Cell Stem Cell, Volume *1*

Supplemental Data

Hematopoietic Fingerprints:

An Expression Database

of Stem Cells and Their Progeny

Stuart M. Chambers, Nathan C. Boles, Kuan-Yin K. Lin, Megan P. Tierney, Teresa V. Bowman, Steven B. Bradfute, Alice J. Chen, Akil A. Merchant, Olga Sirin, David C. Weksberg, Mehveen G. Merchant, C. Joseph Fisk, Chad A. Shaw, and Margaret A. Goodell

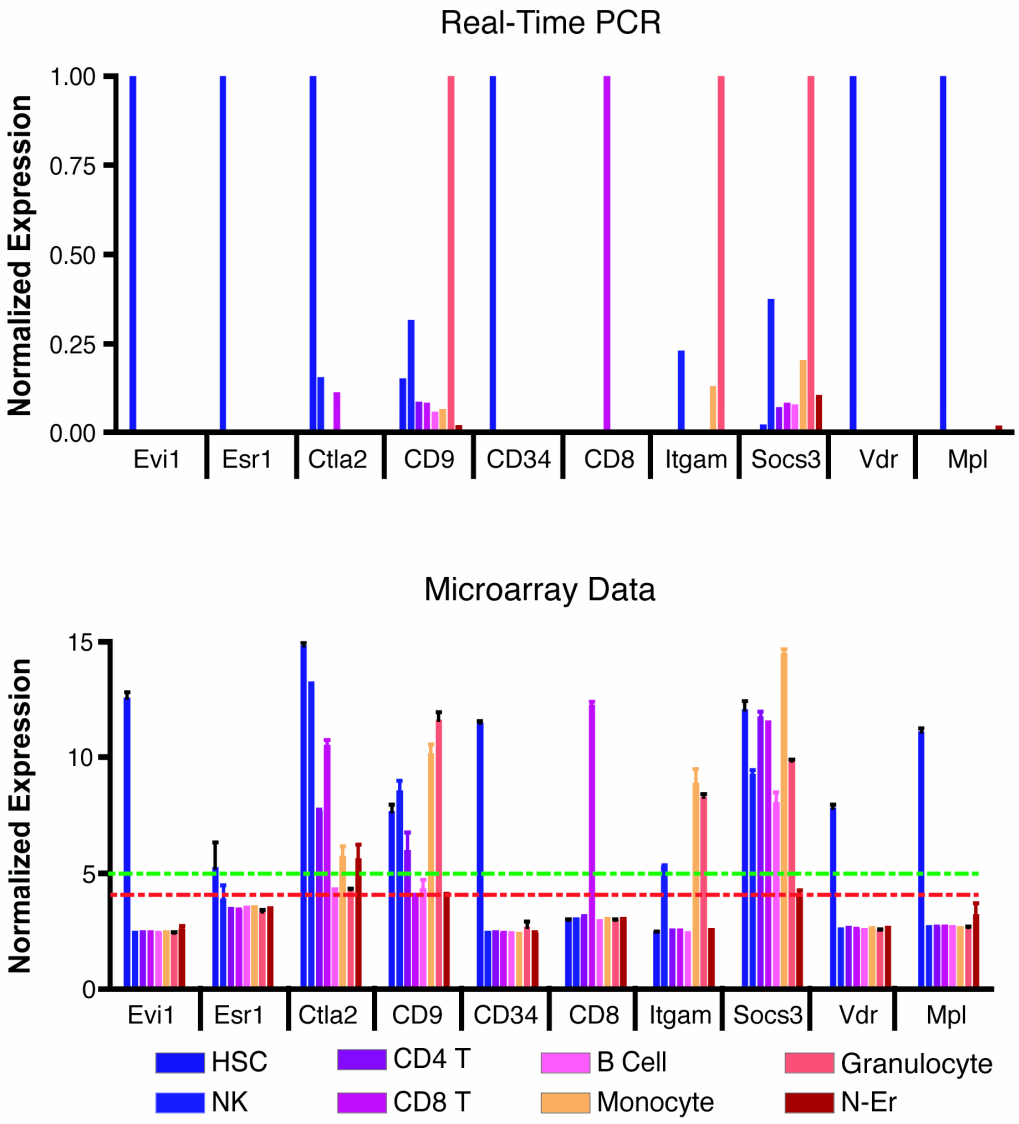

Figure S1. Real-Time qRT-PCR of Genes to Establish Fingerprint Threshold
Genes with both a variable expression and genes that were uniquely expressed were further examined using real-time qRT-PCR. Expression values for selected genes was obtained by real-time PCR and used to calculate thresholds for the microarrays. A microarrray value of greater than 5 for a given cell type and less than 4 for all other cell populations was used to create thresholds by which genes uniquely expressed(fingerprint) in a population were identified. The real-time qRT-PCR results were normalized to a value of 1 for the cell type with the greatest expression. In wells where no amplification was detected beyond 55 PCR cycles, a real-time value of 0 was assigned.

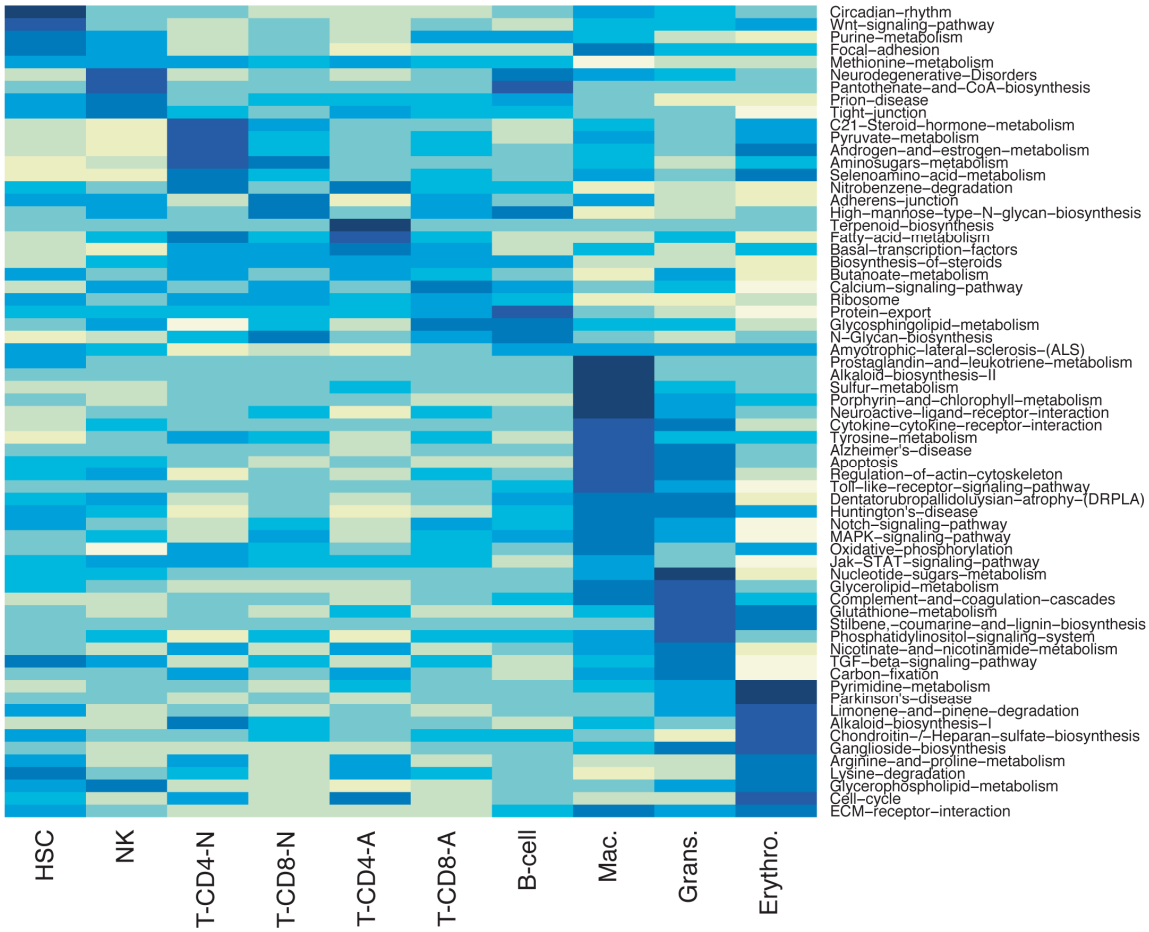

Figure S2. Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis
The KEGG database (http://www.genome.ad.jp/kegg/) is a suite of databases systematically organized into metabolic and signaling pathways. The mean expression value for each hematopoietic cell type was obtained for all KEGG pathways that contained genes found on the microarray. Pathways with a significant variation between the cell types (ANOVA p-value ≤ .05) were identified (65 significant pathways). The pathways were then ranked by maximal abundance for each cell type and a heat map was generated using the centered data.

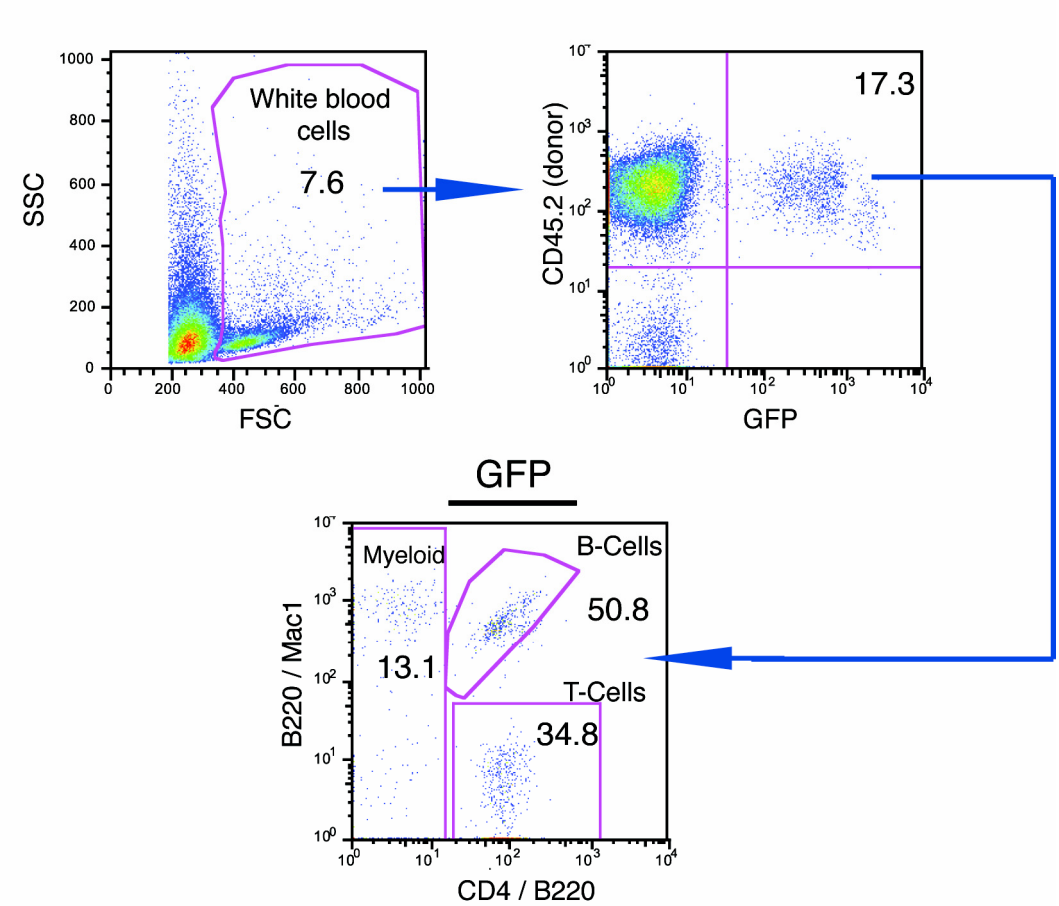

Figure S3. Retroviral Transduction Lineage Analysis
At twelve weeks after transplant/transduction mice were bled and engrafted/transduced cells were identified using CD45.2-APC (donor background) and eGFP expression. B-cells were dual stained with B220-Pacific Blue and B220-PE-Cy7. T-cells were single lablled with CD4- and CD8-Pacific Blue. Myeloid cells were Pacific Blue negative. Transduced and non-transduced cells were simultaneously examined using a FacsAria (BD).

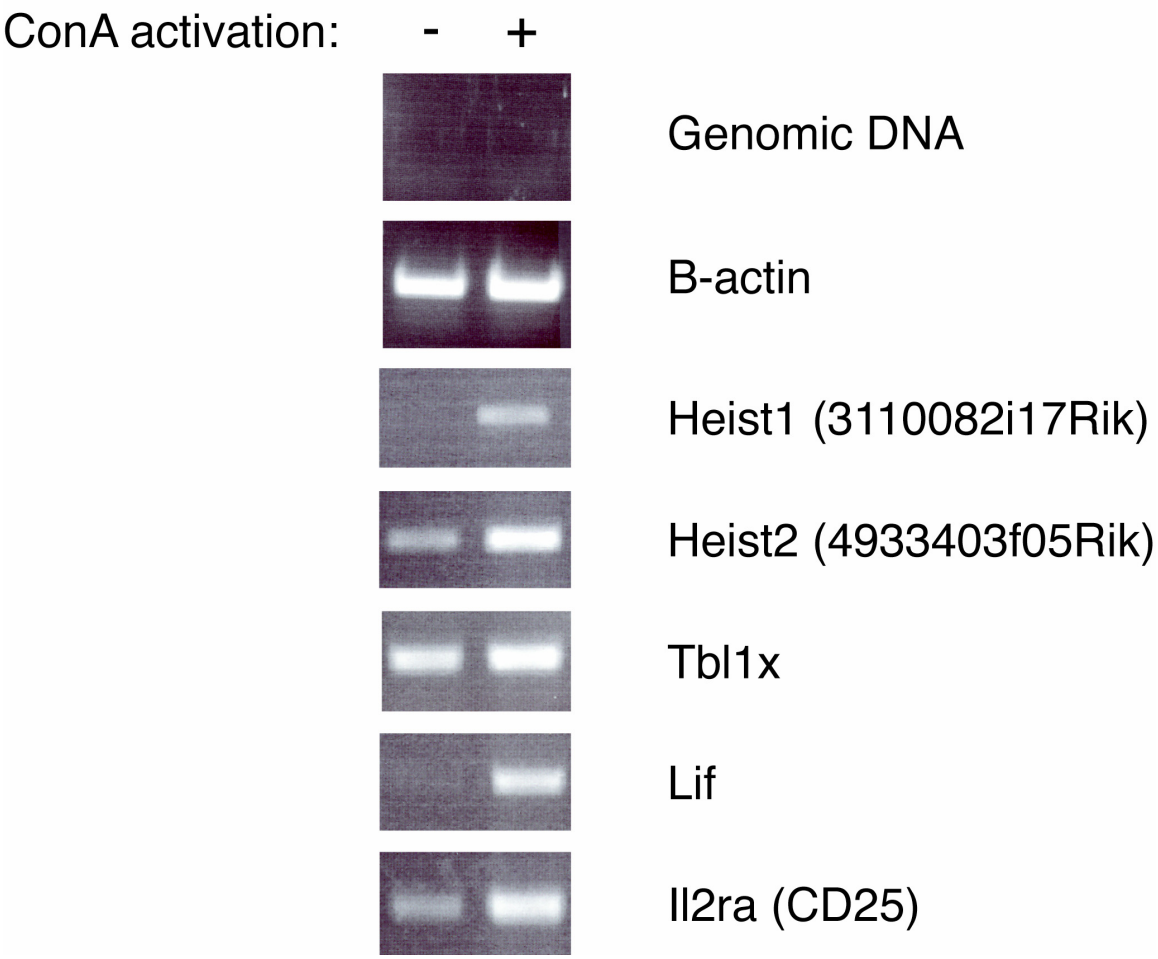

Figure S4. Genes Upregulated by ConA
T-cells were plated in the absence or presence of ConA, a potent activator of naive T-cells. After 24 hours, RNA was isolated and an RT-PCR for genes upregulated according to the microarray results was performed. Genes were selected across a range of differential expression to validate the array data. For all primer pairs that yielded an amplicon, the gene was found to be correctly upregulated.

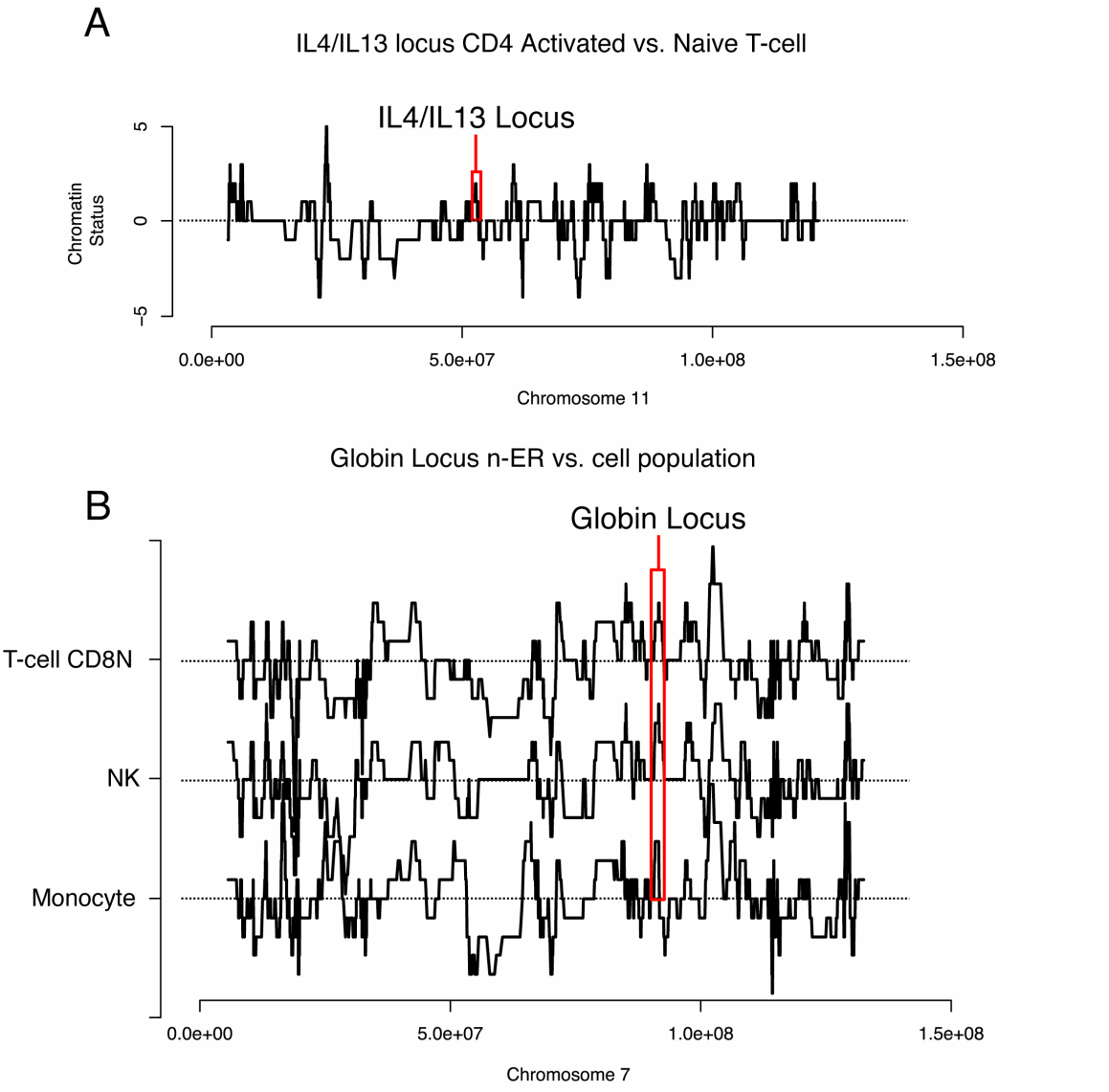

Figure S5. Two Example Loci Detailing Chromosomal Expression Density
(A) IL4/IL13 locus comparing naive and activated T-cells. A positive value (peak above the dashed line), indicates higher gene expression density in activated T-cells compared to naive T-cells. The chromatin within the IL4/IL13 locus opens and the becomes transcriptionally active during T-cell activation(Cai, 2006 and Abni, 2002). A greater density of transcription within this locus is found in activated T-cells compared to naive T-cells. The red box indicates the position of the Il4/Il13 locus. (B) A similar increase in expression density can be found in the globin locus of nucleated erythrocytes when compared to naive CD8 T-cells, NK cells, and monocytes. A positive value (peak above the dashed line), indicates higher gene expression density in erythrocytes when compared to naive CD8 T-cells, NK cells, or monocytes. The red box indicates the position of the globin locus.

Table S1. Purification of Cell Populations Used in This Study

Population	Selection Criteria	Tissue	Array correlation (R^2)
LT-HSC	SP+ c-Kit+ Sca-1+ Lin -	WBM	0.99
NK Cells	Nk1.1+ CD3-	PB	0.99
B-Cells	CD19+ 33D1-	Spleen	0.99
naïve CD4+ T-Cells	CD4+ CD25- CD69-	Spleen	0.98
activated CD4+ T-Cells	Isolated 8-11 Hours After Stimulation CD4+ CD25+ CD69+	Splenocytes treated with ConA	0.99
naïve CD8+ T-Cells	CD8+ CD25- CD69-	Spleen	0.99
activated CD8+ T-Cells	Isolated 8-11 Hours After Stimulation CD4+ CD25+ CD69+	Splenocytes treated with ConA	0.99
Granulocytes	Gr-1+ 7/4 clone+ Lin- (CD2,CD5,B220,F4/80,ICAM-1,Ter-119)	WBM	0.99
Monocytes	Mac-1+ FSC and SSC properties	PB	0.96
Nucleated Erythrocytes	Ter119+ Lin- (CD3, CD4, CD8, Mac1, Gr1, B220)	WBM	0.98

All cells were primary cells freshly isolated using a combination of magnetic and flow cytometric sorting with the indicated markers. Each cell population was isolated from whole bone marrow (WBM), splenocytes, or peripheral blood (PB) of four mice as indicated in the table below. Activated T cells were obtained by stimulation with

conconavalin A for 8-11 hours in vitro. Because of the difficulty in obtaining RNA from terminally differentiated erythroid cells, the erythrocyte population is the only one that contains substantial numbers of cell-type -specific progenitors.

WBM was isolated by flushing the tibias and femurs with Hank's Balanced Salt Solution (HBSS; Invitrogen). Spleens were collected, placed in HBSS, crushed, and filtered in order to obtain splenocytes. PB was collected by retro-orbital bleeds and placed in 2% Dextran T500 in PBS with 10U of heparin. PB was allowed to sit for 30 minutes and the top layer was collected. Cells were then resuspended at 1 x 10^8 cells/ml and stained on ice for 15 minutes with population-specific antibody cocktails followed by an HBSS wash. They were then isolated via a triple-laser instrument (MoFlow, Cytomation). All antibodies were obtained from BD Pharmingen unless otherwise stated. The correlation coefficient (R^2) that indicates the similarity between the two arrays for each population is indicated in the rightmost column.

Cells were purified from 8-12 week-old female mice; each population was purified on two separate occasions from pools of tissue from at least 4 mice (biological replicates). RNA from all samples were processed together and were amplified from approximately the same number of cells prior to hybridization to Affymetrix MOE430 2.0 microarrays, which include ~45,000 probe sets representing about two-thirds of the coding genome. Nucleated erythrocytes from WBM were Ter-119$^+$, CD3$^-$,CD4$^-$,CD8$^-$,Mac-1$^-$,Gr-1$^-$, and B220$^-$. Granulocytes (from WBM) were Gr-1$^+$, clone 7/4$^+$ (Cedarlane Labs), CD2$^-$, CD5$^-$, B220$^-$, F4/80$^-$ (eBiosciences), ICAM-1$^-$, Ter-119$^-$. LT-HSCs were isolated as shown previously (Camargo et al.). Briefly, WBM was stained with Hoescht 33342 and the Sca-1+ cells were enriched by magnetic separation, followed by flow cytometry for side-population (SP) and Sca-1$^+$, c-Kit$^+$, and Lin$^-$ (Mac-1, Gr-1, Ter119, B220, CD4, CD8 (eBiosciences). Naïve T-cells were freshly isolated from spleen as CD4$^+$, CD25$^-$, CD69$^-$ or CD8$^+$, CD25$^-$, and CD69$^-$. Activated T-cells were isolated by enriched naïve T-cells with Concanavalin A (1 µg/ml, Sigma) for eight to eleven hours followed by sorting for CD25$^+$ and CD69$^+$. B-cells were CD19$^+$ and 33D1$^-$ splenocytes. Monocytes were isolated from PB based on Forward and Side Scatter properties as well as being Mac-1$^+$. NK cells from spleen were Nk1.1$^+$ CD3$^-$ cells.

Tables S2 and S3

See the accompanying Excel documents.

Table S4. The Frequency of Fingerprint Genes Was Compared to the Frequency Found by Chance within the Hematopoietic Knockout Data

Fingerprint List	Z-statistic
HSC	17.873
T-Cell, Naïve	50.912
Granulocytes	3.483
Monocytes	29.448
n-Er	11.673
NK Cell	26.178
B-cell	43.222
Lymphoid	12.973
Myeloid	34.583
Differentiatied	57.531
Pooled Fingerprints	24.696
Hematopoietic	-0.922

A Z-statistic was used to determine enrichment significance, where the larger the enrichment the more positive the result. A negative value for the Z-statistic indicates a depletion in hematopoietic specific knockouts.

Table S5
See the accompanying Excel document.