Stem Cell Reports, Volume *5*

Supplemental Information

The Genetic Landscape of Hematopoietic Stem Cell Frequency in Mice

Xiaoying Zhou, Amanda L. Crow, Jaana Hartiala, Tassja J. Spindler, Anatole Ghazalpour, Lora W. Barsky, Brian B. Bennett, Brian W. Parks, Eleazar Eskin, Rajan Jain, Jonathan A. Epstein, Aldons J. Lusis, Gregor B. Adams, and Hooman Allayee

Supplemental Table 1. HMDP Mouse Strains Used in the Present Study Ordered According to LSK Frequency, Related to Figure 1.

1. BXD32/TyJ	29. BXD45/RwwJ	57. C57BL/6J	85. BXD56/RwwJ
2. AXB18/PgnJ	30. BXD42/TyJ	58. BXA7/PgnJ	86. KK/HlJ
3. PL/J	31. BXD34/TyJ	59. DBA/2J	87. BXD16/TyJ
4. BXD14/TyJ	32. BXD15/TyJ	60. BXD64/RwwJ	**88. BXHA1**
5. BUB/BnJ	33. BXD12/TyJ	61. AXB5/PgnJ	89. BXH2/TyJ
6. BXD70/RwwJ	34. AXB13/PgnJ	62. BXD43/RwwJ	**90. CE/J**
7. AXB8/PgnJ	35. BXD50/RwwJ	63. BXD75/RwwJ	**91. AXB6/PgnJ**
8. BXD55/RwwJ	36. BXA11/PgnJ	64. BXD40/TyJ	**92. NOD/LtJ**
9. CXB12/HiAJ	37. BXD39/TyJ	65. BXD84/RwwJ	**93. C3H/HeJ**
10. BXD49/RwwJ	38. AXB19/PgnJ	66. BXD66/RwwJ	**94. BXA8/PgnJ**
11. BXD18/TyJ	39. BXA2/PgnJ	67. 129X1/SvJ	**95. BXH8/TyJ**
12. BXD29/TyJ	40. BXA12/PgnJ	68. CXB8/HiAJ	**96. BXH19/TyJ**
13. BXD11/TyJ	41. BXD5/TyJ	69. C58/J	**97. SEA/GnJ**
14. BXD62/RwwJ	42. AXB12/PgnJ	70. C57BLKS/J	**98. CXB3/ByJ**
15. BXD73/RwwJ	43. BXA1/PgnJ	71. SM/J	**99. CBA/J**
16. BXA13/PgnJ	44. BXA4/PgnJ	72. CXB13/HiAJ	**100. CXB11/HiAJ**
17. AKR/J	45. MA/MyJ	73. BXD86/RwwJ	**101. RIIIS/J**
18. BXD60/RwwJ	46. CXB7/ByJ	74. BXA26/PgnJ	**102. NZB/BlNJ**
19. BXD24/TyJ	47. BXH6/TyJ	75. BXD20/TyJ	**103. BALB/cJ**
20. BXD1/TyJ	48. BXD85/RwwJ	76. BXD61/RwwJ	**104. CXB6/ByJ**
21. BXH10/TyJ	49. BXD68/RwwJ	77. BXD21/TyJ	**105. CXB9/HiAJ**
22. LG/J	50. C57L/J	78. FVB/NJ	**106. NON/LtJ**
23. BXH14/TyJ	51. BXD31/TyJ	79. BXHB2	**107. A/J**
24. BXD38/TyJ	52. BXD48/RwwJ	80. BXD71/RwwJ	**108. I/LnJ**
25. SJL/J	53. BTBRT<+>tf/J	81. CXB1/ByJ	
26. BXA16/PgnJ	54. BXH4/TyJ	**82. BXH9/TyJ**	
27. AXB15/PgnJ	55. SWR/J	83. BXD44/RwwJ	
28. BXD74/RwwJ	56. NZW/LacJ	**84. BXA14/PgnJ**	

Strains are listed from highest to lowest LSK frequency and those carrying the low Sca-1-expressing haplotype (Spangrude & Brooks, 1993), which were excluded from GWAS analyses for LSK and LSKCD150$^-$CD48$^-$ HSPCs, are shown in bold. The 29 classic inbred strains are highlighted in blue.

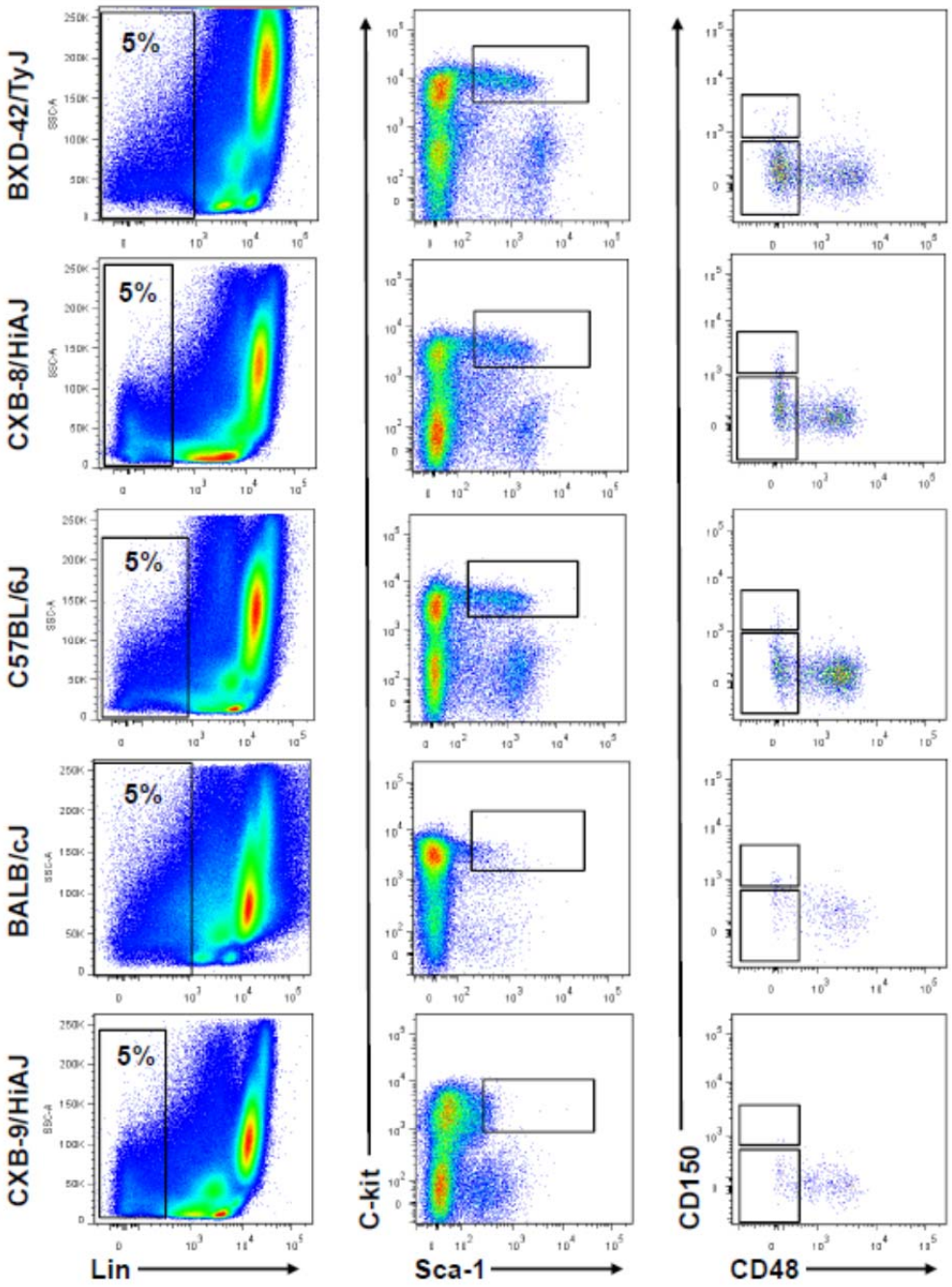
Supplemental Table 2. Correlations of Primitive Hematopoietic Cell Frequency with Blood Cell Parameters, Related to Figure 1.

		WBC	Ly %	Mo %	Gr %	RBC	MCV	HCT	MCH	MCHC	RDW%	RDWa	MPV	HGB	PLT
LSK	r	**0.17**	**0.14**	**0.24**	0.10	0.02	0.028	-0.053	-0.040	-0.050	0.033	0.030	0.066	0.037	0.04
	p-value	**0.005**	**0.04**	**<0.0001**	0.07	0.75	0.657	0.403	0.525	0.425	0.602	0.633	0.295	0.558	0.43
LSKCD150^{-}CD48^{-}	r	0.0	**-0.18**	**-0.20**	**0.24**	-0.05	0.035	-0.016	-0.002	-0.033	0.042	0.064	-0.009	-0.062	0.04
	p-value	0.90	**0.006**	**0.002**	**0.0002**	0.38	0.574	0.798	0.981	0.601	0.502	0.310	0.891	0.325	0.42
LSKCD150^{+}CD48^{-}	r	**0.15**	-0.02	-0.03	0.04	0.10	-0.059	0.035	-0.230	-0.063	0.088	-0.019	-0.026	-0.009	0.07
	p-value	**0.02**	0.79	0.69	0.57	0.12	0.352	0.579	**< 0.001**	0.320	0.161	0.764	0.687	0.891	0.25

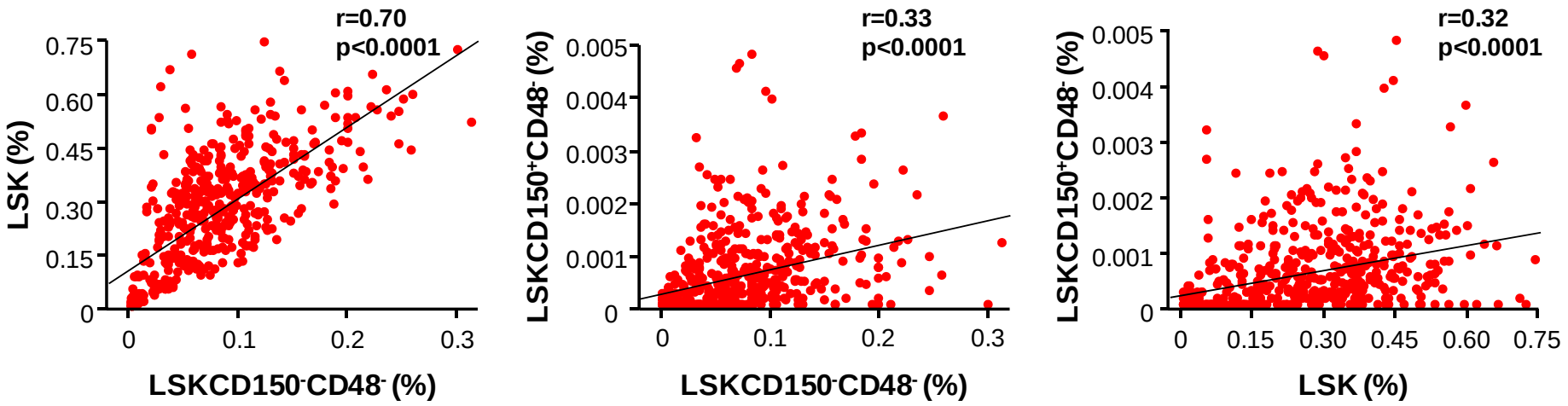
WBC: white blood cell count; Ly %: percent lymphocytes; Mo %: percent monocytes; Gr %: percent granulocyte; RBC, red blood cell count; MCV: mean corpuscular volume; HCT: hematocrit; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW%: red cell distribution width, percent; RDWa: red cell distribution width, area; MPV: mean platelet volume; HGB: hemoglobin concentration; PLT: platelets. Significant correlations are shown in bold.

Supplemental Table 3. List of all Antibodies used for Flow Cytometry Analyses, Related to Figure 1.

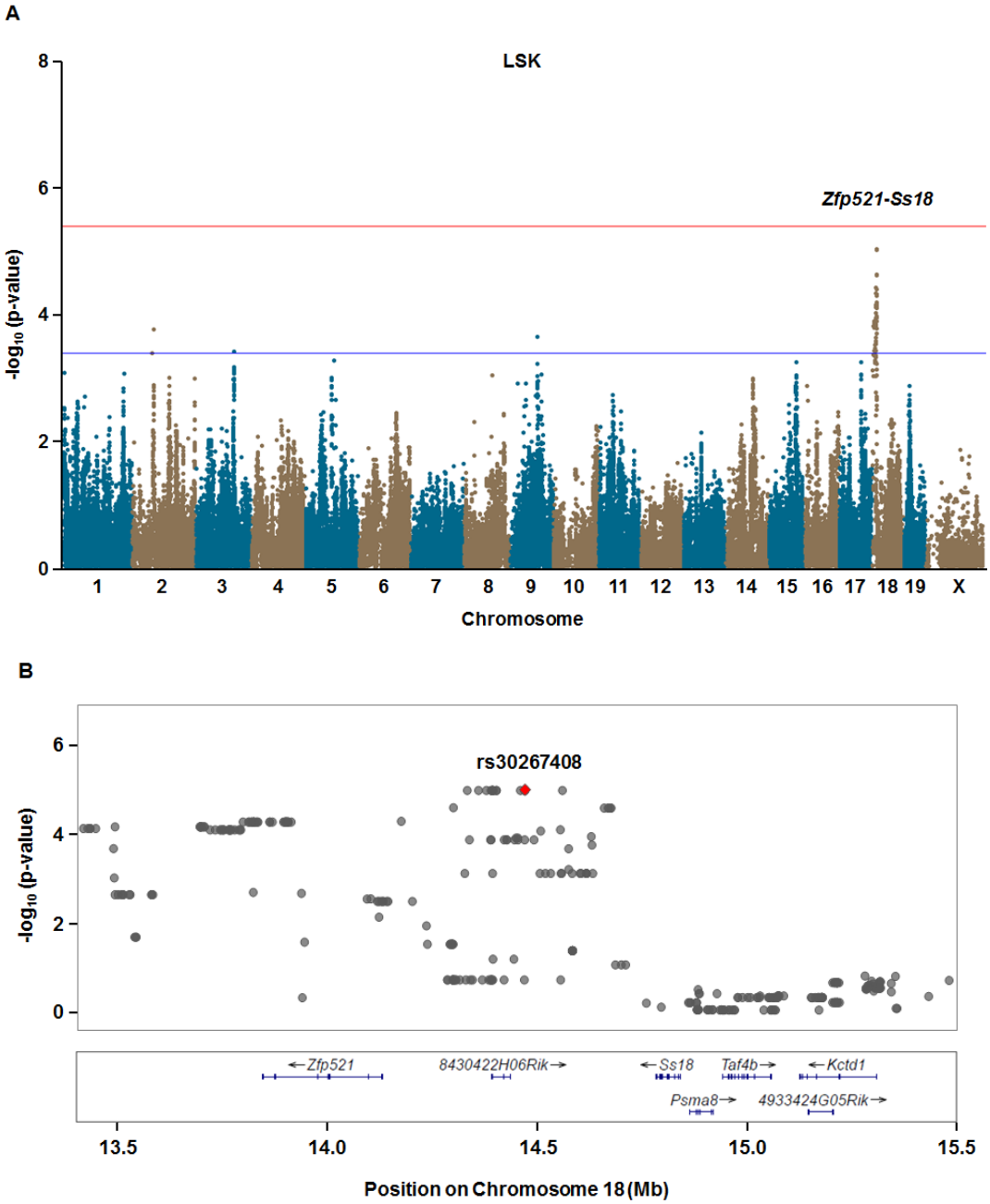
Antibody	Conjugated	Vendor	Catalog number
Lineage Cocktail (CD3e, CD11b, CD45R/B220, TER-119, and Ly6G and Ly-6C)	Streptavidin APC-Cy7	BD Pharmingen	559971
c-kit	APC	eBioscience	17-1172-83
c-kit	PE-Cy7	eBioscience	25-1171-82
Sca-1	PE-Cy7	eBioscience	15-5981-82
Sca-1	AF-700	eBioscience	56-5981-82
Sca-1	FITC	eBioscience	11-5981-82
CD150	APC	eBioscience	17-1501-82
CD150	PE-Cy7	eBioscience	25-1502-82
CD48	FITC	eBioscience	11-0481-82
CD48	PE-Cy5	Biolegend	103420
CD45.1	PE	eBioscience	12-0453-83
CD45.2	FITC	eBioscience	11-0454-85
CD3e	APC	eBioscience	17-0031-83
CD11b	PE-Cy7	eBioscience	25-0112-82
B220	Biotin	eBioscience	13-0452-86



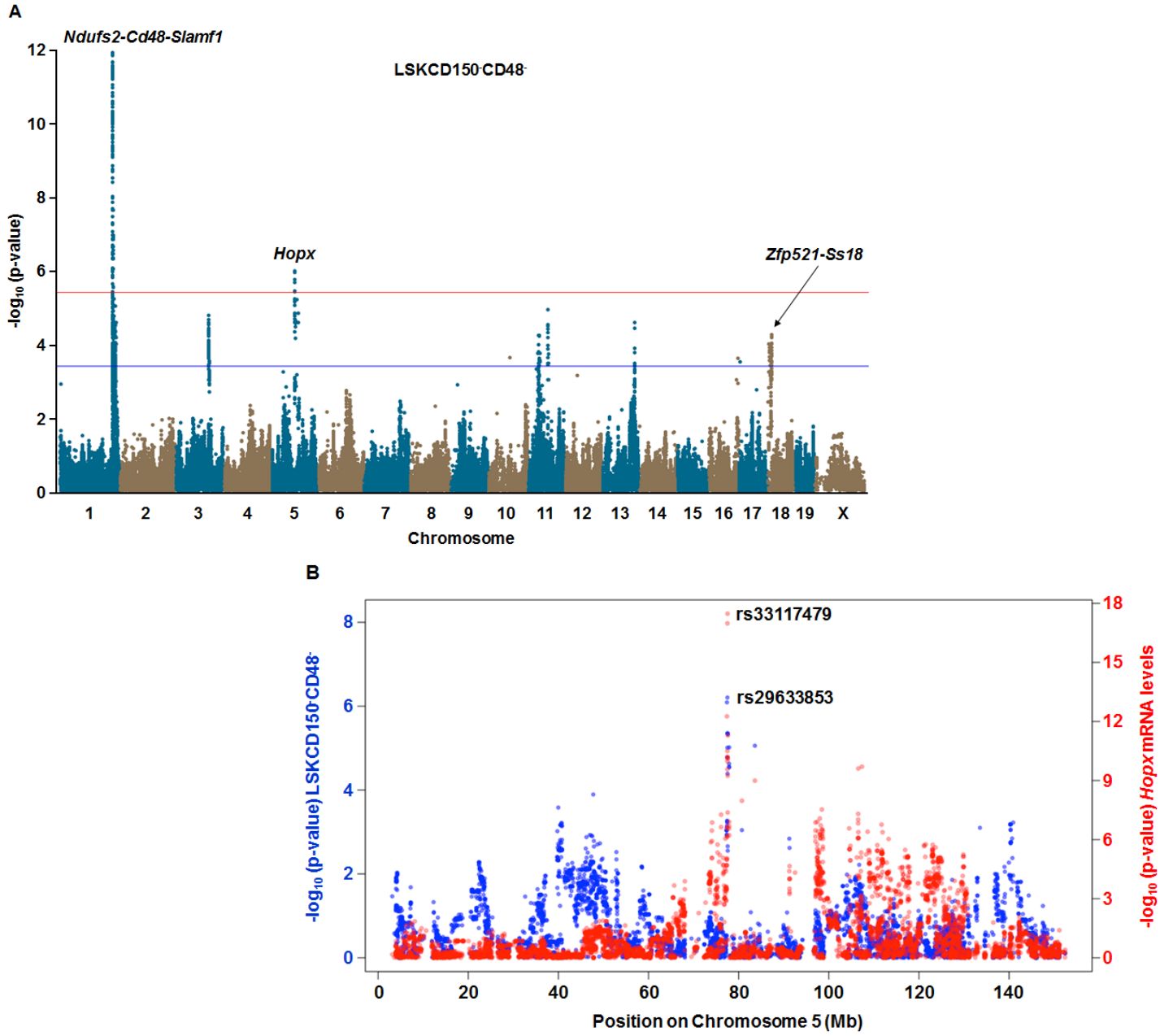
Supplemental Figure 1. Gating strategy for flow cytometric analyses of HSPCs in the HMDP, Related to Figure 1. Representative sequential gating strategy for the Lin$^-$Sca-1$^+$c-Kit$^+$CD150$^+$CD48$^-$ and Lin$^-$Sca-1$^+$c-Kit$^+$CD150$^-$CD48$^-$ cells are demonstrated. Five mouse strains are shown which have among the highest (CXB-8/HiAJ and BXD-42/TyJ) and lowest (CXB-9/HiAJ and BALB/cJ) HSPC frequencies of all strains analyzed. C57BL/6J is also shown for reference. In each case, the Lin$^-$ gate was consistently set as the lowest 5% of BM MNCs.



Supplemental Figure 2. Correlation of three HSPC populations in the HMDP, Related to Figure 1. The frequency of LSK, LSKCD150$^-$CD48$^-$, LSKCD150$^+$CD48$^-$ cells are positively correlated with each other, with a particularly strong association between LSK and LSKCD150$^-$CD48$^-$ cells. Each dot represents an individual mouse from 108 HMDP strains. BM MNCs were isolated from the femurs and tibias of 12-week old male mice (n=3-8 per strain) and the frequency of different HSPC sub-populations was determined by flow cytometry. Data are expressed as a percentage of BM MNCs.

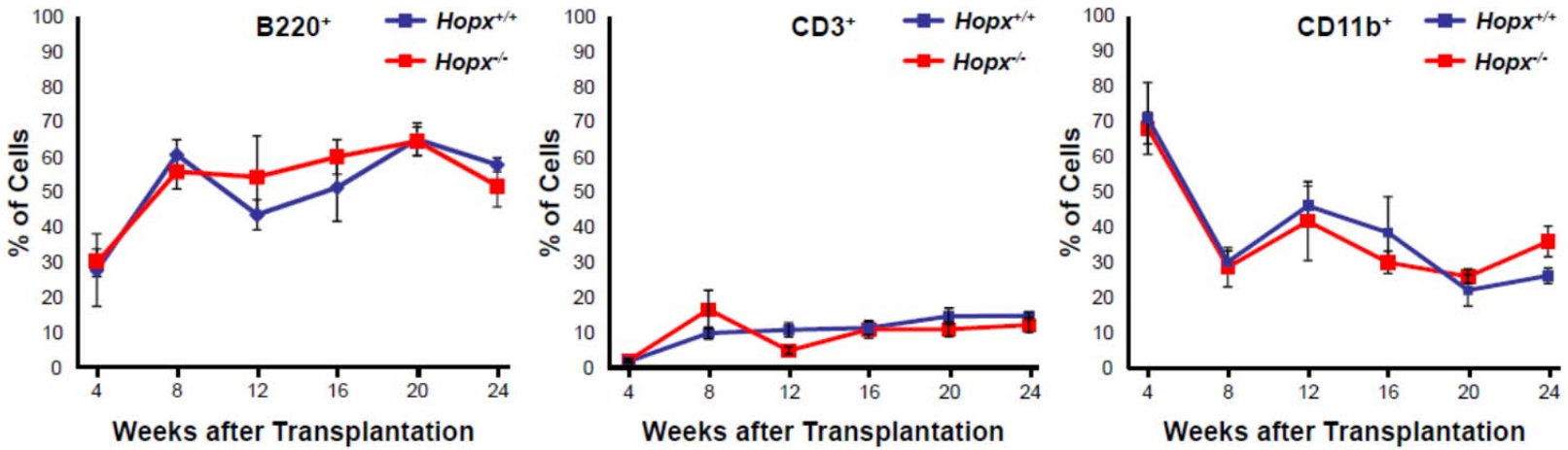


Supplemental Figure 3. GWAS results for LSK frequency after exclusion of strains carrying the uninformative Sca-1-expressing haplotype, Related to Table 1 and Figure 2. **(A)** A Manhattan plot shows no significantly associated locus but a suggestive peak on chromosome 18 harboring *Zfp521* and *Ss18* increased in significance to just below the threshold for genome-wide significance ($p=9.4x10^{-6}$). These analyses included 3-8 mice from 86 strains and 827,406 SNPs, whose genomic positions are shown along the x-axis with their corresponding $-\log_{10}$ p-values indicated by the y-axis. The genome-wide thresholds for significant ($p=4.1x10^{-6}$) and suggestive ($p=4.1x10^{-4}$) evidence of association are indicated by the horizontal red and blue lines, respectively. **(B)** A regional plot of the suggestively associated locus on chromosome 18 shows a 2Mb interval with the peak SNP (rs30267408; indicted by red diamond) mapping to an intergenic region ~363kb distal to *Zfp521* and ~290kb proximal to *Ss18*.



Supplemental Figure 4. GWAS results for LSKCD150$^-$CD48$^-$ frequency after exclusion of strains carrying the uninformative Sca-1-expressing haplotype, Related to Tables 1 and 2 and Figures 2 and 3. (A) A Manhattan plot shows that the association signals on chromosomes 1 ($p=4.4x10^{-12}$) and 18 ($5.7 x10^{-5}$) increased in significance but there was no effect on chromosome 5 ($p=1.1x10^{-6}$). These analyses included 3-8 mice from 86 strains and 827,406 SNPs, whose genomic positions are shown along the x-axis with their corresponding $-\log_{10}$ p-values indicated by the y-axis. The genome-wide thresholds for significant ($p=4.1x10^{-6}$) and suggestive ($p=4.1x10^{-4}$) evidence of association are indicated by the horizontal red and blue lines, respectively. **(B)** A Manhattan plot of the GWAS results on chromosome 5 shows overlap of the association signals for LSKCD150$^-$CD48$^-$ cells and hepatic *Hopx* mRNA expression levels. The lead SNP for the highly significant *cis* eQTL in liver for *Hopx* (rs33117479; $p=3.4x10^{-18}$) maps ~8kb proximal to the lead SNP for LSKCD150$^-$CD48$^-$ cells (rs29633853; $p=1.1x10^{-06}$).

Blue dots correspond to p-values for LSKCD150^{-}CD48^{-} frequency (left y-axis) obtained from the GWAS results for chromosome 5 with 45,050 SNPs and 3-8 mice from 86 strains. Red dots correspond to eQTL p-values for *Hopx* mRNA levels (right y-axis) obtained with 3 mice per strain, as reported previously by Bennett et al (2010).



Supplemental Figure 5. The effect of Hopx deficiency in HSCs on multi-lineage engraftment after competitive repopulation assays, Related to Figure 5. Despite significantly decreased levels of engraftment starting at 16 weeks after transplantation, the relative abundance of circulating B-lymphocytes ($B220^+$), T-lymphocytes ($CD3^+$), or myeloid cells ($CD11b^+$) was not affected after transplantation of HSCs from *Hopx*$^{-/-}$ mice compared to wildtype *Hopx*$^{+/+}$ littermates. Mice were bled from the tail vein every 4 weeks after transplantation and the level of engraftment as a percentage of total circulating MNCs was determined by flow cytometry. Error bars represent s.e.m. with n=3-5 mice per group.

Supplemental Experimental Procedures

Flow cytometry. Femurs and tibias were dissected from the mice and BM MNCs were obtained. To remove any remaining bone fragments or hair, the BM solution was filtered using a 70μm cell strainer (Becton Dickinson). BM MNCs were incubated in PBS with fluorescent labeled anti-mouse CD150, anti-mouse CD48, anti-mouse Sca-1, and anti-mouse c-Kit antibodies (all from eBioscience, San Diego, CA). Concurrently, cells were incubated with the biotinylated lineage cocktail (eBioscience). Red cells were then lysed in 1X FACS lysing solution (Becton Dickinson). The labeled cells were then analyzed by flow cytometry using a LSR II flow cytometer (Becton Dickinson) and HSPC frequencies were calculated according to established cell immunophenotypes using FlowJo flow cytometry analysis software. Lineage-negative (Lin-) cells were consistently defined as the lowest 5% of BM MNCs expressing the lineage cocktail. Gates for other surface markers were standardized and applied across all strains analyzed. **Supplemental Figure 1** illustrates the gating strategy used for the flow cytometry analyses in selected mouse strains that were chosen to represent the range of variation in HSPC frequency. A list of all antibodies used for flow cytometric analyses is shown in **Supplemental Table 3**.

Statistical genetics analysis. Association analyses in the HMDP strains was carried out using SNP genotype data from the Broad Institute (http://www.broadinstitute.org/mouse/hapmap) and the Wellcome Trust Center for Human Genetics (http://mus.well.ox.ac.uk/mouse/). Using these resources and additional variants discovered by the NIEHS/Perlegen mouse resequencing project, we imputed the genotypes of ~4,000,000 SNPs across the genome, with ambiguous genotypes

labeled as "missing." Of these SNPs, 880,924 were informative in the HMDP with a minor allele frequency greater than 5% and used in the present GWAS analyses for HSPC frequency.

We applied the following linear mixed model to account for the population structure and genetic relatedness among strains: $y=\mu+x\beta+u+e$ where μ represents mean HSPC frequency, x represents the SNP effect, u represents random effects due to genetic relatedness with $Var(u) = \sigma g2K$ and $Var(e) = \sigma e2I$, where K represents an identity-by-descent (IBD) kinship matrix across all genotypes. A restricted maximum likelihood (REML) estimate of $\sigma g2$ and $\sigma e2$ were computed using Efficient Mixed Model Association (EMMA) (Kang et al. 2008), and the association mapping was performed based on the estimated variance component with a standard F test to test $\beta \neq 0$. The threshold for genome-wide significance in the HMDP was previously calculations determined by the family-wise error rate (FWER) as the probability of observing one or more false positives across all SNPs per phenotype (Kang et al. 2008; Bennett et al. 2010). These calculations ran 100 different sets of permutation tests and parametric bootstrapping of size 1,000 and observed that the genome-wide significance threshold at a FWER of 0.05 corresponded to a p-value of 4.1×10^{-6}, which has been used in previous studies with the HMDP (Bennett et al. 2010; Farber et al. 2011; Ghazalpour et al. 2011; Park et al. 2011; Bennett et al. 2012; Ghazalpour et al. 2012; Orozco et al. 2012; Davis et al. 2013; Parks et al. 2013; Ghazalpour et al. 2014; Hartiala et al. 2014). This is approximately an order of magnitude larger than the threshold obtained by Bonferroni correction (1.0×10^{-7}), which would be an overly conservative estimate of significance since nearby SNPs among inbred mouse strains are highly correlated with each other.