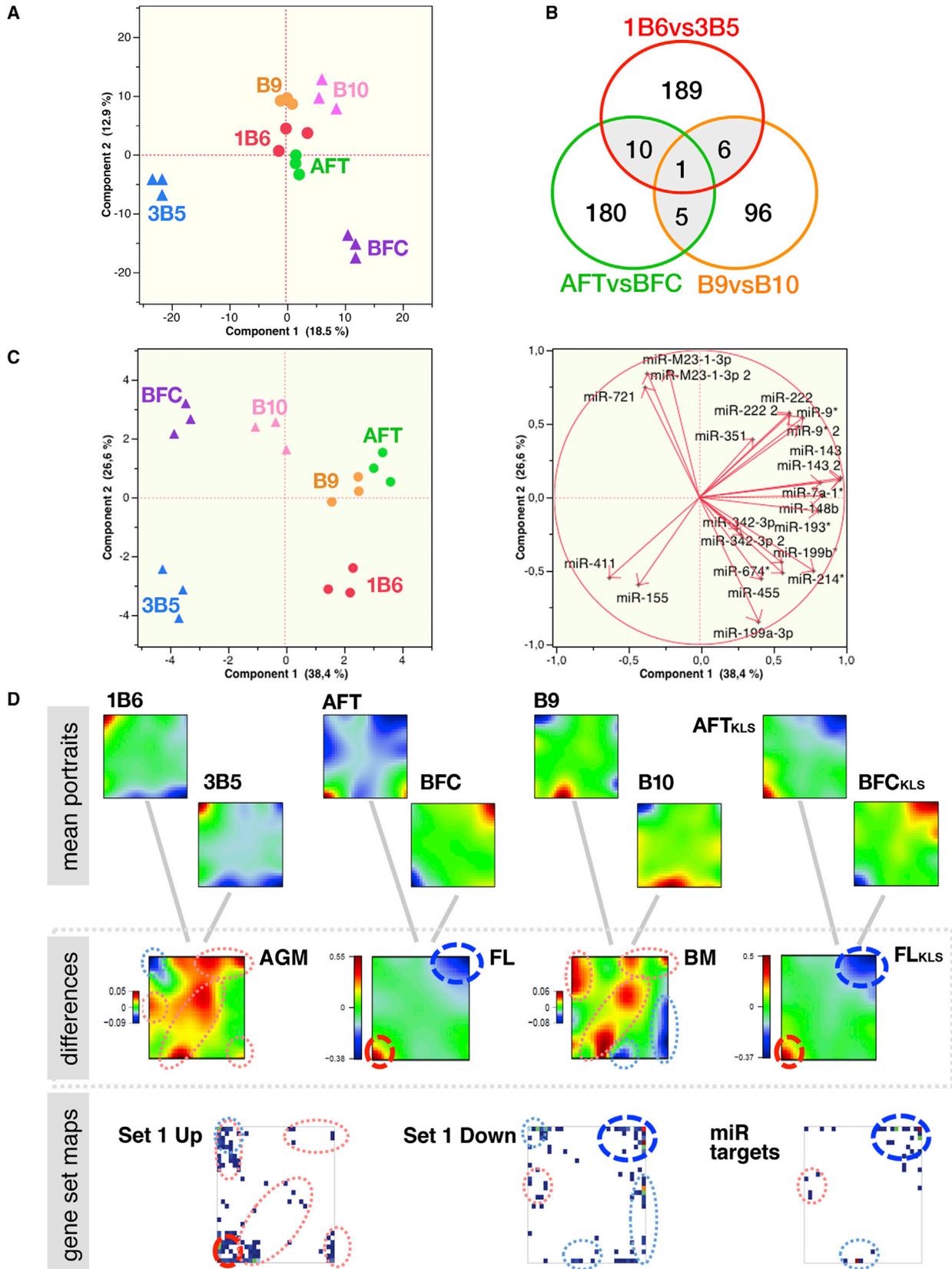
### Figure 5. Modulation of the Transcriptome of FL Stromal Lines after Contact with HSPCs

(A) PCA with 82 GO categories used as variables and four pairwise comparisons (AFT > BFC, AFT<sub>KLS</sub> > BFC<sub>KLS</sub>, AFT < BFC, and AFT<sub>KLS</sub> < BFC<sub>KLS</sub>) used as observations. The 82 categories were selected as differentially expressed in at least one of the AFT lines ( $\pm$ KLS) in comparison to BFC ( $\pm$ KLS). Left, score plot. Right, loading plot (variables as red arrows). Some of the categories upregulated in AFT  $\pm$  BFC are indicated in bold. GF, growth factor activity; Ion, ion transmembrane transport; Phos, phospholipid binding; Sph, Sphingolipid metabolic process.

(B) Literature-based gene network upregulated in AFT<sub>KLS</sub> versus BFC<sub>KLS</sub>. White nodes, genes already upregulated in AFT versus BFC; black nodes, additional genes upregulated in AFT<sub>KLS</sub> versus BFC<sub>KLS</sub>. Green circles indicate genes present in mRNA Set 1.

(C) GSEA with set 1 as reference data set and AFT<sub>KLS</sub> and BFC<sub>KLS</sub> as expression data set.

See also Figure S4 and Tables S5 and S6.



(legend on next page)

indicate how these genes relate one to another so that they can collectively implement the stromal function of HSPC support. To achieve this goal, we have used a number of analytical tools. The final output of these analyses was a gene set of 481 mRNAs (corresponding to 2%–3% of the global transcriptome). This discrete gene set included most of the genes previously identified as positive or negative HSPC regulators. The resulting network included three modules positively correlated to the trait support, with many known regulators presenting as hubs or fuzzy genes. On the basis of the present literature, the gene set specified conserved biological processes and molecular functions implicated in cell communication, ECM remodeling and vessel development. However, additional pathways are likely at play, given that the former did not account for the upregulation in the supportive lines of many intracellular molecules (e.g., *Arnt2* and *Bach2*, see [Table S1](#)). Such molecules might also be implicated in the supportive capacity being exported in microvesicles or secreted and secondarily internalized. Altogether, the stromal gene set 1 essential for HSPC support can be envisioned as a network of connected genes, a set of molecular pathways or a landscape of metagenes.

Such results had to be statistically validated with samples different from the initial set. This has been performed by verifying that set 1 had a predictive value, allowing correct categorization of stromal cells with already characterized supportive capacity and of cells originating from nonhematopoietic sites. Classifications according to PCA and GSEA were completely concordant, allowing clear-cut discrimination of HSPC-supportive from non-supportive cells.

Biological validation of our *in silico* results has been two-pronged. First, we investigated how contact with HSPCs would modulate the gene set. The pairwise comparison of the FL lines before and after exposure to KLS cells indicated that the genes upregulated in the AFT-supportive line before contact were still upregulated after contact, although they were still not induced in the nonsupportive BFC line. Moreover, in the supportive cell line exposed to HSPCs, many genes were specifically induced, the most relevant being those implicated in angiogenesis and cytokine activity. These results confirm previous studies indicating that MSCs and vascular cells play a critical role in the regulation of HSCs ([Ding et al., 2012](#); [Greenbaum et al., 2013](#); [Kiel et al., 2005](#); [Kunisaki et al., 2013](#); [Méndez-Ferrer et al., 2010](#)). Our second approach for biological validation has been loss of function of some of the genes of the gene set. Due to the conservation of major hematopoietic regulations throughout vertebrates, zebrafish is a useful model to study the role of molecules involved in hematopoiesis ([McKinney-Freeman et al., 2012](#); [Tijssen et al., 2011](#)). We selected five genes positively

correlated to the trait support. *Tgfb1*, *snai2*, *pax9*, and *ccdc80*, but not *wnt10b*, morphants displayed a prominent phenotype in the dorsal aorta. The critical implication of *tgfb1* and *ccdc80* known to regulate cell adhesion and matrix assembly suggests an important role of these processes in HSPC specification. The presence of binding sites for the transcription factor Pax9 both in the human and mouse *Kitl* promoter might account for the high correlation between the expressions of these two genes. This result suggests that Pax9 may affect hematopoiesis through direct regulation of cytokine gene expression. *Snai2* is also an interesting candidate because it is known to play a major role in epithelial-to-mesenchyme transition, generation and migration of neural crest cells, and mesenchymal stem cell differentiation. Collectively, our data indicate that transcription factors, as well as ECM and signaling molecules, might cooperate as niche factors to control definitive hematopoiesis in the AGM.

Our study was primarily designed to identify the molecular core of the HSPC support. However, specific signatures preferentially active in the AGM, FL, or BM are expected given that previous studies have reported that mesenchymal cells from different anatomical sites express distinct transcriptional program ([Chang et al., 2002](#)). Moreover, such signatures may also directly reflect the differences between AGM, FL, and BM hematopoietic micro-environments, given that the AGM is the site where HSPCs emerge in the embryo, the FL the site where they are amplified, and the BM the tissue where HSPCs are maintained throughout life. Indeed, our analysis revealed that BM-supportive cells exhibited a mesenchymal phenotype, whereas AGM-supportive cells more specifically expressed genes implicated in blood vessel functions. Interestingly, FL-supportive cells expressed genes related to the cell cycle, suggesting that not only HSCs but also their niches may actively proliferate in the FL.

The expression of the core and associated networks in the supportive lines results from underlying regulatory structures such as chromatin organization ([de Wit et al., 2013](#)) or networks of non-coding RNAs ([Djebali et al., 2012](#)). Therefore, we finalized this work by investigating miRs differentially expressed in the stromal cell lines. The subtractive strategy yielded a short list of 17 miRs that segregated HSPC-supportive from less-supportive stromal lines. Comparison of miR expression in a given HSPC-supportive line and expression of its validated mRNA targets in its less-supportive counterpart, and the converse, suggested that miRs may exert their suppressive effect not only by decreasing the expression of positive hematopoietic regulators but also by inducing adverse priming to nonmesodermal lineages.

In conclusion, our work not only validates previous findings on critical hematopoietic regulators but also provides insights into unexpected ones such as *Tgfb1*, *Pax9*, and *Ccdc80*. Among

#### Figure 6. MicroRNA Transcriptome and SOM Analyses

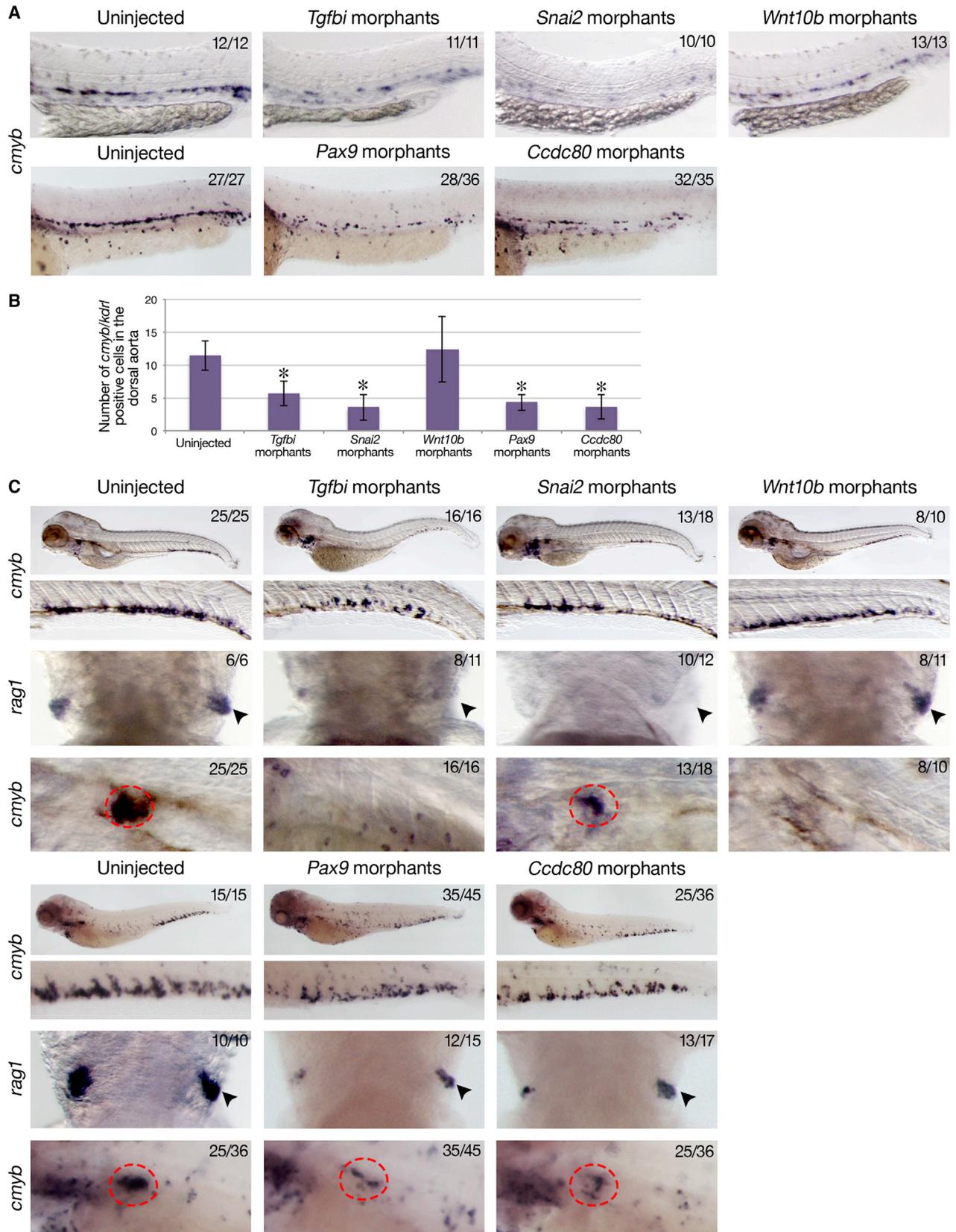
(A) PCA with entire expression data set as variables and basic set of six cell lines as observations. PC2 versus PC1 score plot.

(B) Venn diagram of genes obtained by supervised analysis. miR set 1 is shaded.

(C) PCA (PC2 versus PC1) with miR set 1 as variables and basic set of cell lines as observations. Left, score plot. Right, loading plot.

(D) SOM. Top, mean portraits of the different cell lines obtained by averaging the respective individual portraits. Middle, difference portraits obtained by subtracting mean portraits of less-supportive cell lines from that of their HSPC-supportive counterparts. Red and blue dotted circles indicate regions containing differentially up- and downregulated genes, respectively. Difference in scale for FL and FL + KLS (FLKLS) lines in comparison to AGM, and BM lines indicates larger differences in the FL cell lines (areas encircled by dashed circles) than in AGM and BM cell lines (areas encircled by dotted circles). Bottom, projection of mRNA set 1 genes and of target genes of miR Set 1 onto SOM images (each dot refers to at least one individual gene). Dotted circles indicate highly populated regions.

See also [Figure S5](#) and [Table S7](#).



(legend on next page)

the list of 481 genes, many have unknown functions as HSPC niche factors. To our knowledge, these data represent a comprehensive list of genes involved in the support of HSPCs by stromal cell lines, therefore offering a unique resource for research on the stromal regulation of hematopoiesis. From a clinical standpoint, some of the genes unraveled in this work might prove invaluable for better definition of cell therapy protocols (ex vivo amplification of HSCs and generation of de novo HSCs from pluripotent stem cells) and as optimal targets for pharmaceutical drugs for the treatment of hematologic diseases and primarily leukemias.

## EXPERIMENTAL PROCEDURES

### Microarray Screening and Data Analysis

In this study, we used Affymetrix Mouse Gene 1.0 ST arrays and Agilent miRNA 8X15K arrays. Total RNA was extracted from confluent stromal cultures or sorted stromal cells cultured with or without KLS cells with Trizol reagent (Invitrogen). RNA concentration and integrity were evaluated with the Agilent Bioanalyzer 2100 (Genomic Facility Platform, Cochin Institute). To correct from probe set definition inaccuracy, we used the version 17 of the custom ChIP definition file (Dai et al., 2005). This file eliminated probes with multiple matching sequences. Unsupervised analyses were performed on the global transcriptome, and data were represented with hierarchical clustering, PCA, and SOMs. The global expression profile of the stromal lines was also analyzed with GSEA with the functional (curated) and motif data sets present in the Molecular Signature Database of the Broad Institute ([www.broadinstitute.org/gsea](http://www.broadinstitute.org/gsea)) (Subramanian et al., 2005). In parallel, we performed supervised analyses to identify genes specifically up- or down-regulated in HSPC-supportive stromal cells. Then, the lists of differentially expressed genes (from the supervised and the GSEA analyses) were analyzed for GO with DAVID (<http://david.abcc.ncifcrf.gov>) and to uncover literature-based molecular pathways with Ingenuity Systems ([www.ingenuity.com](http://www.ingenuity.com)) and Genomatix ([www.genomatix.de](http://www.genomatix.de)) databases and conversion with Cytoscape software (<http://www.cytoscape.org>). The expression data were also analyzed with SOM as described previously (Wirth et al., 2011). To find out the network structure of the gene set 1, we used WGCNA (Horvath, 2011). Adjacencies were given according to signed Pearson correlation. The soft threshold power  $\beta$  that resulted in approximate scale-free topology was 36. Modules were constructed with average linkage hierarchical clustering. Minimum module size was 30. For better readability, the connectivity threshold was 0.1. Additional details are given in the [Supplemental Experimental Procedures](#).

### Hematopoietic Assays and Purification of Stromal Cells after Coculture

BM was obtained from adult C57BL/6 female mice (3–10 months of age). Mice were bred at Janvier (Le Genest) and maintained in the animal facility of the Laboratory of Developmental Biology (University Pierre and Marie Curie, UMR7622, CNRS) according to institutional guidelines. Adult BM hematopoietic cells enriched in stem cell activity were purified by fluorescence-activated cell sorting on the basis of the KLS cell-surface phenotype. Details are provided in the [Supplemental Experimental Procedures](#).

### Zebrafish Studies

Zebrafish were maintained according to University of California, San Diego, International Animal Care and Use Committee guidelines. Embryos were collected, staged, fixed, and processed for in situ hybridization as previously described (Thisse et al., 1993). Wild-type or transgenic zygotes were injected with a morpholino solution and incubated at 28.5°C until they reached the stage of interest. Tg(*cmyb*:eGFP) animals (North et al., 2007) were crossed to Tg(*kdr1*:HsHRAS-mCherry)*s896* animals (referred to as *kdr1*:memCherry for clarity) (Chi et al., 2008). Additional details of protocols and analyses are given in the [Supplemental Experimental Procedures](#).

### ACCESSION NUMBERS

The NCBI GEO accession number for the microarray data presented in this paper is GSE44181.

### SUPPLEMENTAL INFORMATION

Supplemental Information contains Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.stem.2014.06.005>.

### AUTHOR CONTRIBUTIONS

P.C., T.J., and C.D. designed research, performed experiments, analyzed data, and wrote the paper. C.P. designed research, performed experiments, and analyzed data. D.T. designed research, analyzed data and wrote the paper. G.S. and P.L. performed experiments and analyzed data. H.B., F.D., F.A., F.L., and H.W. analyzed data. C.M., J.R., and B.U. performed experiments. F.P. and E.D. provided materials.

### ACKNOWLEDGMENTS

We are very grateful to Catherine Robin (Hubrecht Institute) for critical and constructive comments on this study. We thank David Stachura (Department of Cell and Developmental Biology, UCSD) for critical reading of the manuscript. We thank Nicole Boggetto (Institut Jacques Monod, ImagoSeine Bioimaging Core Facility) for cell sorting experiments. We also thank Sophie Goumet (UMR CNRS 7622) for her excellent photographic and drawing assistance. These studies were supported by grants from the Fondation pour la Recherche Médicale (DEQ20100318258) and Agence Nationale pour la Recherche/California Institute for Regenerative Medicine (ANR/CIRM 0001-02).

Received: April 4, 2013

Revised: April 4, 2014

Accepted: June 6, 2014

Published: July 17, 2014

### REFERENCES

Bertrand, J.Y., Cisson, J.L., Stachura, D.L., and Traver, D. (2010). Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. *Blood* 115, 2777–2783.

### Figure 7. Functional Validation of Developmental HSPC Regulators in Zebrafish Embryos

(A) WISH for *cmyb* expression in 36 hr postfertilization morphants embryos and their corresponding uninjected controls. Numbers of embryos with or without (out of total) *cmyb* staining are indicated in the top right corner.

(B) Numbers of *cmyb*<sup>+/kdr1</sup> cells in ventral part of dorsal aorta in double *cmyb*:eGFP; *kdr1*:memCherry transgenic embryos. Data represent mean  $\pm$  SD (n = 6–19 embryos). \*p < 0.05 with two-tailed Student's t test.

(C) WISH for *cmyb* and *rag1* expression in uninjected embryos or embryos injected with morpholinos directed against *tgfb1*, *snai2*, and *wnt10b* (top) or directed against *pax9* and *ccdc80* (bottom). Row 1, *cmyb* expression in CHT in 4 days postfertilization (dpf) embryos. Row 2, magnification of CHT region. Row 3, *rag1* expression in 4 dpf embryos. Arrowheads indicate bilateral thymii. Row 4, *cmyb* expression in 4 dpf embryos. Dotted circles indicate pronephros. Numbers of embryos with or without *cmyb* or *rag1* staining are indicated in the top right corner.

See also [Figure S6](#) and [Table S8](#).

- Boisset, J.C., Clapes, T., Van Der Linden, R., Dzierzak, E., and Robin, C. (2013). Integrin  $\alpha$ 1b (CD41) plays a role in the maintenance of hematopoietic stem cell activity in the mouse embryonic aorta. *Biol. Open* 2, 525–532.
- Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M., Olson, D.P., Knight, M.C., Martin, R.P., Schipani, E., Divieti, P., Bringham, F.R., et al. (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425, 841–846.
- Chang, H.Y., Chi, J.T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., and Brown, P.O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl. Acad. Sci. USA* 99, 12877–12882.
- Chateauvieux, S., Ichanté, J.L., Delorme, B., Frouin, V., Piétu, G., Langonné, A., Gallay, N., Sensebé, L., Martin, M.T., Moore, K.A., and Charbord, P. (2007). Molecular profile of mouse stromal mesenchymal stem cells. *Physiol. Genomics* 29, 128–138.
- Chi, N.C., Shaw, R.M., De Val, S., Kang, G., Jan, L.Y., Black, B.L., and Stainier, D.Y. (2008). Foxn4 directly regulates *tbx2b* expression and atrioventricular canal formation. *Genes Dev.* 22, 734–739.
- Corselli, M., Chin, C.J., Parekh, C., Sahaghian, A., Wang, W., Ge, S., Evseenko, D., Wang, X., Montelatici, E., Lazzari, L., et al. (2013). Perivascular support of human hematopoietic stem/progenitor cells. *Blood* 121, 2891–2901.
- Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33, e175.
- de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231.
- Ding, L., Saunders, T.L., Enikolopov, G., and Morrison, S.J. (2012). Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* 481, 457–462.
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, P.H., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108.
- Durand, C., Robin, C., and Dzierzak, E. (2006). Mesenchymal lineage potentials of aorta-gonad-mesonephros stromal clones. *Haematologica* 91, 1172–1179.
- Durand, C., Robin, C., Bollerot, K., Baron, M.H., Ottersbach, K., and Dzierzak, E. (2007). Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells. *Proc. Natl. Acad. Sci. USA* 104, 20838–20843.
- Dzierzak, E., and Speck, N.A. (2008). Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat. Immunol.* 9, 129–136.
- 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.
- Greenbaum, A., Hsu, Y.M., Day, R.B., Schuettelpelz, 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.
- Hackney, J.A., Charbord, P., Brunk, B.P., Stoeckert, C.J., Lemischka, I.R., and Moore, K.A. (2002). A molecular profile of a hematopoietic stem cell niche. *Proc. Natl. Acad. Sci. USA* 99, 13061–13066.
- Horvath, S. (2011). *Weighted Network Analysis. Applications in Genomics and systems biology.* (New York: Springer).
- Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142, 375–386.
- Issaad, C., Croisille, L., Katz, A., Vainchenker, W., and Coulombel, L. (1993). A murine stromal cell line allows the proliferation of very primitive human CD34+/CD38- progenitor cells in long-term cultures and semisolid assays. *Blood* 81, 2916–2924.
- Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* 121, 1109–1121.
- Kissa, K., and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* 464, 112–115.
- 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.
- Ledran, M.H., Krassowska, A., Armstrong, L., Dimmick, I., Renström, J., Lang, R., Yung, S., Santibanez-Coref, M., Dzierzak, E., Stojkovic, M., et al. (2008). Efficient hematopoietic differentiation of human embryonic stem cells on stromal cells derived from hematopoietic niches. *Cell Stem Cell* 3, 85–98.
- Magnusson, M., Sierra, M.I., Sasidharan, R., Prashad, S.L., Romero, M., Saarikoski, P., Van Handel, B., Huang, A., Li, X., and Mikkola, H.K. (2013). Expansion on stromal cells preserves the undifferentiated state of human hematopoietic stem cells despite compromised reconstitution ability. *PLoS ONE* 8, e53912.
- McKinney-Freeman, S., Cahan, P., Li, H., Lacadie, S.A., Huang, H.T., Curran, M., Loewer, S., Naveiras, O., Kathrein, K.L., Konantz, M., et al. (2012). The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell* 11, 701–714.
- Méndez-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.
- Mirshakar-Syahkal, B., Fitch, S.R., and Ottersbach, K. (2014). From greenhouse to garden: The changing soil of the hematopoietic stem cell microenvironment during development. *Stem Cells* 32, 1691–700.
- Moore, K.A., Ema, H., and Lemischka, I.R. (1997). In vitro maintenance of highly purified, transplantable hematopoietic stem cells. *Blood* 89, 4337–4347.
- Morrison, S.J., and Scadden, D.T. (2014). The bone marrow niche for haematopoietic stem cells. *Nature* 505, 327–334.
- Morrison, S.J., and Spradling, A.C. (2008). Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* 132, 598–611.
- Nakano, T., Kodama, H., and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* 265, 1098–1101.
- North, T.E., Goessling, W., Walkley, C.R., Lengerke, C., Kopani, K.R., Lord, A.M., Weber, G.J., Bowman, T.V., Jang, I.H., Grosser, T., et al. (2007). Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* 447, 1007–1011.
- Oostendorp, R.A., Harvey, K.N., Kusadasi, N., de Brijn, M.F., Saris, C., Ploemacher, R.E., Medvinsky, A.L., and Dzierzak, E.A. (2002). Stromal cell lines from mouse aorta-gonads-mesonephros subregions are potent supporters of hematopoietic stem cell activity. *Blood* 99, 1183–1189.
- Renström, J., Istvanffy, R., Gauthier, K., Shimono, A., Mages, J., Jardón-Alvarez, A., Kröger, 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.
- 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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Thisse, C., Thisse, B., Schilling, T.F., and Postlethwait, J.H. (1993). Structure of the zebrafish *snail1* gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* 119, 1203–1215.
- Tijssen, M.R., Cvejic, A., Joshi, A., Hannah, R.L., Ferreira, R., Forrai, A., Bellissimo, D.C., Oram, S.H., Smethurst, P.A., Wilson, N.K., et al. (2011). Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL

binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell* 20, 597–609.

Wagers, A.J. (2012). The stem cell niche in regenerative medicine. *Cell Stem Cell* 10, 362–369.

Wineman, J., Moore, K., Lemischka, I., and Müller-Sieburg, C. (1996). Functional heterogeneity of the hematopoietic microenvironment: rare stromal elements maintain long-term repopulating stem cells. *Blood* 87, 4082–4090.

Wirth, H., Löffler, M., von Bergen, M., and Binder, H. (2011). Expression cartography of human tissues using self organizing maps. *BMC Bioinformatics* 12, 306.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.G., Ross, J., Haug, J., Johnson, T., Feng, J.Q., et al. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* 425, 836–841.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L., et al. (2009). iPS cells produce viable mice through tetraploid complementation. *Nature* 461, 86–90.

Zipori, D., Duksin, D., Tamir, M., Argaman, A., Toledo, J., and Malik, Z. (1985). Cultured mouse marrow stromal cell lines. II. Distinct subtypes differing in morphology, collagen types, myelopoietic factors, and leukemic cell growth modulating activities. *J. Cell. Physiol.* 122, 81–90.