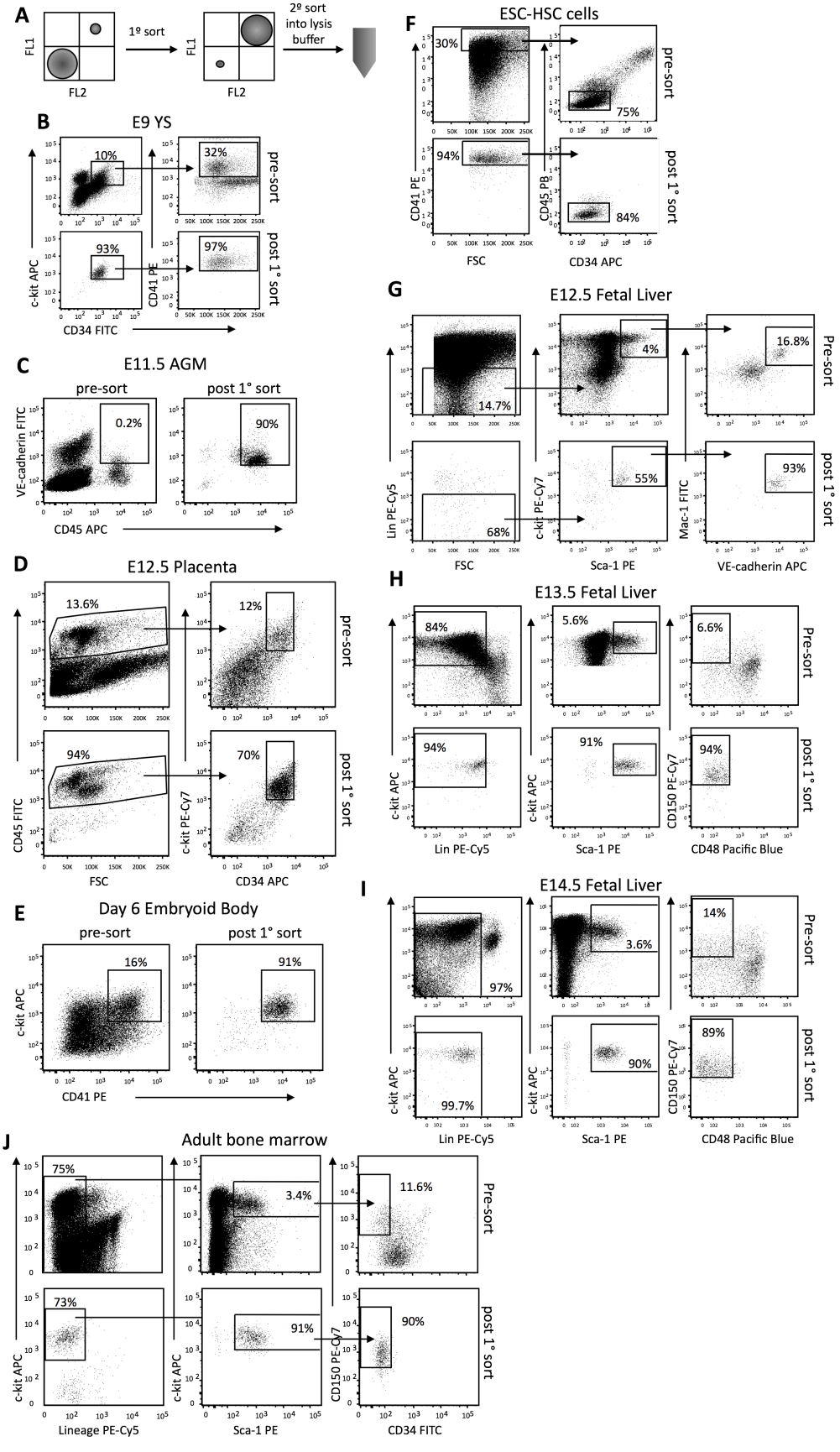
Cell Stem Cell, volume *11*

Supplemental Information

The Transcriptional Landscape of Hematopoietic Stem Cell Ontogeny

Shannon McKinney-Freeman, Patrick Cahan, Hu Li Scott A. Lacadie, Hsuan-Ting Huang, Matthew Curran Sabine Loewer, Olaia Naveiras, Katie L. Kathrein, Martina Konantz, Erin M. Langdon, Claudia Lengerke, Leonard I. Zon, James J. Collins, and George Q. Daley

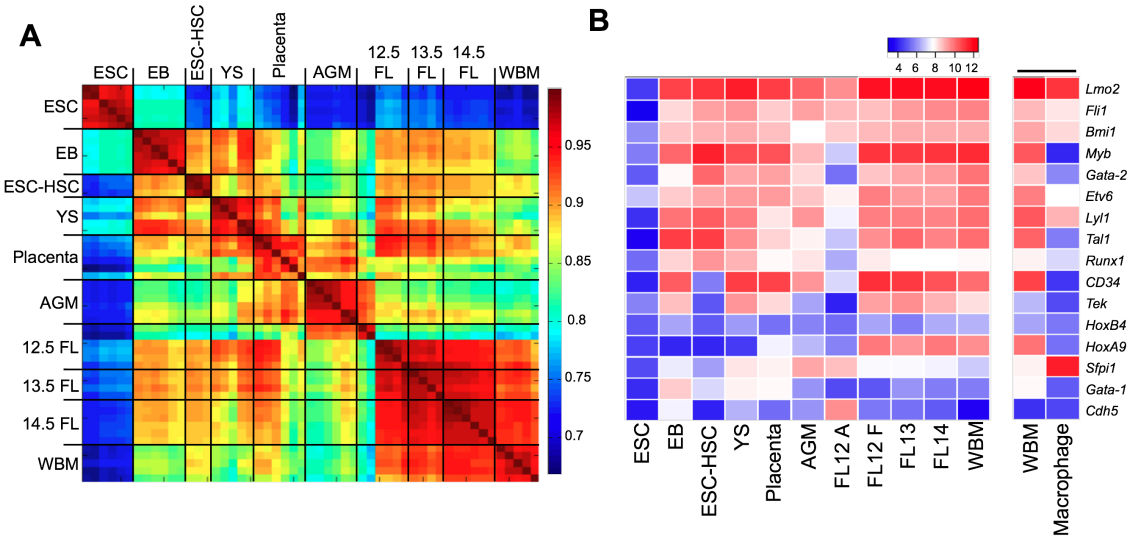
Figure S1.


A
FL1
FL2
1º sort
2º sort into lysis buffer
B
E9 YS
10%
32%
93%
97%
pre-sort
post 1° sort
c-kit APC
CD41 PE
CD34 FITC
C
E11.5 AGM
pre-sort
post 1° sort
0.2%
90%
VE-cadherin FITC
CD45 APC
D
E12.5 Placenta
13.6%
12%
94%
70%
CD45 FITC
c-kit PE-Cy7
FSC
CD34 APC
E
Day 6 Embryoid Body
pre-sort
post 1° sort
16%
91%
c-kit APC
CD41 PE
F
ESC-HSC cells
30%
75%
94%
84%
CD41 PE
CD45 PB
FSC
CD34 APC
G
E12.5 Fetal Liver
14.7%
4%
16.8%
68%
55%
93%
Lin PE-Cy5
c-kit PE-Cy7
Mac-1 FITC
FSC
Sca-1 PE
VE-cadherin APC
H
E13.5 Fetal Liver
84%
5.6%
6.6%
94%
91%
94%
c-kit APC
CD150 PE-Cy7
Lin PE-Cy5
Sca-1 PE
CD48 Pacific Blue
I
E14.5 Fetal Liver
97%
3.6%
14%
99.7%
90%
89%
Lin PE-Cy5
Sca-1 PE
CD48 Pacific Blue
J
Adult bone marrow
75%
3.4%
11.6%
73%
91%
90%
c-kit APC
CD150 PE-Cy7
Lineage PE-Cy5
Sca-1 PE
CD34 FITC
Pre-sort
post 1° sort

Supplemental Figure 1. Isolation of cell populations by FACS

(A) To isolate specific cell populations for array or Fluidigm, cells were first sorted into PBS from their respective tissues and then sorted a second time directly into lysis buffer. The following are representative analyses of the isolation of the indicated cell populations after their initial collection into PBS but prior to their final sort into lysis buffer: (B) $CD41^{+}c\text{-}kit^{+}CD34^{+}$ E9 YS, (C) $CD45^{+}$VE-$cadherin^{+}$ E11.5 AGM, (D) $CD45^{+}c\text{-}kit^{+}CD34^{med}$ E12.5 placenta, (E) $c\text{-}kit^{+}CD41^{+}$ EBs, (F) $CD41^{bright}CD45^{-}CD34^{-}$ EPOCH cells, (G) $Lin^{-}Sca\text{-}1^{+}c\text{-}kit^{+}$VE-$cadherin^{+}Mac\text{-}1^{low}$ E12.5 FL, (H) $Lin^{-}Sca\text{-}1^{+}c\text{-}kit^{+}CD150^{+}CD48^{-}$ E13.5 FL, (I) $Lin^{-}Sca\text{-}1^{+}c\text{-}kit^{+}CD150^{+}CD48^{-}$ E14.5 FL, and (J) $Lin^{-}Sca\text{-}1^{+}c\text{-}kit^{+}CD150^{+}CD34^{-}$ adult WBM.

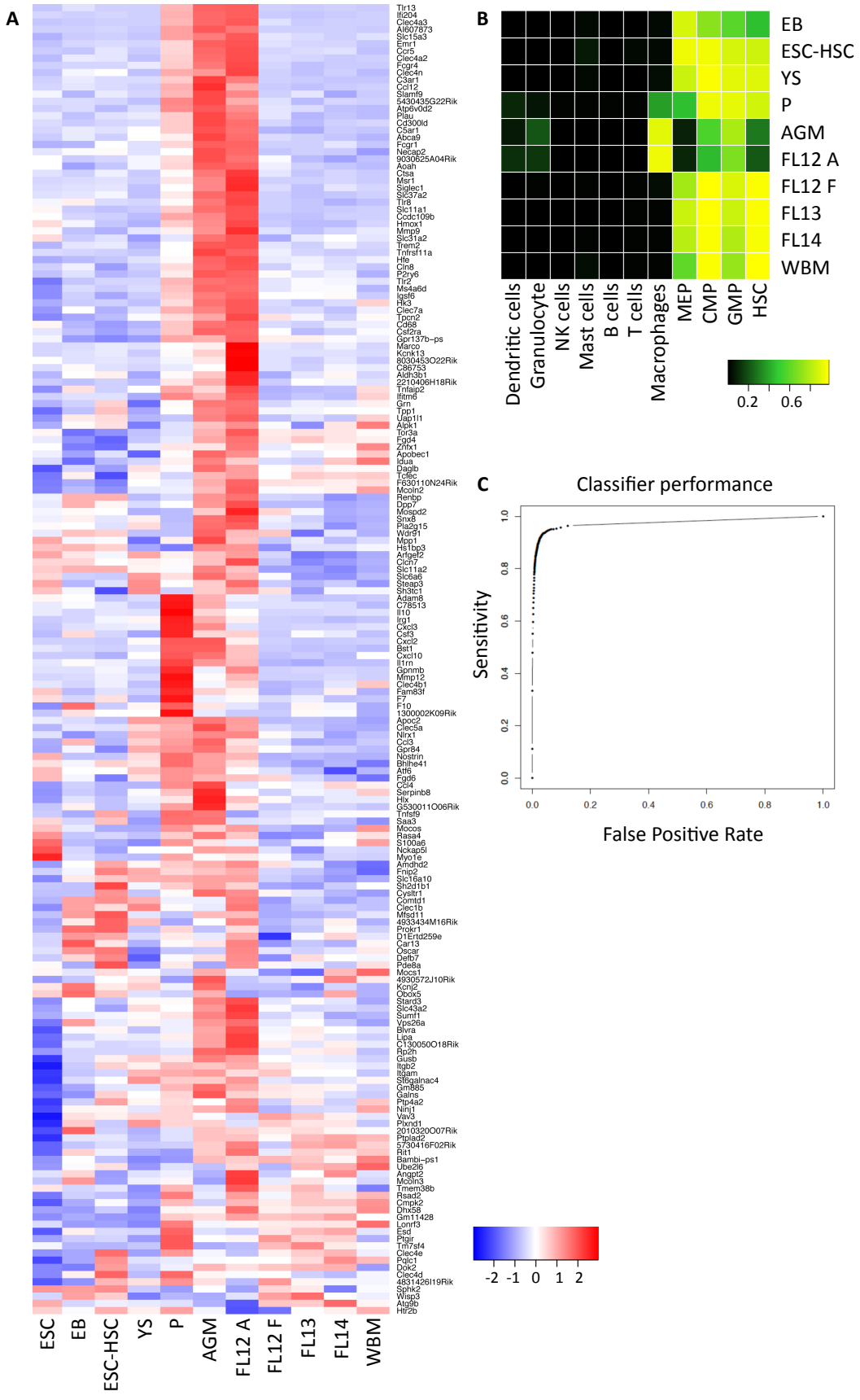
Figure S2.


A
ESC
EB
ESC-HSC
YS
Placenta
AGM
12.5 FL
13.5 FL
14.5 FL
WBM
0.95
0.9
0.85
0.8
0.75
0.7
B
4 6 8 10 12
Lmo2
Fli1
Bmi1
Myb
Gata-2
Etv6
Lyl1
Tal1
Runx1
CD34
Tek
HoxB4
HoxA9
Sfpi1
Gata-1
Cdh5
ESC
EB
ESC-HSC
YS
Placenta
AGM
FL12 A
FL12 F
FL13
FL14
WBM
WBM
Macrophage

Supplemental Figure 2. Pearson correlation and pathway enrichment analysis of pair-wise comparisons between each developmental hematopoietic population

(A) Pearson correlation between all 53 gene expression profiles. Most samples are well correlated. The variability among E12.5 placenta samples was anticipated because the placental HSC phenotype is based on fewer markers than other HSC populations, and thus less defined (Gekas et al., 2005). (B) Examination of the expression pattern of known HSC regulators. S*fpi1* (*PU.1*), critical for definitive hematopoiesis (Scott et al., 1994), was present in YS and placenta, highest in AGM and E12.5 FL F, and lower thereafter. Expression of the erythroid factor *Gata*-1 (Pevny et al., 1991) was high in E9 YS and E12.5 placenta, but low at the FL stage. *Hoxa9* was markedly up-regulated as development progressed: low in the YS, placenta and AGM stage, and high from the FL stage onward, a pattern highlighting the importance of *Hoxa9* as a regulator of definitive HSC (Lawrence et al., 1997). Other known regulators either did not vary dramatically during ontogeny (*e.g. Hoxb4, Runx1, Fli1*) or were consistently expressed during early development and only slightly up-regulated after the FL stage (*Lmo2, Bmi1, Myb, Gata-2, Etv6, Lyl1, Tal1, Cd34. Tek*). These genes are thus characteristic of the global transcriptional identity of primitive hematopoietic populations but cannot account for the developmental transitions that occur as HSC mature during ontogeny.

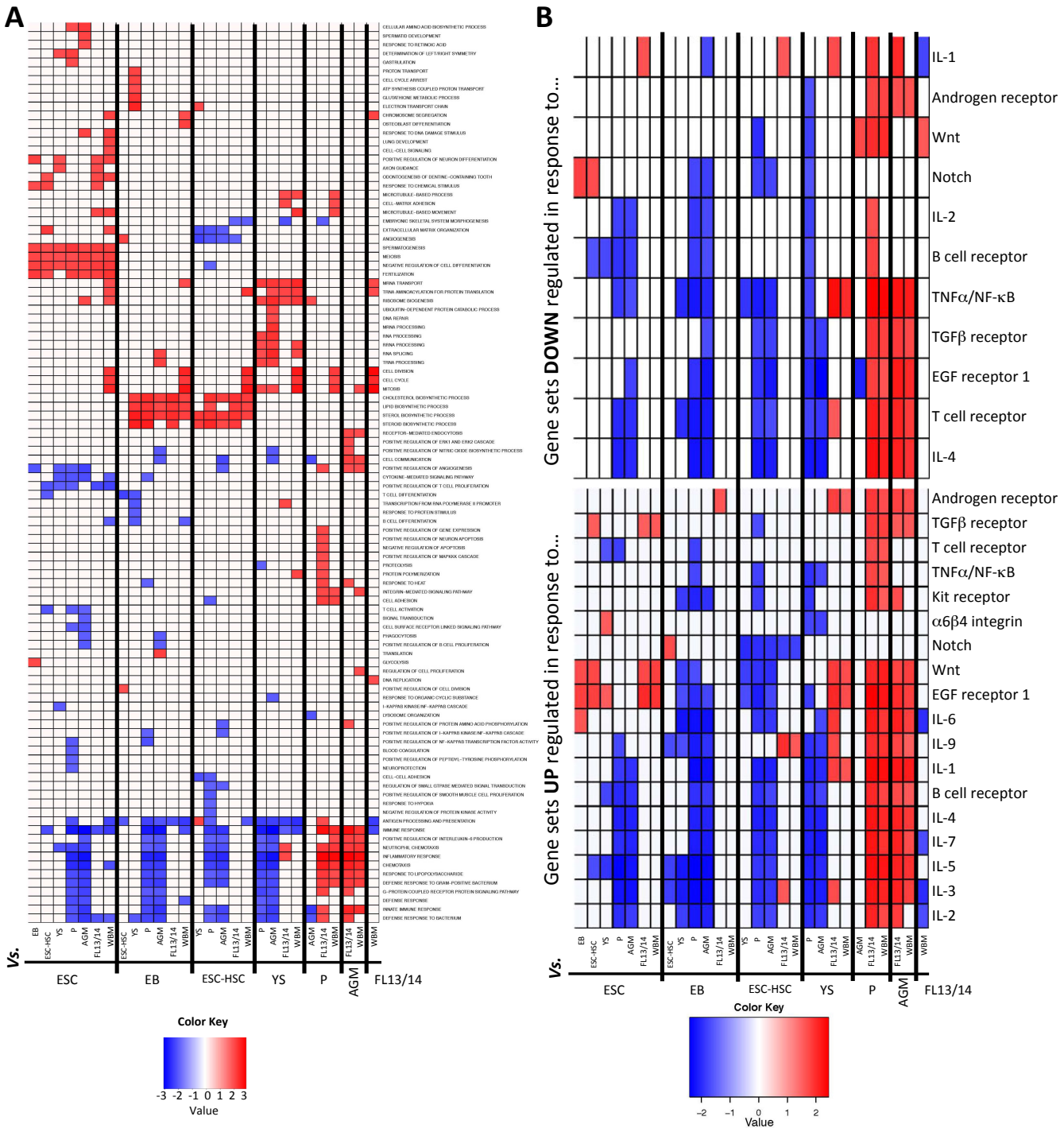
Figure S3.


A
ESC
EB
ESC-HSC
YS
P
AGM
FL12 A
FL12 F
FL13
FL14
WBM
-2 -1 0 1 2
B
EB
ESC-HSC
YS
P
AGM
FL12 A
FL12 F
FL13
FL14
WBM
Dendritic cells
Granulocyte
NK cells
Mast cells
B cells
T cells
Macrophages
MEP
CMP
GMP
HSC
0.2
0.6
C
Classifier performance
Sensitivity
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

Supplemental Figure 3. Macrophage gene set expression across samples and hematopoietic-trained cell type classifier

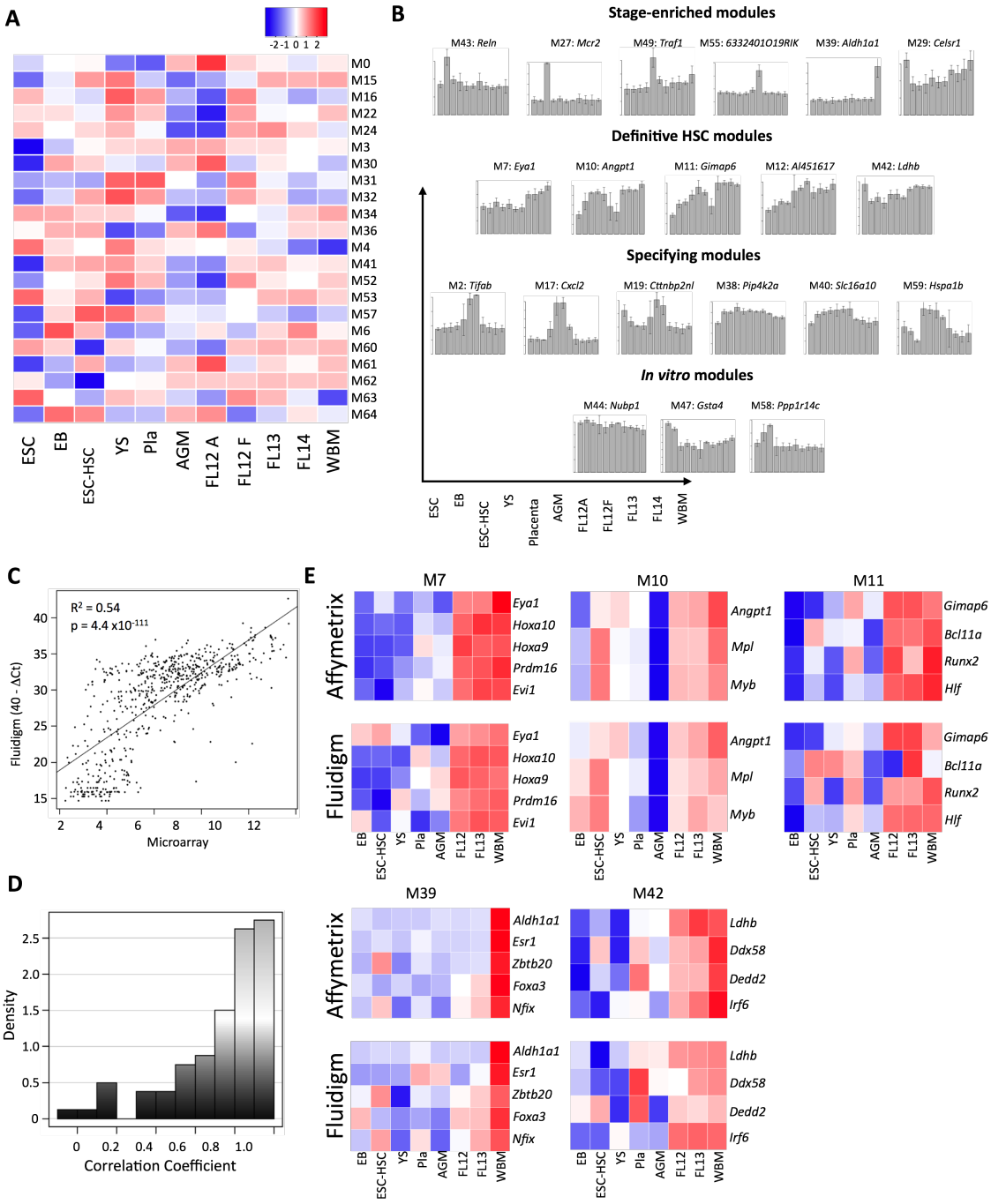
(A) Gene expression heatmap of the gene set whose expression is required to achieve a macrophage classification. (B) To ensure that the classification results were not biased due to incorporation of diverse cell types and tissues, a classifier was trained *de novo* using only hematopoietic populations. As in Figure 3A, each row is a biological group (*i.e.* WBM HSCs), and each column is a known tissue or cell type, in this case, only hematopoietic cell types. The classifier determines the posterior probability that a sample is indistinguishable from each of the tissues or cell types in the reference data set. Higher probabilities are bright yellow and low probabilities are dark green and black. (C) Receiver operating curve displaying the performance of the Naïve Bayesian classifier as determined by applying the algorithm to 1300 gene expression profiles of known origin. The false positive rate, defined as the number of incorrect classifications divided by the total number of classifications, is shown on the x-axis. The sensitivity, defined as the total number of correct classifications divided by the number of samples, is shown on the y-axis. Each point represents the sensitivity and false positive rate at a given probability cutoff.

Figure 4S


A
B
Vs.
ESC
EB
ESC-HSC
YS
P
AGM
FL13/14
Color Key
Value
Gene sets DOWN regulated in response to...
Gene sets UP regulated in response to...
IL-1
Androgen receptor
Wnt
Notch
IL-2
B cell receptor
TNFα/NF-κB
TGFβ receptor
EGF receptor 1
T cell receptor
IL-4
Kit receptor
α6β4 integrin
IL-6
IL-9
IL-7
IL-5
IL-3

Supplemental Figure 4. GSEA and Pathway Enrichment Analysis of Populations
(A) GSEA pathway enrichment analysis of pair-wise comparisons to find GO Biological Processes enriched or depleted between developmental populations. (B) NetPath analysis of pair-wise comparisons to identify signaling pathway-specific transcriptional responses activated or suppressed in each population.

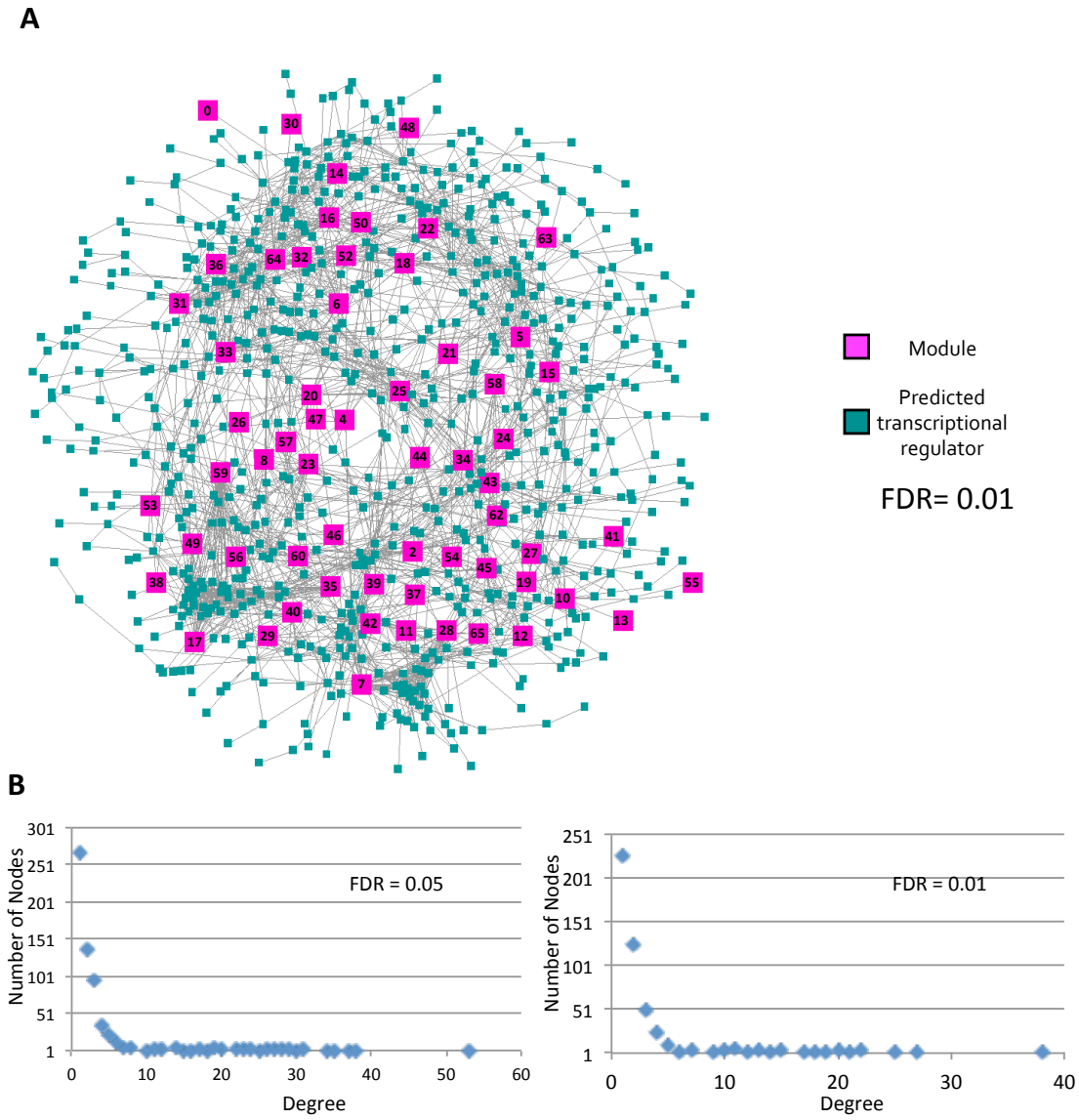
Figure S5.


A
-2 -1 0 1 2
M0
M15
M16
M22
M24
M3
M30
M31
M32
M34
M36
M4
M41
M52
M53
M57
M6
M60
M61
M62
M63
M64
ESC
EB
ESC-HSC
YS
Pla
AGM
FL12 A
FL12 F
FL13
FL14
WBM
B
Stage-enriched modules
M43: Reln
M27: Mcr2
M49: Traf1
M55: 6332401O19RIK
M39: Aldh1a1
M29: Celsr1
Definitive HSC modules
M7: Eya1
M10: Angpt1
M11: Gimap6
M12: Al451617
M42: Ldhb
Specifying modules
M2: Tifab
M17: Cxcl2
M19: Cttnbp2nl
M38: Pip4k2a
M40: Slc16a10
M59: Hspa1b
In vitro modules
M44: Nubp1
M47: Gsta4
M58: Ppp1r14c
ESC
EB
ESC-HSC
YS
Placenta
AGM
FL12A
FL12F
FL13
FL14
WBM
C
R² = 0.54
p = 4.4 x10-111
Fluidigm (40 - ΔCt)
Microarray
D
Density
Correlation Coefficient
E
M7
M10
M11
Affymetrix
Fluidigm
Eya1
Hoxa10
Hoxa9
Prdm16
Evi1
Angpt1
Mpl
Myb
Gimap6
Bcl11a
Runx2
Hlf
M39
M42
Aldh1a1
Esr1
Zbtb20
Foxa3
Nfix
Ldhb
Ddx58
Dedd2
Irf6

Supplemental Figure 5. Identification of stage-specific gene sets and Fluidigm validation

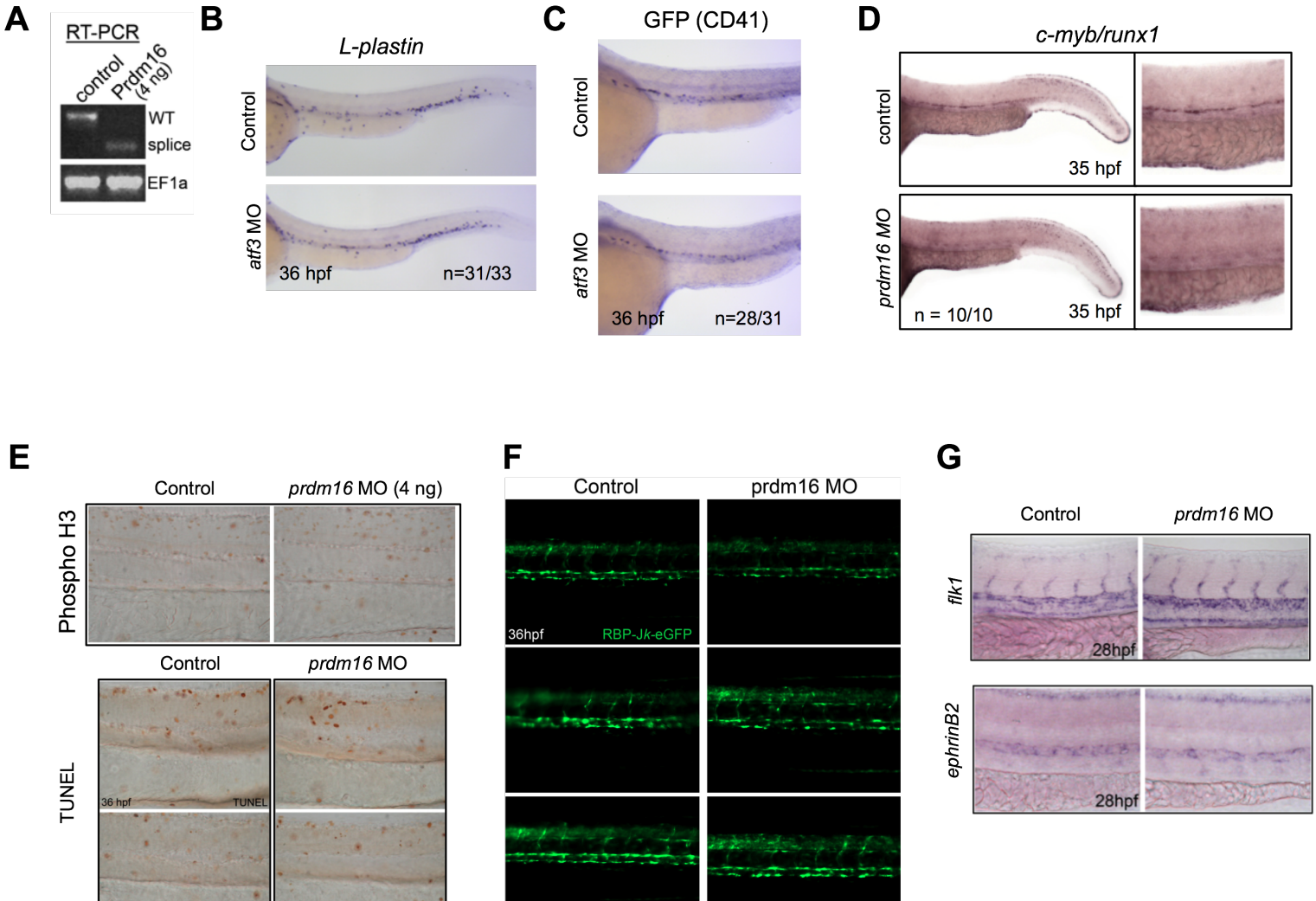
(A) Presentation of 22 WGCNA-determined co-regulated modules of genes that were not assigned as “stage enriched”, “definitive HSC”, “YS-like”, “specifying”, or “*in vitro*”. (B) The expression pattern across the dataset of the exemplars for stage-enriched modules (M43, M27, M49, M55, M39, and M29), select definitive HSC modules (M7, M10, M11, M12, M42), select specifying modules (M2, M17, M19, M38, M40, and M59), and in vitro modules (M44, M47, M58). (C) Results via gene array were highly correlated with results via Fluidigm analysis. (D) The majority of samples validated via Fluidigm showed an R^2 value of greater than 0.7. (E) Representative analysis of genes from five modules showing that the pattern of expression seen across development via array is preserved when independent samples of the same populations are examined via Fluidigm.

Figure S6.


A
Module
Predicted transcriptional regulator
FDR= 0.01
B
Number of Nodes
Degree
FDR = 0.05
FDR = 0.01

Supplemental Figure 6: The transcriptional regulatory network of HSCs in development. (A) The CLR (Context Likelihood of Relatedness) algorithm was applied to each WGCNA-derived gene module to identify putative transcriptional regulators for each gene set. The resulting predictions at the 0.01 FDR are presented as a network. Pink squares represent modules and blue squares represent predicted regulators. (B) The topology of the developing HSC gene regulatory network is scale-free. Histograms illustrating the node degree for the CLR network at FDR thresholds <0.05 (top) and < 0.01 (bottom). Node degree indicates the number of edges or putative regulatory relationships connected to each node. Most nodes have fewer than five edges.

Figure S7.


A
RT-PCR
control
Prdm16 (4 ng)
WT
splice
EF1a
B
L-plastin
Control
atf3 MO
36 hpf
n=31/33
C
GFP (CD41)
Control
atf3 MO
36 hpf
n=28/31
D
c-myb/runx1
control
prdm16 MO
35 hpf
n = 10/10
35 hpf
E
Control
prdm16 MO (4 ng)
Phospho H3
Control
prdm16 MO
TUNEL
36 hpf
TUNEL
F
Control
prdm16 MO
36hpf
RBP-Jk-eGFP
G
Control
prdm16 MO
flk1
28hpf
ephrinB2
28hpf

Supplemental Figure 7. Functional validation of select gene candidates implicated in definitive HSC regulation.

(A) RT-PCR for *prdm16* verify the splicing activity of the *prdm16*-MO. Wild-type product is 350bp, and the splicing out of exon 4 by the *prdm16*-MO results in a smaller 200bp product**.** (B) Whole mount *in situ* hybridization for *l-plastin* was performed on 36 hpf wild-type embryos that were either not injected or injected with *aft3*-MO. (C) Whole mount *in situ* hybridization for GFP was performed 36 hpf on embryos carrying a CD41-GFP transgene that were either not injected or injected with *atf3*-MO (Lin et al., 2005). (D) Whole-mount *in situ* hybridization was performed on 35 hpf wild-type embryos that were either not injected or injected with an independent *prdm16*-MO. (C) Phospho-histone H3 and TUNEL staining show no difference in mitosis or apoptosis, respectively, in *prdm16* morphants at 36hpf (n=4 control, n=4 morphants). (D) Knockdown of *prdm16* did not affect Notch signaling. The average GFP fluorescence between wild-type uninjected and *prdm16*-MO injected Notch reporter embryos were similar (n=5 control, n=5 morphants). (E) Whole-mount *in situ* hybridization of embryos 28 hpf either injected or not injected with *prdm16*-MO for *Flk1* and *ephrinB2*. (F) Whole-mount *in situ* hybridization of embryos at the 16 somite stage either injected or not injected with *prdm16*-MO for *b-globin e3*.

Supplementary Table 1.

Annotation			Microarray data							Fluidigm			
Symbol	Module	Type	Correlation	Z.score	Minimum	Maximum	Mean	Median	Variance	TaqMan.ID	p-value	MA/Flu Correlation	Efficiency
rps29	NA	control	NA	NA	NA	NA	NA	NA	NA	Mm02342448_gH*	NA	NA	0.944
angpt1	ME10	exemplar	0.97	0	3.43	10.64	7.64	8.49	4.31	Mm00456503_m1*	1.07E-03	-0.923	1.128
erg	ME10	Regulator	0.95	6.62	4.59	10.1	8.12	8.56	2.63	Mm00504897_m1	5.32E-01	0.261	-1
mpl	ME10	Regulator	0.91	5.64	4.68	10.34	8.24	8.68	2.74	Mm00440310_m1*	3.69E-04	-0.946	1.011
myb	ME10	Regulator	0.92	5.57	4.74	11.68	9.89	10.77	3.54	Mm00501741_m1*	5.10E-03	-0.869	1.223
gimap6	ME11	exemplar	0.93	0	3.5	11	7.9	8.43	4.82	Mm00462641_m1*	1.07E-04	-0.965	0.815
bcl11a	ME11	Regulator	0.93	5.73	4.23	9.18	7.07	7.43	2.28	Mm00479358_m1*	3.91E-01	-0.353	1.139
hlf	ME11	Regulator	0.87	5.34	3.97	11.22	7.53	7.14	6.04	Mm00723157_m1*	9.32E-04	-0.927	0.756
runx2	ME11	Regulator	0.82	5.41	3.7	8.26	5.65	5.69	1.44	Mm00501584_m1*	7.13E-04	-0.933	1.003
ai451617	ME12	exemplar	0.97	0	3.43	8.44	6	6.42	2.15	Mm01162558_m1*	1.85E-03	-0.907	1.149
irf9	ME12	Regulator	0.93	7.42	4.71	9.1	7.4	7.75	1.25	Mm00492673_m1	4.44E-03	-0.875	0.98
nmi	ME12	Regulator	0.89	5.72	5.15	10.4	8.47	9.08	2.14	Mm00803857_m1*	6.91E-03	-0.854	1.092
osgin1	ME12	Regulator	0.89	6.13	4.27	8.09	6.71	7.17	1.11	Mm00660947_m1*	7.67E-03	-0.849	0.805
cxcl2	ME17	exemplar	0.95	0	2.69	13.5	5.52	3.48	12.44	Mm00436450_m1*	3.33E-04	-0.948	0.965
fosl2	ME17	Regulator	0.84	7.21	5.29	9.2	6.64	6.41	0.84	Mm00484442_m1*	2.05E-02	-0.787	1.029
nr4a3	ME17	Regulator	0.91	7.52	3.83	6.91	4.54	4.28	0.44	Mm00450074_m1*	4.25E-03	-0.877	1.157
rel	ME17	Regulator	0.89	7.71	3.68	9.06	5.98	5.73	1.73	Mm01239661_m1*	3.37E-02	-0.746	0.912
cttnbp2nl	ME19	exemplar	0.96	0	3.25	8.69	5.14	4.4	2.3	Mm00518765_m1*	1.46E-02	-0.811	0.777
rps6ka4	ME19	Regulator	0.78	6.05	3.83	7.17	5.25	5.17	0.67	Mm00451280_m1*	9.27E-03	-0.839	1.324
tcfe3	ME19	Regulator	0.87	5.56	6.4	10.55	8.32	7.94	1.16	Mm01341186_m1*	1.51E-02	-0.808	1.24
wwp1	ME19	Regulator	0.86	5.4	5.98	8.51	7.07	6.98	0.34	Mm01210682_m1*	2.23E-02	-0.78	1.073
tifab	ME2	exemplar	0.97	0	3.17	8.32	4.48	3.89	2.49	Mm04210261_m1*	1.25E-01	-0.588	1.012
ifi204	ME2	Regulator	0.91	6.33	2.21	9.58	3.8	2.66	5.35	Mm00492602_m1*	1.59E-03	-0.912	0.976
irf8	ME2	Regulator	0.86	6.15	3.42	9.78	6.52	6.3	3.03	Mm01250091_m1	1.22E-02	-0.823	0.966
maf	ME2	Regulator	0.92	6.18	3.18	9	5	4.31	2.97	Mm02581355_s1*	3.32E-02	-0.747	0.595
mc2r	ME27	exemplar	0.97	0	2.22	10.16	3.32	2.87	2.9	Mm01262510_m1*	1.61E-06	-0.991	1.021
dpf3	ME27	Regulator	0.67	4.96	4.34	6.03	4.91	4.82	0.13	Mm00475440_m1*	2.75E-03	-0.894	0.871
mbd2	ME27	Regulator	0.55	3.32	8.12	10.41	9.15	9.16	0.26	Mm00521967_m1*	1.64E-01	-0.544	1.05
tcf4	ME27	Regulator	0.53	4.03	5.56	8.81	7.37	7.43	0.38	Mm00443210_m1*	3.93E-01	-0.351	0.958
celsr1	ME29	exemplar	0.93	0	2.85	7.16	4.74	4.42	1.33	Mm00464808_m1*	3.55E-02	-0.741	0.977
b930041f14rik	ME29	Regulator	0.84	6.83	4.21	7.5	5.88	5.8	0.74	Mm00844649_s1*	5.23E-02	-0.702	0.751
klf12	ME29	Regulator	0.84	6.12	3.17	5.47	4.2	4.12	0.32	Mm00516098_m1*	1.79E-02	-0.797	1.033
pml	ME29	Regulator	0.87	5.43	5.32	8.76	6.57	6.35	0.77	Mm00476969_m1*	6.83E-01	-0.173	1.146
evi2a	ME3	exemplar	0.98	0	2.61	8.07	6.54	6.84	2.02	Mm03030107_s1*	5.80E-02	-0.69	1.124
hcls1	ME3	Regulator	0.97	4.39	2.46	9.11	7.21	7.83	3.27	Mm00468528_m1*	7.45E-01	0.138	0.196
ifnar2	ME3	Regulator	0.95	4.23	4.87	9.9	8.12	8.35	1.67	Mm00494916_m1*	1.89E-01	-0.517	0.917
tal1	ME3	Regulator	0.67	3.95	2.75	11.23	8.79	9.57	4.88	Mm01187033_m1*	1.36E-01	-0.574	0.983
pip4k2a	ME38	exemplar	0.94	0	5.71	9.51	7.78	8.11	0.77	Mm00435721_m1*	6.46E-01	-0.194	0.56
foxn2	ME38	Regulator	0.78	4.74	5.48	8.32	6.91	6.83	0.49	Mm00839106_g1	8.00E-01	-0.108	0.882
zdhhc16	ME38	Regulator	0.75	4.26	7.62	9.12	8.52	8.57	0.13	Mm00470108_m1*	7.98E-02	-0.652	1.085
zdhhc21	ME38	Regulator	0.78	4.39	6.3	8.8	7.54	7.64	0.35	Mm00509795_m1*	4.40E-02	-0.72	0.751
aldh1a1	ME39	exemplar	0.96	0	2.5	10.55	3.51	2.91	3.71	Mm00657317_m1*	2.98E-06	-0.989	0.894
esr1	ME39	Regulator	0.95	7.94	2.71	7.13	3.4	3.1	0.99	Mm00433149_m1*	6.72E-03	-0.856	0.976
foxa3	ME39	Regulator	0.89	6.5	3.25	7.85	4.55	4.31	1.26	Mm00484714_m1*	4.91E-03	-0.871	1.049
nfix	ME39	Regulator	0.76	5.31	4.35	8.42	6.23	6.18	1.11	Mm00477796_m1	2.41E-04	-0.954	1.097
zbtb20	ME39	Regulator	0.88	6.51	3.72	7.13	4.75	4.57	0.57	Mm00457765_m1*	1.06E-02	-0.831	1.186
slc16a10	ME40	exemplar	0.91	0	4.24	7.37	5.88	5.95	0.85	Mm00661045_m1*	5.43E-03	-0.866	0.902
cebpa	ME40	Regulator	0.87	6.76	3.36	9.56	6.35	6.17	2.95	Mm00514283_s1*	1.75E-01	-0.532	1.218
klf7	ME40	Regulator	0.83	6.29	5.04	8.61	6.69	6.57	0.98	Mm00728361_s1*	5.08E-01	-0.276	2.144
zbtb16	ME40	Regulator	0.72	4.71	4.23	7.56	5.83	5.64	1.22	Mm01176868_m1*	3.66E-02	-0.738	1.04
ldhb	ME42	exemplar	0.93	0	6.53	12.67	9.52	9.82	2.36	Mm00493146_m1*	8.19E-02	-0.649	1.4
ddx58	ME42	Regulator	0.81	6.05	3.95	7.29	5.75	5.86	0.6	Mm00554529_m1*	1.05E-01	-0.615	1.197
dedd2	ME42	Regulator	0.7	5.76	5.14	7.47	6.57	6.6	0.38	Mm01149726_m1*	2.69E-01	-0.445	1.094
irf6	ME42	Regulator	0.76	5.53	2.75	8.71	5.86	6.13	2.6	Mm00516797_m1*	1.43E-03	-0.915	0.449
reln	ME43	exemplar	0.96	0	2.85	7.36	3.69	3.28	1.12	Mm00465200_m1*	2.27E-06	-0.99	0.812
hoxd1	ME43	Regulator	0.95	8.64	2.06	7.19	3.03	2.56	1.87	Mm00439370_g1*	1.44E-04	-0.961	0.945
hoxd8	ME43	Regulator	0.9	7.49	3.26	7.85	4.3	3.87	1.29	Mm03016337_m1*	3.33E-03	-0.887	0.803
klf1	ME43	Regulator	0.78	5.74	4.16	10.29	6.45	6.23	1.99	Mm00516096_m1*	4.52E-02	-0.717	0.86
nubp1	ME44	exemplar	0.92	0	6.35	8.66	7.51	7.49	0.33	Mm00478752_m1*	4.22E-01	-0.332	1.134
gtf2f2	ME44	Regulator	0.83	6.08	7.83	10.64	9.17	9.1	0.49	Mm01310683_m1*	2.19E-01	-0.489	0.775
psmc3	ME44	Regulator	0.83	6.28	9.71	11.45	10.89	10.95	0.12	Mm00477177_m1*	9.60E-01	-0.021	1.063
rnf141	ME44	Regulator	0.75	5.39	6.09	8.06	6.95	6.84	0.28	Mm01130671_g1*	6.37E-01	-0.199	0.949
gsta4	ME47	exemplar	0.95	0	4.68	13.41	8.61	8	5.15	Mm00494803_m1*	1.20E-02	-0.823	0.778
bc003267	ME47	Regulator	0.8	5.32	3.96	8.03	6.34	6.38	0.92	Mm00728712_s1*	6.91E-01	-0.168	0.758
gtf2h3	ME47	Regulator	0.8	4.73	5.97	9.01	7.8	7.88	0.47	Mm00619444_m1*	5.06E-03	-0.869	1.03
uba3	ME47	Regulator	0.8	3.97	8.25	10.77	9.42	9.41	0.32	Mm00495866_m1*	1.78E-02	-0.797	1.136
traf1	ME49	exemplar	0.93	0	3.39	9.38	4.96	4.61	2.12	Mm00493827_m1*	7.84E-03	-0.848	1.579
foxa1	ME49	Regulator	0.62	6.83	2.7	5.85	3.39	3.16	0.47	Mm00484713_m1*	6.27E-02	-0.681	0.18
irf4	ME49	Regulator	0.8	6.5	3.54	6.28	4.32	4.09	0.4	Mm00516431_m1*	3.25E-02	-0.749	1.009
mxd1	ME49	Regulator	0.9	7.19	5.64	9.08	6.73	6.6	0.6	Mm00487504_m1*	4.89E-03	-0.871	1.176
6332401o19rik	ME55	exemplar	0.97	0	2.97	7.16	3.53	3.39	0.49	Mm00844775_s1*	8.74E-01	0.067	1.25
mafb	ME55	Regulator	0.78	4.27	3.96	8.72	4.9	4.42	1.34	Mm00627481_s1*	1.40E-01	-0.57	0.976
nr1h3	ME55	Regulator	0.91	4.88	4.7	10.65	5.75	5.46	1.29	Mm00443451_m1*	7.87E-02	-0.654	0.802
spic	ME55	Regulator	0.76	3.7	3.13	10.55	4.5	3.94	2.76	Mm00488428_m1*	6.44E-05	-0.97	1.203
ppp1r14c	ME58	exemplar	0.94	0	2.9	7.3	4.35	3.86	1.13	Mm00652462_m1	1.08E-01	-0.611	0.821
etv2	ME58	Regulator	0.6	5.21	3.99	9.53	5.53	4.95	2.39	Mm00468389_m1*	9.21E-03	-0.839	1.024
hmgcs1	ME58	Regulator	0.81	7.43	3.06	7.02	4.41	4	0.98	Mm00524111_m1	5.91E-03	-0.862	0.937
zfpm2	ME58	Regulator	0.8	5.92	3.08	8.03	4.35	3.79	1.46	Mm00496074_m1*	5.57E-04	-0.938	1.094
hspa1b	ME59	exemplar	0.95	0	3.01	11.4	6.83	5.57	7.46	Mm03038954_s1*	2.31E-01	-0.478	0.818
egr3	ME59	Regulator	0.92	8.68	4.06	8.27	5.65	4.92	1.8	Mm00516979_m1*	2.17E-03	-0.902	0.822
fosb	ME59	Regulator	0.88	8.28	2.69	10.63	5.59	3.8	8.33	Mm00500401_m1*	2.15E-03	-0.903	1.195
klf6	ME59	Regulator	0.93	8.67	5.91	10.9	8.12	7.5	2.32	Mm00516184_m1*	1.17E-03	-0.921	1.235
ampd3	ME62	exemplar	0.85	0	2.61	8.11	6.59	6.81	1.16	Mm00477495_m1*	1.15E-03	-0.921	1.156
bach2	ME62	Regulator	0.71	4.48	3.96	6.44	5.43	5.35	0.25	Mm00464379_m1*	2.72E-03	-0.894	1.212
fmnl2	ME62	Regulator	0.73	5.04	3.8	9.72	8.02	8.27	1.24	Mm00549621_m1	6.20E-02	-0.683	0.986
htatip2	ME62	Regulator	0.68	4.69	4	7.38	5.5	5.58	0.6	Mm00457476_m1*	1.69E-03	-0.91	0.826
eya1	ME7	exemplar	0.97	0	3.42	7.63	4.97	4.44	1.52	Mm00438796_m1*	6.38E-02	-0.679	1.212
evi1	ME7	Regulator	0.87	7.05	2.92	7.89	5.56	5.33	2.49	Mm00514814_m1*	4.58E-02	-0.716	0.827
hoxa10	ME7	Regulator	0.94	7.97	2.41	5.58	3.95	3.49	1.26	Mm00433966_m1*	2.79E-05	-0.978	0.903
hoxa9	ME7	Regulator	0.89	7.23	3.53	10.22	6.94	7.06	5.61	Mm00439364_m1*	2.46E-05	-0.978	0.753
prdm16	ME7	Regulator	0.91	7.15	3.86	7.71	5.75	5.41	1.3	Mm00712556_m1*	3.29E-03	-0.887	0.775

Type indicates id the candidate is an exemplar (most higly correlated with module profile), or one of the top three predicted regulators, or a control genes.

Correlation is the Pearson correlation of the candidate gene expression wih the module profile in the microarray data.

Z.score: the CLR-calculated Z-score representing the significance of the association between the candidate regulator and module.

Minimum, Maximum, Mean, Median, and Variance: Range, center and variance estimates for the gene expression from the microarray data.

TaqMan.ID: Assay identifier used in the Fluidigm experiments.

MA/Flu Correlation: Pearson correlation between the biological-grouped averaged microarray expression values and Fluidigm deltaCts.

Efficiency: TaqMan efficiency estimate as determined by estimating the slope of a standard curve with six ten fold dilutions of a pooled sample and the formula: effeciency = 10^(-1/slope)-1.

Supplemental Methods

Antibodies used for cell sorting

CD41 (MWreg-30), c-kit (2B8), CD34 (RAM34), CD45 (30-F11), VE-cadherin (11D4.1), ter119 (ter119), CD3(145-2C11), CD4 (GK1.5), CD8 (536.7), IgM (11/41), CD19 (1D3), Gr-1 (RB6-8C5), Mac-1 (M1/70), CD48 (HM48-1, Biolegend), CD150 (TC15-12F122, Biolegend), Sca-1 (E13-161.7), anti-rat PE-Cy5, anti-rat FITC, and anti-rat APC. Unless otherwise indicated, all antibodies were obtained from Ebiosciences.

Embryo dissections and cell preparations

The Children's Hospital Boston institutional review board approved these studies. All embryonic tissues were dissected from C57Bl/6 mice and treated as previously described (McKinney-Freeman et al., 2009).

Embryoid body and EPOCH cell culture

iNotch or *iCdx4* ESC were created, maintained, and differentiated into EBs or ESC-HSC cells as previously described (Kyba et al., 2002; McKinney-Freeman et al., 2009). For *iNotch* cells, ESC were differentiated as previously described with the addition of 0.5 μg/mL doxycycline between days4-day6 EB differentiation. At day 6, EB-derived cells were isolated, transduced with retroviral *HoxB4*, and expanded on OP9 stroma exactly as previously described for *iCdx4* ESC (McKinney-Freeman et al., 2009). ESC-HSC isolated from *iNotch* or *iCdx4* ESC were functionally, phenotypically, and transcriptionally indistinguishable (McKinney-Freeman S, data not shown).

Zebrafish maintenance and microinjection

Zebrafish *(Danio rerio)* were maintained according to Animal Research guidelines at Children's Hospital Boston. Tüebingen strain and the Notch reporter line *TP1bglob:gfp* (Parsons et al 2009) were used. Embryos were developed at 28.5°C and staged according to hpf and morphological features (Kimmel et al 1995). Morpholinos were injected at the 1-cell stage. Splice blocking morpholino targeting exon 4 of *prdm16* (5'-ACTCACACTATCACCCACCTTATCA-3' or 5'-ATGACTTGCATTAATCTCACCTTCC-3') and translation morpholino targeting rfx5 (5'-CTGCCTTTAACTGATCTTCTGCCAT-3') were ordered from Gene-Tools and dissolved in water.

Protein sequence and genomic synteny revealed locus LOC100151531 as the most likely candidate for a zebrafish orthologue of the gene GFI1B. *Gfi1b* morpholino sequence targeting the intron-exon boundary of exon 5: 5'-GCTTCTTTCCTGTAAACACAAAACA-3'. Antisense splicing morpholino with sequence TGTTTATTTAACTTACCACCTCTGT targeting the exon2/intron2 boundary of *msrb2* was injected into 1-4 cell embryos and assayed at 36hpf for *runx1* and *c-myb* expression. Antisense splicing morpholino with sequence TGTAGTAATGGAGTGTTTACCCTGC targeting the exon4/intron4 boundary of *tulp4* was injected into 1-4 cell embryos and assayed at 36hpf for *runx1* and *c-myb* expression. Disrupted splicing was verified by RT-PCR with primers 5-GTGGTGCTGGTGCGCTGGAACGAGCCCTTCC-3 and 5-GGGTCGGAAGTCACGAGTCTCTCCATCCC-3. For *gfi1b*, *msrb2*, and *tulp4* morpholinos, 4 ng was injected. 0.5 ng of antisense splicing morpholino with sequence GGGACAGCCTGAAATAACAACATCT targeting the intron2/exon3 boundary of *atf3* was injected into 1-4 cell embryos and assayed at 36hpf for *runx1* and *c-myb* expression. Disrupted splicing was verified by RT-PCR with primers 5-ATTTCGGCATGATGCTTCAGCACCCTGG-3 and 5-TCTTCGGGGGTCTGGCCGTTCTGAGCG-3. Total RNA was extracted from embryos staged at 36hpf in 500 μL as previously reported. Primers to amplify exons 3 to 5 of *prdm16* were as follows: forward 5'-AAGCAGGAGCGGGAAGACAG-3' and reverse 5'-TGTGCTTGTGCTGCTTGAGG-3' (Draper et al., 2001).

Whole-mount in situ hybridization, phosphoH3, and TUNEL staining

Whole-mount in situ hybridization in zebrafish embryos was performed as described previously (Thisse and Thisse, 2008). Stained embryos were mounted in glycerol and imaged on a Nikon E600 compound microscope with a Nikon Coolpix 4500 camera. Phospho-histone H3 staining was performed as decribed previously (Shepard et al 2005) with diaminobenzidine/H_2O_2 (Invitrogen). TUNEL staining was performed on fixed embryos as described (Shepard et al., 2005). Equal number of control and morpholino injected embryos were imaged in the tail region above the yolk sac extension, and the images were used for counting the number of mitotic or apoptotic cells, averaging the results.

Fluidigm

The exemplar and top 3 CLR-predicted regulators for 22 modules were validated via qRT-PCR on the Fluidigm BioMark 96x96 platform (Fluidigm, San Francisco, CA, Supplemental Table 2). cDNA from 200 cells per sample was pre-amplified using the pool of TaqMan probes per the Fluidigm instructions. As a quality control for each assay, we performed a standard curve of six 10-fold dilutions of pre-amplified cDNA pooled from all samples. The 80 of 92 non-control assays with amplification efficiencies between 0.75 and 1.25 were considered reliable and assessed for correlation with the microarray data. Delta Cts were calculated in reference to the housekeeping gene *Rps29*. To determine whether a gene replicated the expression pattern observed in the microarray data, we calculated the correlation coefficient between the average of the deltaCts for each replicate of a biological group and the average gene value from the microarrays. We considered genes validated when the p-value of the correlation coefficient <0.10.

Microarray

Raw microarray signal intensities were RMA-summarized and quantile normalized using R/BioConductor (Bolstad et al., 2003; Gentleman et al., 2004; Irizarry et al., 2003). To correct for batch effects, we applied the ComBat batch correction algorithm to the normalized data (Johnson et al., 2007), then multiple probesets mapping to the same gene were averaged, resulting in 21,308 gene expression measures per sample. The expression of all genes was used in the analyses displayed in Figure 2. We used hierarchical clustering with the 'average' linkage method for the dendrogram in Figure 2A. In subsequent analyses steps we used the 13,530 genes detected as expressed in at least one biological group. To find genes differentially expressed between biological groups, we performed pair-wise Student's T-tests, treating genes with nominal p-values<0.05 and fold change>2 as differentially expressed. Stage-specific and stage-enriched modules were identified as described in the Methods section. To find sets of positively correlated genes (modules), we used WGCNA, setting beta=15. To find enrichment of Gene Ontology and NetPath gene sets, we applied Gene Set Enrichment Analysis to gene lists ranked by fold change for each specified comparison. All analysis results, including differentially expressed gene lists, genes in modules, and transcription factor regulators of modules, and raw data, are freely available through the companion website, http://hsc.hms.harvard.edu/. The Context likelihood of relatedness (CLR) algorithms uses mutual information to determine the statistical dependence between

transcription factor and putative target module expression. To compute mutual information, we used a B-spline smoothing and discretization method, implementing the Freedman-Diaconis rule to estimate the bin width size. All mutual information values were computed using 6 bins and third order B-splines and a MatLab interface to B-spline mutual information estimation code library is available at the companion website.

Naïve Bayesian classifier and context-dependent gene regulatory networks

We trained a Naïve Bayesian classifier (Mitchell, 1997) using 130 gene expression profiles from three independent experiments representing 44 cell types and tissues (GSE10246, GSE14012, and GSE10806). The classifier uses Bayes' Theorem to compute the posterior probability that a query gene expression profile is indistinguishable from biological replicates of each of 44 distinct cell types and tissues that are in the training data set, given distributions of inter- and intra-cell type distances:

$Pr(que = Ref_i \mid Dist_{q,i} = X) = Pr(Dist_{q,i} = X \mid que = Ref_i) * Pr(que = Ref_i) / (Pr(Dist_{q,i} = X \mid que = Ref_i) + Pr(Dist_{q,i} = X \mid que \neq Ref_i))$,

where que = query sample, Ref_i = reference sample i, $Dist_{q,i}$ = distance between query and Ref_i profiles. Assuming equivalent prior probabilities that a query is indistinguishable from each reference sample, this becomes:

$Pr(que = Ref_i \mid Dist_{q,i} = X) = Pr(Dist_{q,i} = X \mid que = Ref_i) / (Pr(Dist_{q,i} = X \mid que = Ref_i) + Pr(Dist_{q,i} = X \mid que \neq Ref_i))$

The likelihoods $Pr(Dist_{q,i} = X \mid que = Ref_i)$ and $Pr(Dist_{q,i} = X \mid que \neq Ref_i)$ are computed directly from the estimated distributions of module profile distances between replicates (approximated using an exponential distribution) and between different tissues and cell types (approximated using a normal distribution) from the training data. Distances are based on gene set (rather than individual gene) profiles, and gene sets were determined by clustering approximately 20,000 genes across the 44 cell types and tissues using WGCNA (Zhang and Horvath, 2005). The output of the classifier is vector of probabilities. To test whether the classification results were biased due to the inclusion of diverse cell types in the training data, we constructed the classifier de novo using only hematopoietic populations (HSCs, CMPs, GMPs, MEPs, macrophages, granulocytes, T-cells, B-cells, natural killer cells, mast cells, and dendritic cells). When we applied this classifier to our developmental gene expression profiles, we found that the results were highly similar to the original classification results, with AGM and FL12 A still classified as macrophages, and Definitive samples as HSPCs (Figure S3B). To test the overall

performance of the classifier, we applied it to a validation data set of 1,300 publicly available gene expression profiles (GSE10733, GSE10744, GSE11056, GSE11110, GSE11207, GSE11220, GSE12464, GSE12545, GSE13032, GSE13155, GSE13224, GSE13402, GSE13526, GSE13753, GSE13805, GSE13873, GSE14270, GSE15129, GSE16073, GSE16150, GSE16364, GSE16994, GSE17263, GSE17923, GSE18500, GSE18669, GSE18746, GSE19299, GSE19403, GSE20352, GSE21754, GSE21842, GSE22527, GSE22935, GSE2389, GSE2869, GSE3203, GSE3440, GSE3554, GSE3653, GSE4035, GSE4040, GSE4142, GSE4413, GSE4816, GSE5127, GSE5296, GSE5763, GSE6210, GSE6461, GSE6466, GSE6506, GSE6514, GSE6591, GSE6623, GSE6676, GSE6686, GSE7069, GSE7196, GSE7333, GSE7381, GSE7407, GSE7764, GSE7793, GSE7798, GSE8000, GSE8044, GSE8199, GSE8249, GSE8582, GSE9630, GSE9711, GSE9810, GSE9913, GSE9954, GSE24637, GSE25140, GSE10627, GSE21018, GSE12982), demonstrating that the classifier achieved a sensitivity of 94% at a false positive rate <5% (data not shown).

We used the same training data (130 gene expression profiles) and gene sets described above to reconstruct the gene regulatory networks of adult cell types and tissues. We searched for potential gene set regulators by computing the Pearson correlation coefficient between the expression level of each transcription factor and each gene set profile both globally (using all cell types and tissues) and in a context-specific manner (using only the subset of cell types and tissues that share a developmental origin). Reasoning that TFs central to module regulation would also be highly correlated to the same gene set profiles in independent data sets of the same cell types and tissues, we computed TF and gene set correlations in the validation data set (1,300 publicly available gene expression profiles) and removed relationships in which either the direction of the correlation differed or was not significant.

Companion website

There is more information within our data set than can be described in detail here. Furthermore, we believe that this data and analyses will be informative beyond our focus on the development of the HSC. For example, our analyses can be further leveraged to determine the extent to which the developmental programs are re-activated in cancer, or in diverse physiological conditions. In the hope that other investigators will leverage this resource to complement their research, we created a website and database

(http://hsc.hms.harvard.edu) to facilitate data mining and data integration. We designed the website to be user-friendly so that visitors can quickly perform one-off queries (e.g. "What is the expression profile of my gene in HSC development?"). We also implemented a feature to allow visitors to upload a gene list and the website will determine the modules in which the gene list is enriched. We also provide access to all the data and analysis results, so that visitors can use more sophisticated tools to mine the data (e.g. GSEA). A complete list of the website features is below:

(1) Download normalized data, the transcriptional regulatory network in Cytoscape format, the sample annotation table, and a .GMT file containing the modules for use with Gene Set Enrichment Analysis (GSEA).

(2) View the expression profiles of selected genes.

(3) Explore the genes differentially expressed by pair-wise comparison between individual biological groups (e.g. AGM vs YS) and between HSC states (e.g. Definitive vs Specifying).

(4) Explore each module by viewing module profiles, and by viewing the gene members.

(5) Find the predicted transcriptional regulators of modules.

(6) Determine whether a user-supplied gene list is enriched in any of the modules.

The website's FAQ fully describes how the resource can be used. All genes listed on the website are linked to NCBI's Entrez Gene and MGI, providing a source of up-to-date annotation.

Supplemental References

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Draper, B.W., Morcos, P.A., and Kimmel, C.B. (2001). Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. Genesis *30*, 154-156.

Gekas, C., Dieterlen-Lievre, F., Orkin, S.H., and Mikkola, H.K. (2005). The placenta is a niche for hematopoietic stem cells. Dev Cell *8*, 365-375.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Kyba, M., Perlingeiro, R.C., and Daley, G.Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. Cell *109*, 29-37.

Lawrence, H.J., Helgason, C.D., Sauvageau, G., Fong, S., Izon, D.J., Humphries, R.K., and Largman, C. (1997). Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. Blood *89*, 1922-1930.

Lin, H.F., Traver, D., Zhu, H., Dooley, K., Paw, B.H., Zon, L.I., and Handin, R.I. (2005). Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood *106*, 3803-3810.

McKinney-Freeman, S.L., Naveiras, O., Yates, F., Loewer, S., Philitas, M., Curran, M., Park, P.J., and Daley, G.Q. (2009). Surface antigen phenotypes of hematopoietic stem cells from embryos and murine embryonic stem cells. Blood *114*, 268-278.

Mitchell, T. (1997). Machine Learning (McGraw Hill).

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257-260.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573-1577.

Shepard, J.L., Amatruda, J.F., Stern, H.M., Subramanian, A., Finkelstein, D., Ziai, J., Finley, K.R., Pfaff, K.L., Hersey, C., Zhou, Y.*, et al.* (2005). A zebrafish bmyb mutation causes genome instability and increased cancer susceptibility. Proc Natl Acad Sci U S A *102*, 13194-13199.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc *3*, 59-69.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.