Stem Cell Reports
Article
OPEN ACCESS

The Genetic Landscape of Hematopoietic Stem Cell Frequency in Mice

Xiaoying Zhou,[1,7] Amanda L. Crow,[2,7] Jaana Hartiala,[2] Tassja J. Spindler,[1] Anatole Ghazalpour,[3] Lora W. Barsky,[1] Brian J. Bennett,[4] Brian W. Parks,[3] Eleazar Eskin,[5] Rajan Jain,[6] Jonathan A. Epstein,[6] Aldons J. Lusis,[3] Gregor B. Adams,[1,8,*] and Hooman Allayee[2,8,*]

[1]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[2]Department of Preventive Medicine and Institute for Genetic Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
[3]Departments of Human Genetics, Medicine, and Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA
[4]Department of Genetics and Nutrition Research Institute, University of North Carolina, Chapel Hill, Kannapolis, NC 28081, USA
[5]Department of Computer Science and Inter-Departmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095, USA
[6]Department of Cell and Developmental Biology and Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Co-first author
[8]Co-senior author
*Correspondence: gregorad@med.usc.edu (G.B.A.), hallayee@usc.edu (H.A.)
http://dx.doi.org/10.1016/j.stemcr.2015.05.008

SUMMARY

Prior efforts to identify regulators of hematopoietic stem cell physiology have relied mainly on candidate gene approaches with genetically modified mice. Here we used a genome-wide association study (GWAS) strategy with the hybrid mouse diversity panel to identify the genetic determinants of hematopoietic stem/progenitor cell (HSPC) frequency. Among 108 strains, we observed ~120- to 300-fold variation in three HSPC populations. A GWAS analysis identified several loci that were significantly associated with HSPC frequency, including a locus on chromosome 5 harboring the homeodomain-only protein gene (*Hopx*). *Hopx* previously had been implicated in cardiac development but was not known to influence HSPC biology. Analysis of the HSPC pool in $Hopx^{-/-}$ mice demonstrated significantly reduced cell frequencies and impaired engraftment in competitive repopulation assays, thus providing functional validation of this positional candidate gene. These results demonstrate the power of GWAS in mice to identify genetic determinants of the hematopoietic system.

INTRODUCTION

In the hematopoietic stem cell (HSC) field, the identification of regulators of HSC physiology has relied mainly on candidate gene approaches with genetically modified mouse models (Rossi et al., 2012). However, these studies have not necessarily provided a complete picture of the complex network of signals that govern HSC proliferation, differentiation, and function. Therefore non-biased strategies, such as transcriptome analyses, have been used, but these approaches also sometimes are restricted by the screen itself (Hope et al., 2010; Karlsson et al., 2013). For example, many genes have been shown to be expressed at specific differentiation stages, including the most primitive HSCs (Chambers et al., 2007), but they do not provide information as to the key regulators of HSC physiology.

An alternative unbiased strategy to elucidate the determinants of HSC frequency/function is to leverage naturally occurring variation in a forward genetics approach. In this regard, prior studies have shown that hematopoietic stem/progenitor cell (HSPC) frequency in mice differs as a function of genetic background (de Haan et al., 1997), and attempts to identify the underlying genes have employed linkage analysis in crosses between inbred mouse strains. For example, using the long-term culture-initiating cell (LTC-IC; Müller-Sieburg and Riblet, 1996) or cobblestone area-forming cell (CAFC; de Haan and Van Zant, 1997) assays, quantitative trait loci (QTLs) for HSPC frequency have been identified on chromosomes 1 and 18 among a panel of recombinant inbred (RI) strains derived from C57Bl/6 × DBA/2 (BXD). This strategy also has been used to map loci for dynamic changes in CAFC frequency associated with aging (de Haan and Van Zant, 1999; Geiger et al., 2001). With the development of immunophenotypic markers that could identify functional HSPCs (Morrison and Weissman, 1994), flow cytometric analysis has been used to identify QTLs for HSPC frequency in the BXD RI panel, but this approach has only been successful in aged mice (Henckaerts et al., 2002, 2004). In addition, QTLs for HSPCs distinct from those mapped in the BXD RI panel have been identified using different inbred mouse strains (Morrison et al., 2002; van Os et al., 2006; Jawad et al., 2008), suggesting that the limited variation in crosses between two strains restricts the identification of other genetic factors that play a role in HSPC physiology. Furthermore, the classical QTL approach has inherently

low mapping resolution, since the regions of interest typically span large chromosomal intervals and can contain hundreds to thousands of genes. To overcome these obstacles, a genetical genomics approach has been proposed, where transcriptomics analysis is combined with QTL mapping (de Haan et al., 2002; Bystrykh et al., 2005; Gerrits et al., 2008). However, to date, the incorporation of intermediate molecular traits into linkage analysis has led to the identification of only one gene, latexin, where differential levels of expression influence the pool of HSPCs in the bone marrow (BM) (Liang et al., 2007).

Genome-wide association studies (GWASs) represent another unbiased approach that has proven particularly successful in humans for gene discovery of numerous disease phenotypes (Hindorff et al., 2015). Encouraged by this success, GWASs also have been proposed in mice using a recently developed panel of classic inbred and RI mouse strains, termed the Hybrid Mouse Diversity Panel (HMDP). Our prior studies with the HMDP have identified key associations for a variety of complex traits that are highly relevant to human diseases, which collectively illustrate the power of this approach for gene discovery in mice (Bennett et al., 2010, 2012; Farber et al., 2011; Ghazalpour et al., 2011, 2012, 2014; Park et al., 2011; Orozco et al., 2012; Davis et al., 2013; Parks et al., 2013; Hartiala et al., 2014). In the present study, we used this mouse platform to carry out a comprehensive genetic screen for HSPC frequency in the adult BM. We identified multiple loci associated with the frequencies of three HSPC populations and functionally validated one positional candidate gene that previously had not been known to play a role in the hematopoietic system.

RESULTS

Variation in HSPC Frequencies in the HMDP

While the competitive repopulation assay is the gold standard for measuring HSPC frequency in a given population of cells (Purton and Scadden, 2007), this method is influenced by a number of parameters, including radiation sensitivity of the host, homing and lodgment of the injected cells in the BM microenvironment, and differences in compensatory mechanisms involved in the differentiation of the mature hematopoietic cells. To limit variation due to any one or a combination of these parameters, we chose immunophenotypic analysis to characterize the frequency of primitive HSPC subsets in the adult BM. We obtained BM mononuclear cells (MNCs) from 12-week-old male mice (Table S1) and used flow cytometry to determine the frequency of several HSPC populations. These included Lineage[Lin]$^-$Sca-1$^+$c-Kit$^+$ (LSK) HSPCs, the more immature LSKCD150$^-$CD48$^-$ multipotent progenitors (MPPs), which have been shown to be able to differentiate into all lineages of the hematopoietic system but have limited potential to self-renew, and LSKCD150$^+$CD48$^-$ cells, which are the most primitive long-term HSCs (Kiel et al., 2008; Oguro et al., 2013). In addition, we chose CD150 and CD48 for our analyses, since these *Slam* markers have been validated to identify the different sub-populations of HSPCs across multiple inbred mouse strains (Kiel et al., 2005). Representative plots that illustrate the gating strategy used for these analyses are shown in Figure S1.

Among the 108 HMDP strains screened, the frequency of LSK, LSKCD150$^-$CD48$^-$, and LSKCD150$^+$CD48$^-$ cells varied by approximately 120- to 300-fold (Figure 1). Since all mice were age- and sex-matched and kept under identical environmental conditions, we attributed these differences, at least in part, to naturally occurring genetic variation. This notion was confirmed by calculating the heritability for each of the three HSPC sub-populations, which yielded values of 0.90, 0.92, and 0.70 for LSK, LSKCD150$^-$CD48$^-$, and LSKCD150$^+$CD48$^-$ cells, respectively. We note, however, that these heritability estimates are somewhat higher than what would be typically expected for complex traits in humans, since phenotype measurements in the HMDP are obtained from multiple animals of the same genotype (strain).

Relationship between HSPC Frequencies and Other Hematological Parameters

We next explored the relationship between LSK, LSKCD150$^-$CD48$^-$, and LSKCD150$^+$CD48$^-$ cells and other hematological parameters. The three types of primitive HSPCs were all significantly correlated with each other (Figure S2), with a particularly strong association between LSK and LSKCD150$^-$CD48$^-$ cells ($r = 0.70$; $p < 0.0001$). LSK cells exhibited modestly positive, but significant, correlations with total white blood cell (WBC) count and with the numbers of lymphocytes and monocytes (Table S2). By comparison, LSKCD150$^-$CD48$^-$ cells were negatively correlated with lymphocyte and monocyte counts and positively associated with granulocytes. With the exception of a weakly positive association with WBC count and a negative relationship with mean corpuscular hemoglobin, no correlations were observed with the most primitive LSKCD150$^+$CD48$^-$ cells. Furthermore, no significant correlations were observed between any of the three HSPC populations and other red blood cell (RBC) traits, such as hemoglobin and hematocrit levels (Table S2). These data suggest that variation in LSK and LSKCD150$^-$CD48$^-$ cells and mature WBCs could be controlled, in part, by similar genetic mechanisms, whereas variation in LSKCD150$^+$CD48$^-$ cells as well as RBC parameters may be driven by distinct factors.

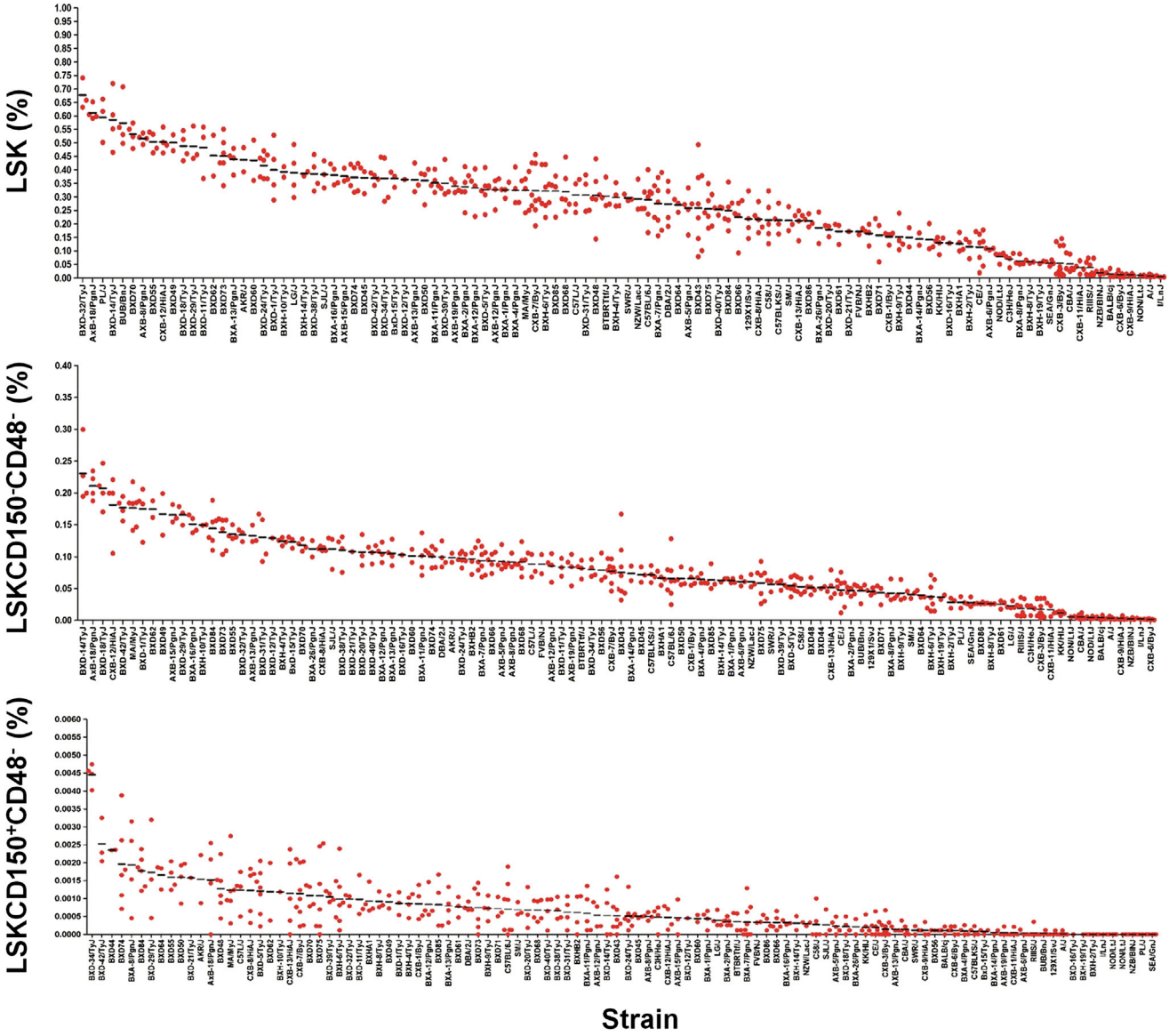


Figure 1. Variation in Three HSPC Populations in the HMDP

The frequency of LSK, LSKCD150^{-}CD48^{-}, and LSKCD150^{+}CD48^{-} cells exhibits 120- to 300-fold variation among 108 HMDP strains. Each dot represents an individual mouse from the respective strain and the mean values are indicated by the horizontal black bars. BM MNCs were isolated from the femurs and tibias of 12-week-old male mice (n = 3–8 per strain; N = 467), and the frequency of different HSPC sub-populations was determined by flow cytometry. Data are expressed as a percentage of BM MNCs. See also Tables S1–S3 and Figures S1 and S2.

GWAS for HSPC Frequencies

To identify the genetic determinants of HSPC frequency, we used the phenotype data to carry out a GWAS for the three cell populations (Figures 2A–2C). One significantly associated locus for LSKCD150^{+}CD48^{-} cells was identified at the distal end of chromosome 18 (Figure 2A; Table 1), where the lead SNP (rs36866074; $p = 3.2 \times 10^{-6}$) mapped to intron 1 of the mitogen-activated protein kinase 4 (*Mapk4*) gene (Figure 3A). On chromosome 11 (Figure 2A; Table 1), we also identified suggestive association with an intergenic region located ~464 kb downstream of the neuromedin-U receptor 2 (*Nmur2*) and ~507 kb upstream of the glutamate receptor 1 isoform 1 precursor (*Gria1*) genes (rs29434264; $p = 2.3 \times 10^{-5}$).

A GWAS for LSK cells revealed several suggestively associated loci, including those on chromosomes 2, 4, 6, and 18,

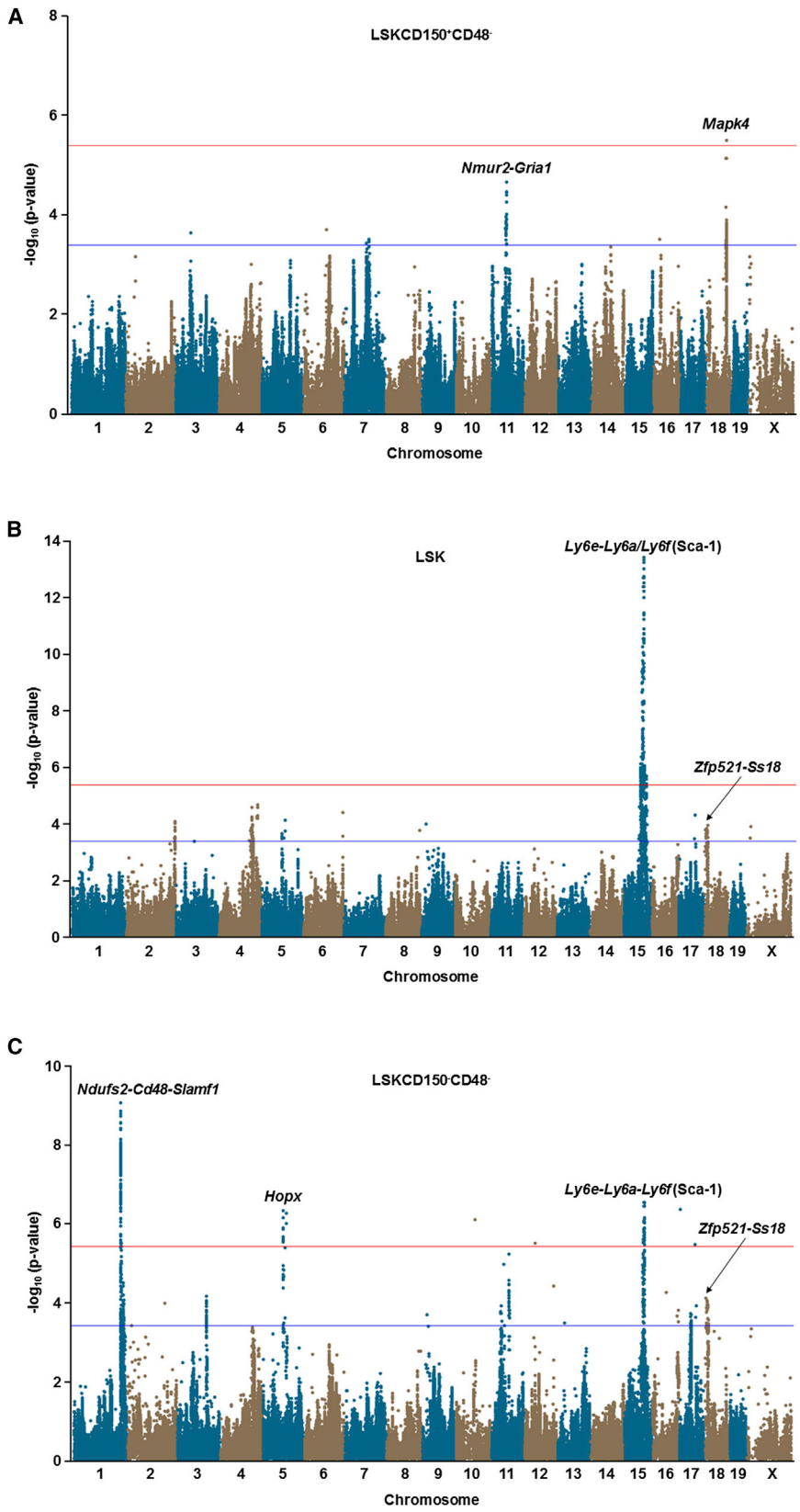


Figure 2. Manhattan Plots of GWAS Results for HSPC Frequency in the HMDP

(A) The frequency of LSKCD150^{+}CD48^{-} cells was significantly associated with a locus on chromosome 18 harboring *Mapk4* and suggestively associated with a region on chromosome 11 near *Nmur2* and *Gria1*.

(B) A GWAS for LSK cells revealed several suggestively associated loci on chromosomes 2, 4, 6, and 18 as well as a highly significant locus on chromosome 15 containing *Ly6e-Ly6a-Ly6f*.

(C) Three significantly associated loci on chromosomes 1, 5, and 15 were identified in the GWAS analysis for LSKCD150^{-}CD48^{-} cells, as well as several suggestively associated regions on chromosomes 3, 11, 17, and 18. For each significantly associated locus, the gene(s) nearest to the peak SNP is indicated. The GWAS analyses for each of the three HSPCs included 3–8 mice per strain (N = 467) and 880,924 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 ($p = 4.1 \times 10^{-6}$) and suggestive ($p = 4.1 \times 10^{-4}$) evidence of association are indicated by the horizontal red and blue lines, respectively.

See also Figure S3.

Table 1. Loci Identified in GWAS for HSPCs in the HMDP

HSPC Population	Chromosome	Position (bp)[a]	Lead SNP	Nearest Gene(s)	Minor Allele Frequency	p Value	p Value[b]
LSKCD150^{+}CD48^{-}	11	56318423	rs29434264	*Nmur2-Gria1*	0.06	2.3×10^{-5}	NA
	18	74202262	rs36866074	*Mapk4*	0.13	3.2×10^{-6}	NA
LSK	15	75203270	rs31675052	*Ly6e-Ly6a-Ly6f*	0.47	3.7×10^{-14}	NA
	18	14494409	rs30267408	*Zfp521-Ss18*	0.41	4.3×10^{-4}	9.4×10^{-6}
LSKCD150^{-}CD48^{-}	1	173168046	rs8242728	*Ndufs2-Cd48-Slamf1*	0.50	9.3×10^{-10}	4.4×10^{-12}
	5	77528483	rs29633853	*Hopx*	0.15	5.1×10^{-7}	1.1×10^{-6}
	15	74669094	rs32350275	*Ly6e-Ly6a-Ly6f*	0.40	3.0×10^{-7}	NA
	18	14474199	rs31073841	*Zfp521-Ss18*	0.11	3.8×10^{-4}	5.7×10^{-5}

NA, not applicable.
[a]Base pair position of lead SNP given according to NCBI build 37 of the reference mouse genome sequence.
[b]p value obtained from GWAS analysis that excluded strains carrying the uninformative haplotype of Sca-1 at the chromosome 15 locus. See also Figures S3 and S4.

as well as a highly significant ($p = 3.7 \times 10^{-14}$) locus on chromosome 15 (Figure 2B; Table 1), where the lead SNP (rs31675052) is located approximately 100 kb downstream of *Ly6f* (Figure 3B). *Ly6f* is part of a family of genes located at this locus that encode Sca-1, which is one of the surface markers used to immunophenotypically quantitate HSPC frequency. While Sca-1 is known to play a role in the function of HSPCs (Ito et al., 2003), some studies have suggested that it is not an informative cell surface marker for flow cytometry analysis in certain mouse strains (Spangrude and Brooks, 1993). To address this potential issue and remove the effect of the chromosome 15 locus, we re-performed the GWAS analysis after excluding strains carrying the low Sca-1-expressing haplotype (Table S1). Importantly, exclusion of these strains did not appreciably decrease the heritability for variation in LSK cells (0.90 versus 0.82). Furthermore, the Sca-1 locus did not yield an association signal in this analysis, as expected, but the suggestive peak on chromosome 18 increased in significance from $p = 4.3 \times 10^{-4}$ to just below the threshold for genome-wide significance with $p = 9.4 \times 10^{-6}$ (Table 1; Figure S3A). The peak SNP (rs30267408) on chromosome 18 is not located within a known gene, but it maps ~363 kb distal to the zinc-finger protein 521 (*Zfp521*) and ~290 kb proximal to the synovial sarcoma-associated Ss18-alpha (*Ss18*) genes (Figure S3B).

We next performed a GWAS for LSKCD150^{-}CD48^{-} cells and identified three significantly associated loci on chromosomes 1 (rs8242728; $p = 9.3 \times 10^{-10}$), 5 (rs29633853; $p = 5.1 \times 10^{-7}$), and 15 (rs32350275; $p = 3.0 \times 10^{-7}$), as well as several suggestively associated loci on chromosomes 3, 11, 17, and 18 (Figure 2C; Table 1). Since the chromosome 15 locus for LSKCD150^{-}CD48^{-} cells was the same as that identified for LSKs, we re-performed the GWAS after excluding the same strains carrying the low Sca-1-expressing haplotype. Similar to LSK cells, excluding these strains only had a marginal effect on the heritability of LSKCD150^{-}CD48^{-} cells (0.92 versus 0.88). This analysis strengthened the association signal on chromosome 1 ($p = 4.4 \times 10^{-12}$), but had no effect on chromosome 5 ($p = 1.1 \times 10^{-6}$) nor revealed any additional regions associated with LSKCD150^{-}CD48^{-} cells (Table 1; Figure S4A). The lead SNP on chromosome 1 (rs8242728) fell within a large ~2-Mb linkage disequilibrium (LD) block containing dozens of genes and numerous SNPs that yielded equivalently significant p values (Figure 3C). Given this LD pattern, it would be difficult to narrow down the underlying causative gene without extensive follow-up. For example, rs8242728 is specifically located within the NADH dehydrogenase [ubiquinone] iron-sulfur protein 2 (*Ndufs2*) gene. However, this interval also encompasses the *Slam* locus and contains the *Cd48* and *Slamf1* genes, which encode two of the other flow cytometry markers used for quantitating HSPCs (CD48 and CD150). By comparison, the most significantly associated SNPs on chromosome 5 are located within a small LD block, with the peak SNP (rs29633853) localizing to the homeodomain-only protein (*Hopx*) gene (Figure 3D). Interestingly, the suggestive association with the peak SNP on chromosome 18 (rs31073841), which maps ~20 kb proximal to the peak SNP of the same *Zfp521-Ss18* locus identified for LSKs, also increased in significance by nearly one order of magnitude from $p = 3.8 \times 10^{-4}$ to $p = 5.7 \times 10^{-5}$ after exclusion of the strains carrying the low Sca-1-expressing haplotype (Table 1; Figure S4A). This observation suggests

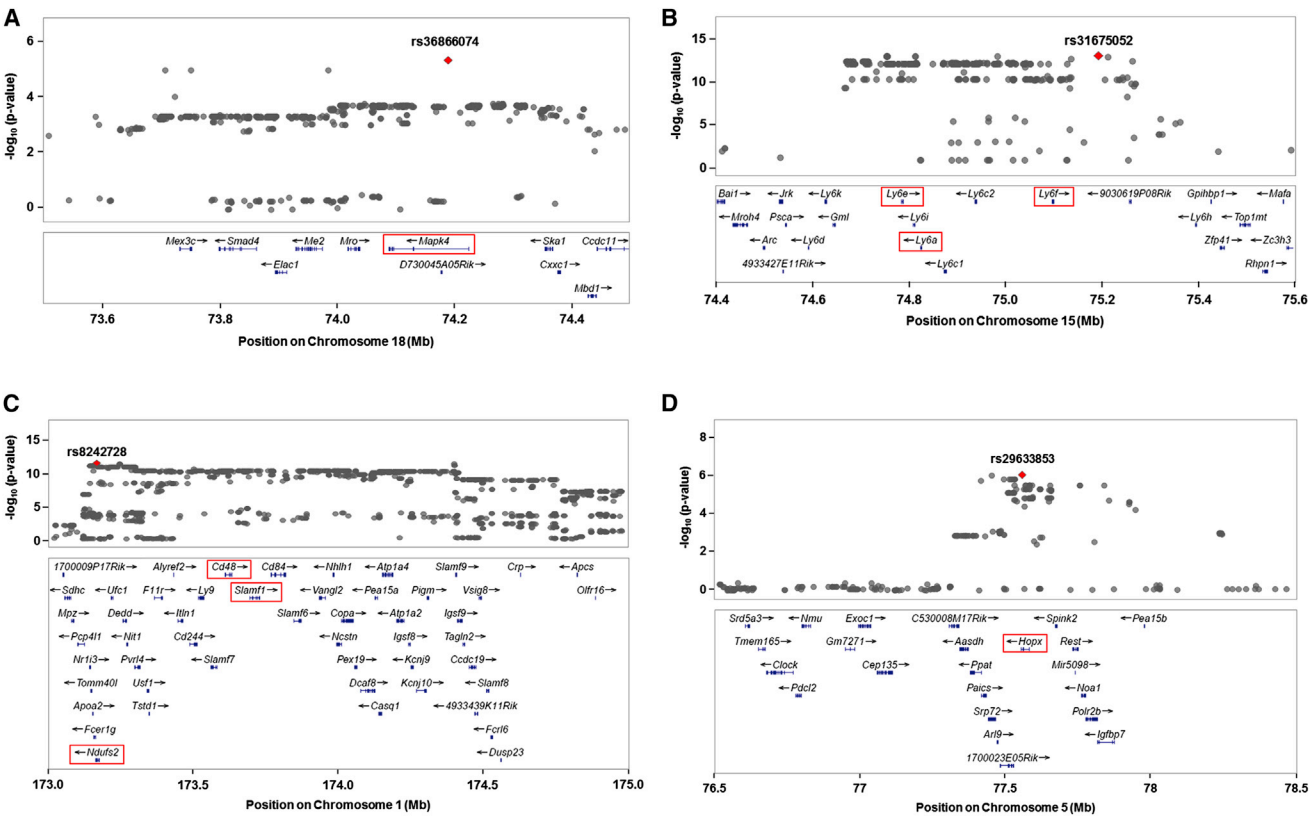


Figure 3. Regional Plots of Loci Significantly Associated with HSPCs in the HMDP

(A) The lead SNP at the chromosome 18 locus for LSKCD150^{+}CD48^{-} cells (rs36866074) maps to intron 1 of *Mapk4* (boxed in red).

(B) The lead SNP on chromosome 15 for LSKs (rs31675052) maps to a region harboring *Ly6e*, *Ly6a*, and *Ly6f* (boxed in red), which are part of a family of genes at this locus that encode one of the surface markers used to immunophenotypically quantitate HSPC frequency (Sca-1).

(C) The chromosome 1 locus identified for LSKCD150^{-}CD48^{-} cells encompasses a large ~2-Mb LD block containing dozens of genes and numerous SNPs that yielded equivalently significant p values. Although the lead SNP (rs8242728) is specifically located in *Ndufs2* (boxed in red), this interval also includes *Cd48* and *Slamf1* (boxed in red), which encode two of other flow cytometry markers used for quantitating HSPCs (CD48 and CD150).

(D) The peak SNP at the chromosome 5 locus (rs29633853) associated with LSKCD150^{-}CD48^{-} cells is located within a small LD block and localizes to *Hopx* (boxed in red). For each plot, a 1–2-Mb region is shown and the lead SNP is indicated by a red diamond.

(A–D) (Top) SNP positions are shown on the x axis with their corresponding $-\log_{10}$ p values indicated by the y axis. (Bottom) The locations of genes in the selected intervals are given. See also Figure S4.

that this chromosome 18 locus may exert pleiotropic effects on multiple HSPC subsets.

Prioritization of Positional Candidate Genes at Chromosome 5 Locus for LSKCD150^{-}CD48^{-} Frequency

Of the loci identified, we focused on the chromosome 5 locus because of the increased mapping resolution and the possibility that this region contains a previously unknown gene(s) for LSKCD150^{-}CD48^{-} cells. In particular, rs29633853 is located within intron 1 of *Hopx*, which has been shown to play important roles in cardiac development (Chen et al., 2002; Shin et al., 2002) and lymphoid regulation (Albrecht et al., 2010) and to serve as a marker for quiescent intestinal epithelial stem cells (Takeda et al., 2011) and multipotent hair follicle stem cells (Takeda et al., 2013). However, *Hopx* has otherwise not been implicated in the biological regulation of HSPCs. To evaluate *Hopx* as a positional candidate, we first searched the publicly available genomic sequences of ~60 inbred mouse strains, but we did not identify any amino acid substitutions and/or other protein structure-altering variations. We next used gene expression data across multiple tissues in the HMDP (http://geneeqtl.genetics.ucla.edu/) to determine whether the chromosome 5 locus exhibited any *cis* expression QTL (eQTL), defined as those mapping within a 2-Mb interval centered on the lead GWAS SNP (rs29633853). Of the ~20 genes in this region (Figure 3D),

Table 2. Genes Exhibiting *cis* eQTLs across Multiple Tissues at the Chromosome 5 Locus for LSKCD150^{-}CD48^{-} Cells

Gene	Gene Position (bp)[a]	SNP	SNP Position (bp)[a]	GWAS p Value[b]	eQTL p Value			
					Bone	Macrophage	Liver	Heart
Srd5a3	76569296-76584529	rs13478330	75857166	0.26	5.8×10^{-7}	ns	ns	ns
		rs3153753	76574923	0.99	4.0×10^{-15}	ns	ns	ns
		rs4225290	76583956	0.76	ns	ns	ns	5.0×10^{-8}
		rs33559924	76254389	0.47	ns	ns	9.0×10^{-12}	ns
		rs3697530	76575062	0.82	ns	ns	3.9×10^{-16}	ns
		rs13478335	76955455	0.41	ns	2.0×10^{-17}	ns	ns
Tmem165	76612905-76638270	rs4225290	76583956	0.76	ns	1.5×10^{-11}	ns	ns
Exoc1	76958336-76999319	rs4225290	76583956	0.76	9.4×10^{-10}	ns	ns	ns
		rs13478334	76817411	0.64	ns	9.4×10^{-12}	ns	ns
		rs33139946	76913636	0.64	ns	9.4×10^{-12}	ns	ns
		rs31555016	76915601	0.64	ns	9.4×10^{-12}	ns	ns
		rs31555860	76921199	0.64	ns	9.4×10^{-12}	ns	ns
		rs29561143	77013954	0.86	ns	ns	5.2×10^{-7}	ns
Cep135	77075352-77075401	rs29824030	75977692	0.36	6.2×10^{-8}	ns	ns	ns
		rs13478331	76154830	0.38	1.7×10^{-9}	ns	ns	ns
Hopx	77516011-77544181	rs33077631	77509306	6.1×10^{-6}	4.0×10^{-8}	ns	ns	ns
		rs33175967	77511463	2.6×10^{-5}	ns	ns	ns	6.2×10^{-19}
		rs33750358	77534348	6.1×10^{-6}	4.0×10^{-8}	ns	ns	ns
		rs33117479	77536186	4.8×10^{-5}	ns	1.6×10^{-6}	3.4×10^{-18}	ns

A genomic region ±1 Mb around the peak SNP for LSKCD150^{-}CD48^{-} cells on chromosome 5 (rs29633853; bp position 77528483) was interrogated for the presence of *cis* eQTLs in multiple tissues using the UCLA Systems Genetics Resource (http://systems.genetics.ucla.edu/). Only genes exhibiting *cis* eQTLs in bone, macrophage, liver, or heart are listed. Ns, not significant.

[a]Base pair (bp) positions of genes and SNPs are given according to NCBI build 37 of the reference mouse genome sequence.

[b]p values are given from GWAS results that included all HMDP strains. See also Figure S4.

Srd5a3, *Tmem165*, *Exoc1*, *Cep135*, and *Hopx* exhibited *cis* eQTLs in either bone, macrophages, liver, or heart (Table 2). However, only variants in *Hopx* yielded both significant associations with LSKCD150^{-}CD48^{-} cells and *cis* eQTLs across all four tissues examined. In particular, the lead eQTL SNP in liver (rs33117479; $p = 3.4 \times 10^{-18}$) was located ~8 kb proximal to the lead SNP for LSKCD150^{-}CD48^{-} cells (rs29633853), whereas the lead eQTL SNP in heart (rs33175967; $p = 6.2 \times 10^{-19}$) was ~17 kb distal. The overlap between the association signals for LSKCD150^{-}CD48^{-} frequency and hepatic *Hopx* mRNA expression at the chromosome 5 locus is illustrated in Figure S4B.

To further evaluate whether *Hopx* plays a role in the regulation of HSPCs, we performed gene expression analysis by RT-PCR in BM MNCs of 25 different mouse strains with varying LSKCD150^{-}CD48^{-} frequencies. These analyses demonstrated that *Hopx* was expressed in BM and that its mRNA levels were both variable and correlated ($r = 0.44$) with LSKCD150^{-}CD48^{-} frequency in these strains (Figure 4A). This latter observation is not entirely surprising since the coincident mapping of *cis* eQTLs for *Hopx* across multiple tissues and LSKCD150^{-}CD48^{-} frequency to the same chromosome 5 could be expected to result in a correlation between *Hopx* gene expression and cell number in BM. We next purified various HSPC subsets and other hematopoietic-derived lineages from C57BL/6 mice and determined the cell-specific expression pattern of *Hopx*. These analyses revealed that *Hopx* was expressed in all hematopoietic cells and at particularly high levels in mature hematopoietic lineages, such as CD3^{+} T lymphocytes (Figure 4B). To confirm these findings, we used flow cytometry to analyze BM of genetically modified mice that express

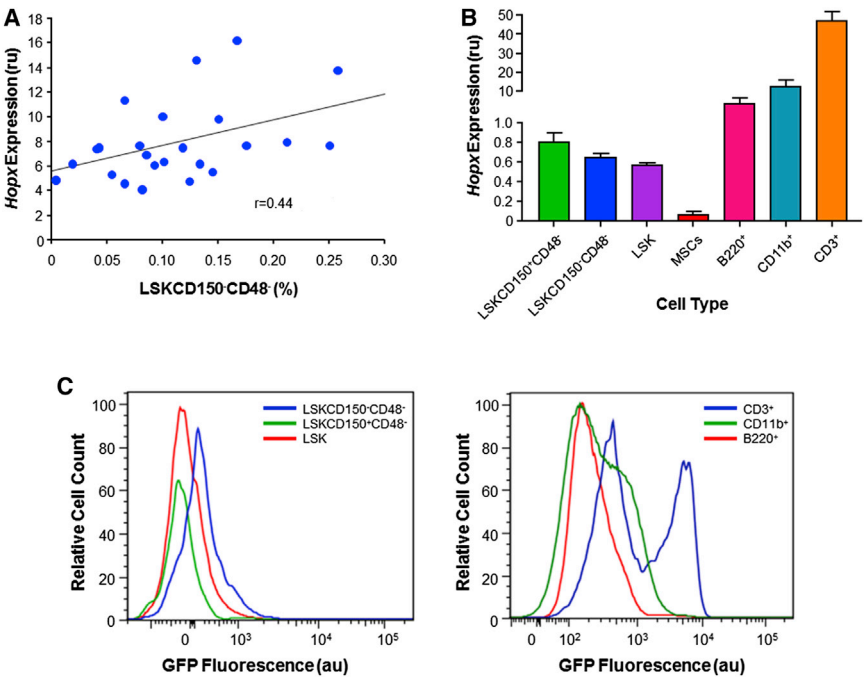


Figure 4. Relationship between *Hopx* Gene Expression and HSPCs

(A) Among 25 HMDP strains, *Hopx* expression levels in BM MNCs are positively correlated (r = 0.44) with LSKCD150^{-}CD48^{-} frequency. Each data point represents the mean expression of each mouse strain (n = 2–3 mice/strain) across two independent experiments. LSKCD150^{-}CD48^{-} frequency is expressed as a percentage of BM MNCs and expression levels are shown relative to the expression of *Hprt*, as an endogenous control, in relative units (ru).

(B) *Hopx* is expressed in various hematopoietic cell subsets of C57Bl/6 mice (n = 3), including all three HSPC populations, but at very low levels in mesenchymal stem cells (MSCs). Expression is also abundant in mature leukocytes, particularly in T lymphocytes (CD3^{+}) compared with B lymphocytes (B220^{+}) or monocytes (CD11b^{+}). Total RNA was isolated from BM and RT-PCR was carried out as described in the Experimental Procedures. Expression levels are shown relative to the expression of *Hprt*, as an endogenous control, in relative units (ru) with error bars representing SEM.

(C) As determined by relative GFP fluorescence intensity, *Hopx* expression was highest in the LSKCD150^{-}CD48^{-} population of HSPCs and especially abundant in CD3^{+} cells, consistent with the RT-PCR results. Fluorescence was quantitated by flow cytometry in different HSPC sub-populations and mature hematopoietic lineages in BM MNCs that were isolated from the femurs and tibias of 12-week-old genetically modified mice (n = 3) expressing GFP under the control of the endogenous *Hopx* promoter (Takeda et al., 2013). Data are expressed in a.u. Note the different x axes in the plots.

GFP under the control of the endogenous *Hopx* promoter (Takeda et al., 2013). As shown in Figure 4C, *Hopx* was expressed highest in the LSKCD150^{-}CD48^{-} population of HSPCs and in CD3^{+} cells, which is consistent with the RT-PCR results. Collectively, these data point to *Hopx* as a strong positional candidate gene since (1) genetic variation at this locus is functional with respect to *Hopx* gene expression in multiple tissues, (2) mRNA levels of *Hopx* in BM MNCs are positively associated with LSKCD150^{-}CD48^{-} number, and (3) *Hopx* is expressed at relatively high levels in the HSPC subset whose frequency maps to the chromosome 5 locus.

Biological Validation of Hopx as a Gene that Regulates LSKCD150^{-}CD48^{-} Cells

To biologically validate *Hopx* as a genetic determinant of HSPCs, we characterized previously generated *Hopx* knockout (*Hopx*$^{-/-}$) mice (Chen et al., 2002) for differences in HSPC frequency and function. Flow cytometric analysis revealed that *Hopx*$^{-/-}$ mice had significantly reduced numbers of LSKCD150^{-}CD48^{-} cells compared to wild-type *Hopx*$^{+/+}$ littermates, but no differences in the frequency of LSK (Figure 5A) or LSKCD150^{+}CD48^{-} cells (data not shown). Since the reduced frequency of LSKCD150^{-}CD48^{-}cells in *Hopx*$^{-/-}$ mice could be due to either intrinsic defects of the cells or extrinsic factors, we performed competitive repopulation assays with HSCs from *Hopx*$^{-/-}$ mice and wild-type littermates. These experiments demonstrated that transplantation of HSCs from *Hopx*$^{-/-}$ mice did not affect levels of circulating CD45.2 cells at early time points, but led to significantly impaired engraftment at 16 weeks after transplantation and extending to 24 weeks (Figure 5B). However, analysis of multi-lineage engraftment in these mice did not reveal significant differences in the relative abundance of circulating B lymphocytes, T lymphocytes, or myeloid cells (Figure S5).

We next analyzed the composition of BM MNCs at the 24-week time point. Compared to transplantation with *Hopx*$^{+/+}$ HSCs, mice transplanted with HSCs from *Hopx*$^{-/-}$ mice had a significant and specific reduction in the frequency of LSKCD150^{-}CD48^{-} cells, but not LSK (Figure 5C) or LSKCD150^{+}CD48^{-} cells (data not shown). To investigate a potential biological mechanism for the functional effects of *Hopx* deficiency on HSPC frequency, we analyzed the cell-cycle status of LSK and LSKCD150^{-}CD48^{-} cells. *Hopx*$^{-/-}$ mice had significantly reduced numbers of quiescent LSKCD150^{-}CD48^{-} cells and increased numbers of cells in the G_1 phase

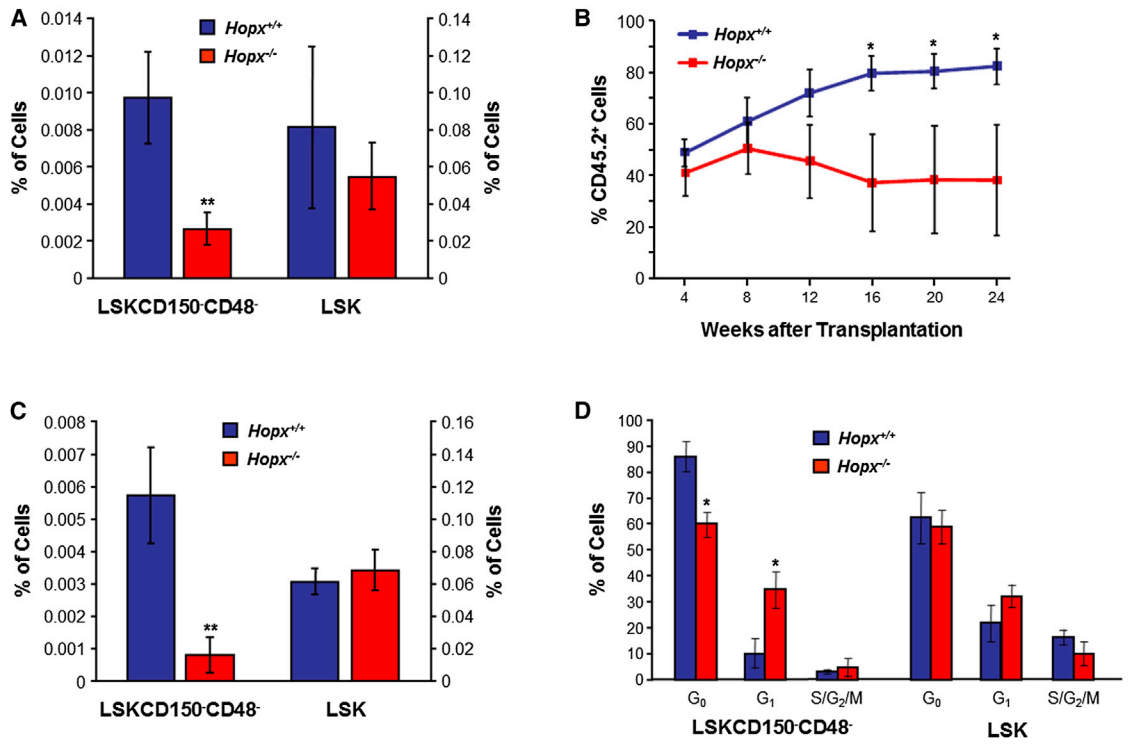


Figure 5. Functional Validation of the Effects of *Hopx* on HSPC Physiology

(A) The frequency of LSKCD150$^-$CD48$^-$ cells, as a percentage of BM MNCs (left y axis), are significantly reduced in *Hopx*$^{-/-}$ mice compared to wild-type littermate controls (*Hopx*$^{+/+}$), whereas LSKs are not affected (right y axis). BM MNCs were isolated from the femurs and tibias of 12-week-old male mice and the frequency of different HSPC sub-populations was determined by flow cytometry.

(B) Competitive repopulation assay shows significantly decreased levels of engraftment starting at 16 weeks after transplantation of HSCs isolated from BM of *Hopx*$^{-/-}$ mice compared to wild-type *Hopx*$^{+/+}$ littermates. Mice were bled from the tail vein every 4 weeks after transplantation and the level of engraftment, indicated by the percentage of circulating CD45.2$^+$ cells, was quantified by flow cytometry.

(C) Frequency of LSKCD150$^-$CD48$^-$ cells in BM, as a percentage of MNCs (left y axis), is significantly reduced in mice transplanted with HSCs from *Hopx*$^{-/-}$ mice compared to *Hopx*$^{+/+}$ littermate controls 6 months following transplantation. There is no difference in the frequency of LSKs (right y axis).

(D) LSKCD150$^-$CD48$^-$ cells of *Hopx*$^{-/-}$ mice exhibit altered cell-cycle status, with decreased and increased percentage of cells in the G_0 and G_1 phases, respectively, compared to wild-type littermates. No effects were observed in LSKs.

For each experiment in (A)–(D), error bars represent SEM and n = 3–5 mice per group. *p < 0.05; **p < 0.01. See also Figure S5.

(Figure 5D). However, *Hopx* deficiency had no cell-cycle effects on LSKs (Figure 5D) or LSKCD150$^+$CD48$^-$ cells (data not shown). Taken together, these data validate the GWAS findings and provide functional evidence that *Hopx* is a positive regulator of HSPC frequency in mice.

DISCUSSION

In the present study, we carried out a phenotypic screen with over 100 inbred strains from the HMDP to investigate how the frequency of three HSPC populations varies as a function of naturally occurring genetic variation. The results revealed substantial variability across this panel of mouse strains, at least based on the specific cell surface markers we used to quantify cell number. However, this may not necessarily reflect the true variation in the pool size of all stem cells, since, for example, the number of HSPCs in strains carrying the Sca-1 null allele could be underestimated. Despite this possibility, we still observed relatively strong correlations between certain HSPC subsets, but, interestingly, not between HSPCs and the number of circulating mature blood cells or other hematological parameters.

Using the strain survey data, we also carried out a GWAS analysis and identified multiple distinct and overlapping loci that were significantly or suggestively associated with HSPC sub-populations. At one particular locus on chromosome 5, we provide strong evidence that LSKCD150$^-$CD48$^-$ cell frequency and function is

regulated by *Hopx*, a gene that was previously not known to influence HSPC biology. The collective results of the functional validation studies with $Hopx^{-/-}$ mice and competitive transplantation assays demonstrated that *Hopx* deficiency led to a specific and intrinsic functional defect in the MPP subset of HSCs ($LSKCD150^{-}CD48^{-}$), possibly related to cell-cycle status, but not in the most primitive long-term subset of cells or more committed progenitor populations. These observations were also entirely consistent with the gene expression analyses, showing that *Hopx* mRNA levels in the BM were positively correlated with $LSKCD150^{-}CD48^{-}$ cells, and the GWAS results, where the *Hopx* locus was only associated with this HSPC subset and not with LSKs or $LSKCD150^{+}CD48^{-}$ cells. Of direct relevance to our findings, a recent study demonstrated how different HSC populations, and $LSKCD150^{-}CD48^{-}$ cells in particular, play distinct temporal roles in the maintenance of the hematopoietic system (Busch et al., 2015). More specifically, it is the $LSKCD150^{-}CD48^{-}$ sub-population of cells that sustains adult hematopoiesis under homeostatic conditions through multiple rounds of self-renewal. Furthermore, proliferation of this cell population accounts for the differentiation of committed progenitors, while only receiving limited input from the more primitive $LSKCD150^{+}CD48^{-}$ sub-population of cells.

Hopx is a homeodomain-only protein that lacks DNA-binding activity yet is predominantly localized to the nucleus (Chen et al., 2002; Shin et al., 2002). There are many different functional roles proposed for this protein depending on the cell type examined. Through direct protein-protein interactions, *Hopx* was shown to repress the expression of serum response factor (SRF)-dependent genes (Kook and Epstein, 2003), and conditional deletion of *Srf* in mouse BM leads to an approximate 3-fold increase in the frequency of all HSPC sub-populations (Ragu et al., 2010). Therefore, one plausible mechanism for why *Hopx* deficiency leads to decreased numbers of $LSKCD150^{-}CD48^{-}$ cells could be as a result of increased *Srf*-mediated transcription of downstream targets. Alternatively, *Hopx* may affect HSPC frequency through its effects on cell-cycle status. For example, we demonstrated that $Hopx^{-/-}$ mice have a decreased proportion of $LSKCD150^{-}CD48^{-}$ cells in the quiescent G_0 state and increased numbers in the G_1 phase. Other groups have shown that *Hopx* promoter methylation frequently is observed in many malignancies, including esophageal, gastric, pancreatic, and colorectal cancers, where it has been suggested that *Hopx* functions as a tumor suppressor gene (Yamashita et al., 2013). Since it is known that deficiency of a number of genes in the hematopoietic system leads to increased entry into cell cycle, with a consequent loss of HSPC frequency and function (Orford and Scadden, 2008), it is possible that perturbation of *Hopx* expression/activity affects the intrinsic homeostasis of HSPCs by altering cellular quiescence and, thus, cell frequency.

Two other interesting observations from our GWAS results were the overlap of loci that previously had been mapped for HSPC phenotypes using linkage analysis and identification of genomic regions that harbored genes encoding cell surface markers used for flow cytometry. In this regard, we identified a major locus for LSK cells over the genes encoding Sca-1 (*Ly6e-Ly6a-Ly6f*) on chromosome 15, which confirms previous reports that certain strains carry a Sca-1 haplotype associated with decreased expression (Spangrude and Brooks, 1993). Exclusion of these strains from the analysis did not significantly decrease the heritability of LSK variation or reveal additional loci, but yielded highly suggestive evidence for association with a chromosome 18 locus containing *Zfp521*. QTL analysis with BXD RI strains also identified a locus for HSC frequency on chromosome 18 near *Zfp521*, although the peak of that linkage signal mapped ~20 Mb distal from our lead GWAS SNP (de Haan and Van Zant, 1997). *Zfp521* encodes a transcription factor protein with 30 zinc fingers that was originally identified as being specifically expressed in primitive human $CD34^{+}$ cells compared to more mature hematopoietic cells (Bond et al., 2004). More recently, *Zfp521* also has been implicated in bone formation (Hesse et al., 2010; Kiviranta et al., 2013), but validation of *Zfp521* as a gene directly influencing HSPCs will require additional functional studies. Notably, the loci on chromosomes 15 (*Ly6e-Ly6a-Ly6f*) and 18 (*Zfp521-Ss18*) were identified in the GWAS analyses for $LSKCD150^{-}CD48^{-}$ cells as well. However, aside from these two regions, there was no overlap between the loci identified for the three HSPC populations or with those previously reported for blood cell traits in the HMDP (Davis et al., 2013). In addition to *Ly6e-Ly6e-Ly6f*, our GWAS analysis revealed highly significant association of $LSKCD150^{-}CD48^{-}$ cells with the *Slam* locus on chromosome 1, the strength of which increased by two orders of magnitude after excluding the low Sca-1-expressing strains. Consistent with our results, Müller-Sieburg and Riblet (1996) also mapped a QTL for HSCs directly over the *Slam* locus using the CAFC assay and linkage analysis in the BXD RI panel. Although we cannot determine with certainty whether *Cd48* and *Slamf1* account for the signal on chromosome 1 due to extensive LD across this region, *Cd48*-deficient mice do exhibit increased numbers of short-term HSCs (Boles et al., 2011).

While GWAS in humans has been successful in identifying genes for many different disease phenotypes, these types of analysis have had limited success for HSPCs. This is due, in part, to the logistical challenges of obtaining the relevant tissue in humans (i.e., BM) in large numbers of subjects in order to have sufficient mapping power. For example, a GWAS in the Framingham Heart Study for levels

of circulating $CD34^+$ cells only identified suggestive loci, none of which could be functionally validated (Cohen et al., 2013). With respect to the present results, the relevance of the identified genes and pathways to humans is still to be determined. However, it is encouraging that recent studies in the HMDP were able to directly correlate mouse association signals with those identified in human GWAS (Davis et al., 2013; Parks et al., 2013). For instance, four of the five loci identified for RBC parameters in the HMDP (Davis et al., 2013) correspond to loci recently reported for analogous phenotypes in a large human GWAS (van der Harst et al., 2012). These observations reinforce the concept that the underlying biological pathways for complex traits are likely to be conserved between mice and humans.

In summary, we demonstrate the power of the HMDP as a platform to identify the genetic determinants of highly relevant biomedical traits that would otherwise not be feasible or logistically challenging to investigate in humans. While our studies only investigated the genetic determinants of HSPC frequency in the adult BM, they provide a proof of principle that similar unbiased approaches can be applied to comparable or more complex phenotypes in other stem cell systems. Such gene discovery efforts may lead to targets that could be studied toward further understanding the biological regulation of stem cells and/or manipulated for therapeutic development.

EXPERIMENTAL PROCEDURES

Animals

The 12-week-old male mice of 108 inbred strains (n = 3–8) from the HMDP were obtained (The Jackson Laboratory) and used in accordance with the USC Institutional Animal Care and Use Committee guidelines. Mice were maintained on a 12-hr light/dark cycle in sterilized microisolator cages and received autoclaved food and water ad libitum. Biological validation of *Hopx* was carried out using previously generated genetically targeted mice (Chen et al., 2002; Takeda et al., 2013).

Flow Cytometry

BM MNCs were incubated in PBS with fluorescent labeled anti-mouse antibodies. The labeled cells were then analyzed by flow cytometry using an LSR II flow cytometer (Becton Dickinson), and HSPC frequencies were calculated according to established cell immunophenotypes using FlowJo flow cytometry analysis software. Figure S1 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. Details regarding the flow cytometric analyses are described in the Supplemental Experimental Procedures and a list of all antibodies used is given in Table S3.

Complete Blood Count Analysis

Mice (n = 3/strain) were fasted for 4–5 hr and bled from the retro-orbital plexus under isoflurane anesthesia 2–3 hr after the beginning of the light cycle. Blood was collected in 20-μl EDTA-coated glass capillaries, processed using standard procedures, and profiled with respect to hematological parameters using the Heska HemaTrue Veterinary Hematology Analyzer, as described previously (Davis et al., 2013).

Statistical Genetics Analysis

GWAS analyses utilized ~880,000 genotyped and imputed SNPs and implemented an efficient mixed model algorithm (EMMA) that takes into account the underlying population structure and genetic relatedness across the strains, as described previously (Kang et al., 2008; Bennett et al., 2010). Details regarding the statistical genetics methodology are described in the Supplemental Experimental Procedures.

Gene Expression Analysis

Total RNA was extracted using RNA Miniprep Kits (Stratagene) and was reverse-transcribed into cDNA using SuperScript VILO cDNA synthesis kits (Invitrogen). Taqman gene expression assay primers and probe sets (Applied Biosystems and Roche Diagnostics) were used to quantify expression of *Hopx* (assay ID Mm00558630_m1) and *Hprt* (assay ID Mm01545399_m1) as an endogenous control. Gene expression levels were measured in triplicate using a 7900HT real-time PCR system (Applied Biosystems). Standard curves in these experiments were created with the use of qPCR mouse reference total RNA (Stratagene).

Competitive Repopulation Assays

BM MNCs were obtained from $Hopx^{+/+}$ and $Hopx^{-/-}$ mice and the primitive $LSKCD150^+CD48^-$ HSCs were isolated using cell sorting. To quantify the ability to reconstitute hematopoiesis, 100 $LSKCD150^+CD48^-$ cells from $Hopx^{+/+}$ or $Hopx^{-/-}$ mice, along with 250,000 competitor BM MNCs from a B6.SJL congenic mouse, were injected into the tail veins of B6.SJL mice that were lethally irradiated with 9 Gy approximately 24 hr prior to transplantation. Peripheral blood samples were obtained from the tail vein of hosts starting at 4 weeks after transplantation and every 4 weeks thereafter, for a total of 24 weeks. PE anti-mouse CD45.1, fluorescein isothiocyanate (FITC) anti-mouse CD45.2, APC anti-mouse CD3e, APC-Cy7 anti-mouse CD11b, and biotinylated anti-mouse B220 antibodies (all from Becton Dickinson) were used to assess the degree of multi-lineage reconstitution.

Cell-Cycle Experiments

For cell-cycle analysis, cells were incubated with 10 μg/ml Hoechst 33342 (Sigma-Aldrich) at 37°C for 45 min, then stained with lineage and stem cell antibodies as described above. The stained cells were resuspended and fixed in 10% buffered formalin and incubated at 4°C overnight. To stain for RNA content, pyronin Y (Polysciences) was added to the cells at a final concentration of 0.75 μg/ml and incubated at 4°C for 30 min. Cell-cycle status was examined using an LSR II flow cytometer (Becton Dickinson) and FlowJo software.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be

found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.05.008.

AUTHOR CONTRIBUTIONS

G.B.A. and H.A. conceived the study. X.Z., A.L.C., J.H., T.J.S., A.G., L.W.B., B.B.B., B.W.P., and R.J. performed experiments, collected data, and analyzed data. E.E., J.A.E., A.J.L., G.B.A., and H.A. contributed to experimental design and data analysis and interpretation. X.Z., A.L.C., G.B.A., and H.A. wrote the paper.

ACKNOWLEDGMENTS

This work was supported in part by California Institute for Regenerative Medicine grant TG2-01161; NIH grants T32ES013678, K99HL102223, K99HL123021, R01ES022282, R01HL071546, P01HL30568, P01HL28481, R01ES021801, 3R01ES021801-03S1, UL1TR000130, and P30ES007048; and the Margaret E. Early Medical Research Trust.

Received: September 15, 2014
Revised: May 7, 2015
Accepted: May 8, 2015
Published: June 4, 2015

REFERENCES

Albrecht, I., Niesner, U., Janke, M., Menning, A., Loddenkemper, C., Kühl, A.A., Lepenies, I., Lexberg, M.H., Westendorf, K., Hradilkova, K., et al. (2010). Persistence of effector memory Th1 cells is regulated by Hopx. Eur. J. Immunol. *40*, 2993–3006.

Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., et al. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res. *20*, 281–290.

Bennett, B.J., Orozco, L., Kostem, E., Erbilgin, A., Dallinga, M., Neuhaus, I., Guan, B., Wang, X., Eskin, E., and Lusis, A.J. (2012). High-resolution association mapping of atherosclerosis loci in mice. Arterioscler. Thromb. Vasc. Biol. *32*, 1790–1798.

Boles, N.C., Lin, K.K., Lukov, G.L., Bowman, T.V., Baldridge, M.T., and Goodell, M.A. (2011). CD48 on hematopoietic progenitors regulates stem cells and suppresses tumor formation. Blood *118*, 80–87.

Bond, H.M., Mesuraca, M., Carbone, E., Bonelli, P., Agosti, V., Amodio, N., De Rosa, G., Di Nicola, M., Gianni, A.M., Moore, M.A., et al. (2004). Early hematopoietic zinc finger protein (EHZF), the human homolog to mouse Evi3, is highly expressed in primitive human hematopoietic cells. Blood *103*, 2062–2070.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542–546.

Bystrykh, L., Weersing, E., Dontje, B., Sutton, S., Pletcher, M.T., Wiltshire, T., Su, A.I., Vellenga, E., Wang, J., Manly, K.F., et al. (2005). Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat. Genet. *37*, 225–232.

Chambers, S.M., Boles, N.C., Lin, K.-Y.K., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C., et al. (2007). Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell *1*, 578–591.

Chen, F., Kook, H., Milewski, R., Gitler, A.D., Lu, M.M., Li, J., Nazarian, R., Schnepp, R., Jen, K., Biben, C., et al. (2002). Hop is an unusual homeobox gene that modulates cardiac development. Cell *110*, 713–723.

Cohen, K.S., Cheng, S., Larson, M.G., Cupples, L.A., McCabe, E.L., Wang, Y.A., Ngwa, J.S., Martin, R.P., Klein, R.J., Hashmi, B., et al. (2013). Circulating CD34(+) progenitor cell frequency is associated with clinical and genetic factors. Blood *121*, e50–e56.

Davis, R.C., van Nas, A., Bennett, B., Orozco, L., Pan, C., Rau, C.D., Eskin, E., and Lusis, A.J. (2013). Genome-wide association mapping of blood cell traits in mice. Mamm. Genome *24*, 105–118.

de Haan, G., and Van Zant, G. (1997). Intrinsic and extrinsic control of hemopoietic stem cell numbers: mapping of a stem cell gene. J. Exp. Med. *186*, 529–536.

de Haan, G., and Van Zant, G. (1999). Dynamic changes in mouse hematopoietic stem cell numbers during aging. Blood *93*, 3294–3301.

de Haan, G., Nijhof, W., and Van Zant, G. (1997). Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. Blood *89*, 1543–1550.

de Haan, G., Bystrykh, L.V., Weersing, E., Dontje, B., Geiger, H., Ivanova, N., Lemischka, I.R., Vellenga, E., and Van Zant, G. (2002). A genetic and genomic analysis identifies a cluster of genes associated with hematopoietic cell turnover. Blood *100*, 2056–2062.

Farber, C.R., Bennett, B.J., Orozco, L., Zou, W., Lira, A., Kostem, E., Kang, H.M., Furlotte, N., Berberyan, A., Ghazalpour, A., et al. (2011). Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis. PLoS Genet. *7*, e1002038.

Geiger, H., True, J.M., de Haan, G., and Van Zant, G. (2001). Age- and stage-specific regulation patterns in the hematopoietic stem cell hierarchy. Blood *98*, 2966–2972.

Gerrits, A., Dykstra, B., Otten, M., Bystrykh, L., and de Haan, G. (2008). Combining transcriptional profiling and genetic linkage analysis to uncover gene networks operating in hematopoietic stem cells and their progeny. Immunogenetics *60*, 411–422.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., et al. (2011). Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. *7*, e1001393.

Ghazalpour, A., Rau, C.D., Farber, C.R., Bennett, B.J., Orozco, L.D., van Nas, A., Pan, C., Allayee, H., Beaven, S.W., Civelek, M., et al. (2012). Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mamm. Genome *23*, 680–692.

Ghazalpour, A., Bennett, B.J., Shih, D., Che, N., Orozco, L., Pan, C., Hagopian, R., He, A., Kayne, P., Yang, W.P., et al. (2014). Genetic regulation of mouse liver metabolite levels. Mol. Syst. Biol. *10*, 730.

Hartiala, J., Bennett, B.J., Tang, W.H., Wang, Z., Stewart, A.F., Roberts, R., McPherson, R., Lusis, A.J., Hazen, S.L., and Allayee, H.; CARDIoGRAM Consortium (2014). Comparative genome-wide association studies in mice and humans for trimethylamine N-oxide, a proatherogenic metabolite of choline and L-carnitine. Arterioscler. Thromb. Vasc. Biol. *34*, 1307–1313.

Henckaerts, E., Geiger, H., Langer, J.C., Rebollo, P., Van Zant, G., and Snoeck, H.W. (2002). Genetically determined variation in the number of phenotypically defined hematopoietic progenitor and stem cells and in their response to early-acting cytokines. Blood *99*, 3947–3954.

Henckaerts, E., Langer, J.C., and Snoeck, H.W. (2004). Quantitative genetic variation in the hematopoietic stem cell and progenitor cell compartment and in lifespan are closely linked at multiple loci in BXD recombinant inbred mice. Blood *104*, 374–379.

Hesse, E., Kiviranta, R., Wu, M., Saito, H., Yamana, K., Correa, D., Atfi, A., and Baron, R. (2010). Zinc finger protein 521, a new player in bone formation. Ann. N Y Acad. Sci. *1192*, 32–37.

Hindorff, L.A., Junkins, H.A., Hall, P.N., Mehta, J.P., and Manolio, T.A. (2015). A catalog of published genome-wide association studies. http://www.genome.gov/gwastudies/.

Hope, K.J., Cellot, S., Ting, S.B., MacRae, T., Mayotte, N., Iscove, N.N., and Sauvageau, G. (2010). An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell *7*, 101–113.

Ito, C.Y., Li, C.Y., Bernstein, A., Dick, J.E., and Stanford, W.L. (2003). Hematopoietic stem cell and progenitor defects in Sca-1/Ly-6A-null mice. Blood *101*, 517–523.

Jawad, M., Cole, C., Zanker, A., Giotopoulos, G., Fitch, S., Talbot, C.J., and Plumb, M. (2008). QTL analyses of lineage-negative mouse bone marrow cells labeled with Sca-1 and c-Kit. Mamm. Genome *19*, 190–198.

Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. Genetics *178*, 1709–1723.

Karlsson, C., Baudet, A., Miharada, N., Soneji, S., Gupta, R., Magnusson, M., Enver, T., Karlsson, G., and Larsson, J. (2013). Identification of the chemokine CCL28 as a growth and survival factor for human hematopoietic stem and progenitor cells. Blood *121*, 3838–3842.

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.

Kiel, M.J., Yilmaz, O.H., and Morrison, S.J. (2008). CD150- cells are transiently reconstituting multipotent progenitors with little or no stem cell activity. Blood *111*, 4413–4414.

Kiviranta, R., Yamana, K., Saito, H., Ho, D.K., Laine, J., Tarkkonen, K., Nieminen-Pihala, V., Hesse, E., Correa, D., Määttä, J., et al. (2013). Coordinated transcriptional regulation of bone homeostasis by Ebf1 and Zfp521 in both mesenchymal and hematopoietic lineages. J. Exp. Med. *210*, 969–985.

Kook, H., and Epstein, J.A. (2003). Hopping to the beat. Hop regulation of cardiac gene expression. Trends Cardiovasc. Med. *13*, 261–264.

Liang, Y., Jansen, M., Aronow, B., Geiger, H., and Van Zant, G. (2007). The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice. Nat. Genet. *39*, 178–188.

Morrison, S.J., and Weissman, I.L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity *1*, 661–673.

Morrison, S.J., Qian, D., Jerabek, L., Thiel, B.A., Park, I.K., Ford, P.S., Kiel, M.J., Schork, N.J., Weissman, I.L., and Clarke, M.F. (2002). A genetic determinant that specifically regulates the frequency of hematopoietic stem cells. J. Immunol. *168*, 635–642.

Müller-Sieburg, C.E., and Riblet, R. (1996). Genetic control of the frequency of hematopoietic stem cells in mice: mapping of a candidate locus to chromosome 1. J. Exp. Med. *183*, 1141–1150.

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

Orford, K.W., and Scadden, D.T. (2008). Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nat. Rev. Genet. *9*, 115–128.

Orozco, L.D., Bennett, B.J., Farber, C.R., Ghazalpour, A., Pan, C., Che, N., Wen, P., Qi, H.X., Mutukulu, A., Siemers, N., et al. (2012). Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. Cell *151*, 658–670.

Park, C.C., Gale, G.D., de Jong, S., Ghazalpour, A., Bennett, B.J., Farber, C.R., Langfelder, P., Lin, A., Khan, A.H., Eskin, E., et al. (2011). Gene networks associated with conditional fear in mice identified using a systems genetics approach. BMC Syst. Biol. *5*, 43.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., Rau, C.D., Bennett, B.J., et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab. *17*, 141–152.

Purton, L.E., and Scadden, D.T. (2007). Limiting factors in murine hematopoietic stem cell assays. Cell Stem Cell *1*, 263–270.

Ragu, C., Elain, G., Mylonas, E., Ottolenghi, C., Cagnard, N., Daegelen, D., Passegué, E., Vainchenker, W., Bernard, O.A., and Penard-Lacronique, V. (2010). The transcription factor Srf regulates hematopoietic stem cell adhesion. Blood *116*, 4464–4473.

Rossi, L., Lin, K.K., Boles, N.C., Yang, L., King, K.Y., Jeong, M., Mayle, A., and Goodell, M.A. (2012). Less is more: unveiling the functional core of hematopoietic stem cells through knockout mice. Cell Stem Cell *11*, 302–317.

Shin, C.H., Liu, Z.P., Passier, R., Zhang, C.L., Wang, D.Z., Harris, T.M., Yamagishi, H., Richardson, J.A., Childs, G., and Olson, E.N. (2002). Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. Cell *110*, 725–735.

Spangrude, G.J., and Brooks, D.M. (1993). Mouse strain variability in the expression of the hematopoietic stem cell antigen Ly-6A/E by bone marrow cells. Blood *82*, 3327–3332.

Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science *334*, 1420–1424.

Takeda, N., Jain, R., Leboeuf, M.R., Padmanabhan, A., Wang, Q., Li, L., Lu, M.M., Millar, S.E., and Epstein, J.A. (2013). Hopx expression defines a subset of multipotent hair follicle stem cells and a progenitor population primed to give rise to K6+ niche cells. Development *140*, 1655–1664.

van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D.S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature *492*, 369–375.

van Os, R., Ausema, A., Noach, E.J., van Pelt, K., Dontje, B.J., Vellenga, E., and de Haan, G. (2006). Identification of quantitative trait loci regulating haematopoietic parameters in B6AKRF2 mice. Br. J. Haematol. *132*, 80–90.

Yamashita, K., Katoh, H., and Watanabe, M. (2013). The homeobox only protein homeobox (HOPX) and colorectal cancer. Int. J. Mol. Sci. *14*, 23231–23243.

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.

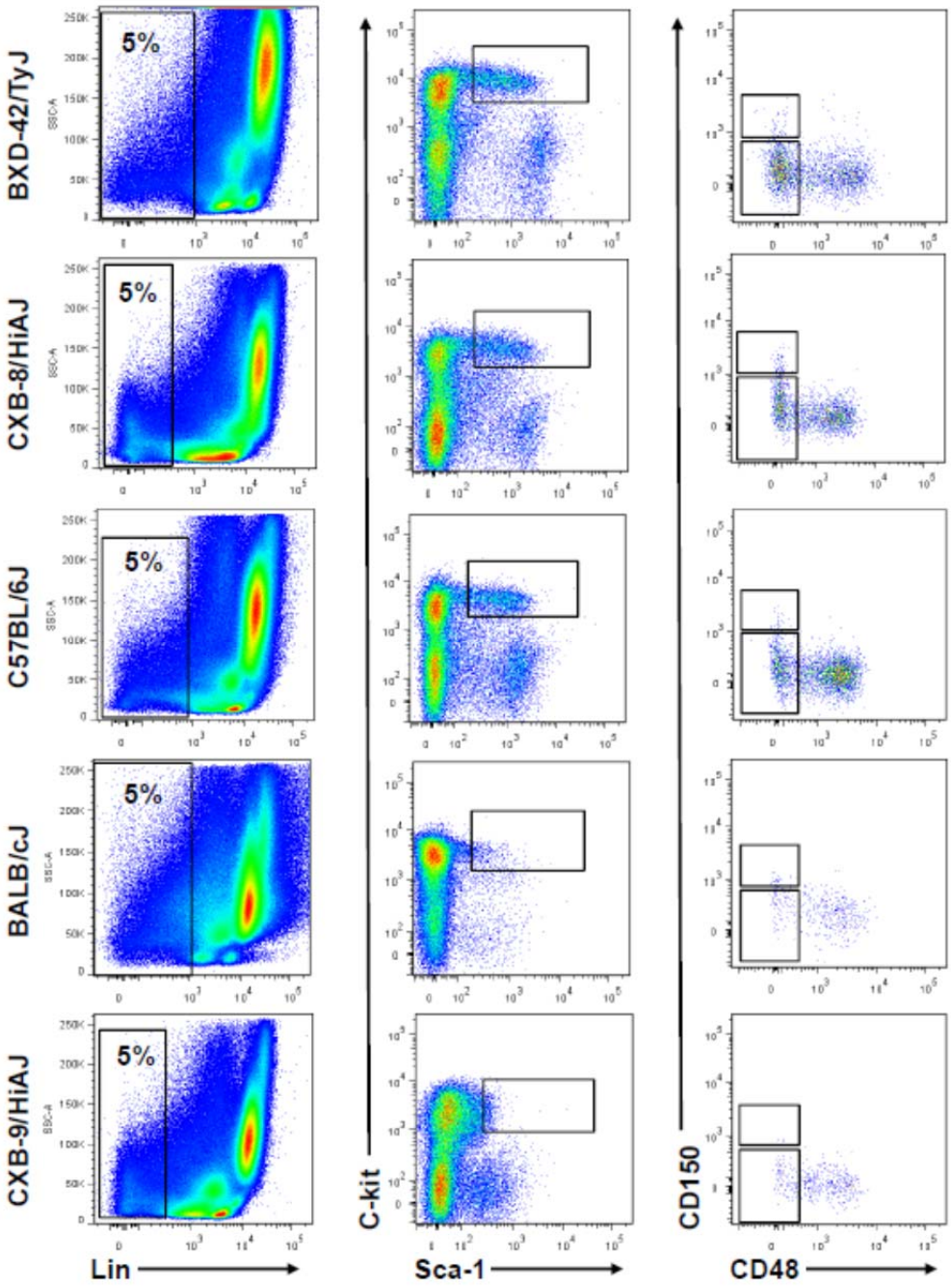
Supplemental Table 2. Correlations of Primitive Hematopoietic Cell Frequency with Blood Cell Parameters, Related to Figure 1.

		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

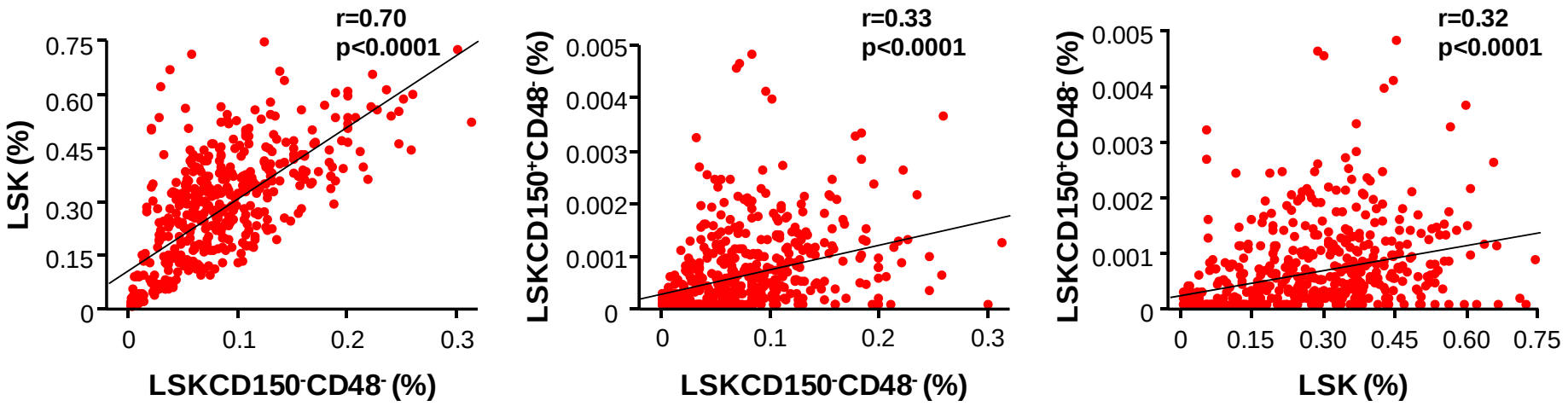
WBC: white blood cell count; Ly %: percent lymphocytes; Mo %: percent monocytes; Gr %: percent granulocyte; RBC, red blood cell count; MCV: mean corpuscular volume; HCT: hematocrit; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW%: red cell distribution width, percent; 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.

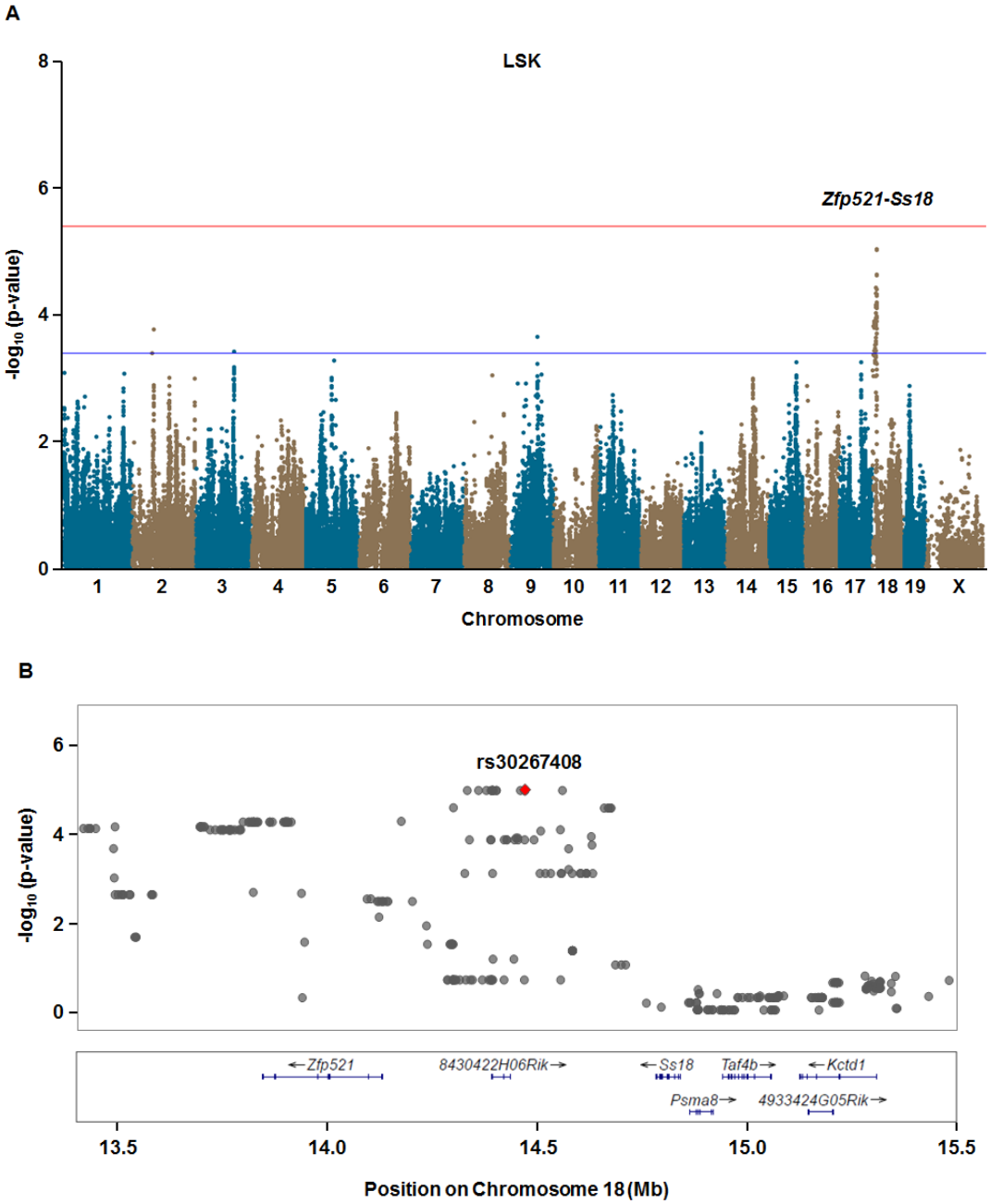
Antibody	Conjugated	Vendor	Catalog number
Lineage Cocktail (CD3e, CD11b, CD45R/B220, TER-119, and Ly6G and Ly-6C)	Streptavidin APC-Cy7	BD Pharmingen	559971
c-kit	APC	eBioscience	17-1172-83
c-kit	PE-Cy7	eBioscience	25-1171-82
Sca-1	PE-Cy7	eBioscience	15-5981-82
Sca-1	AF-700	eBioscience	56-5981-82
Sca-1	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



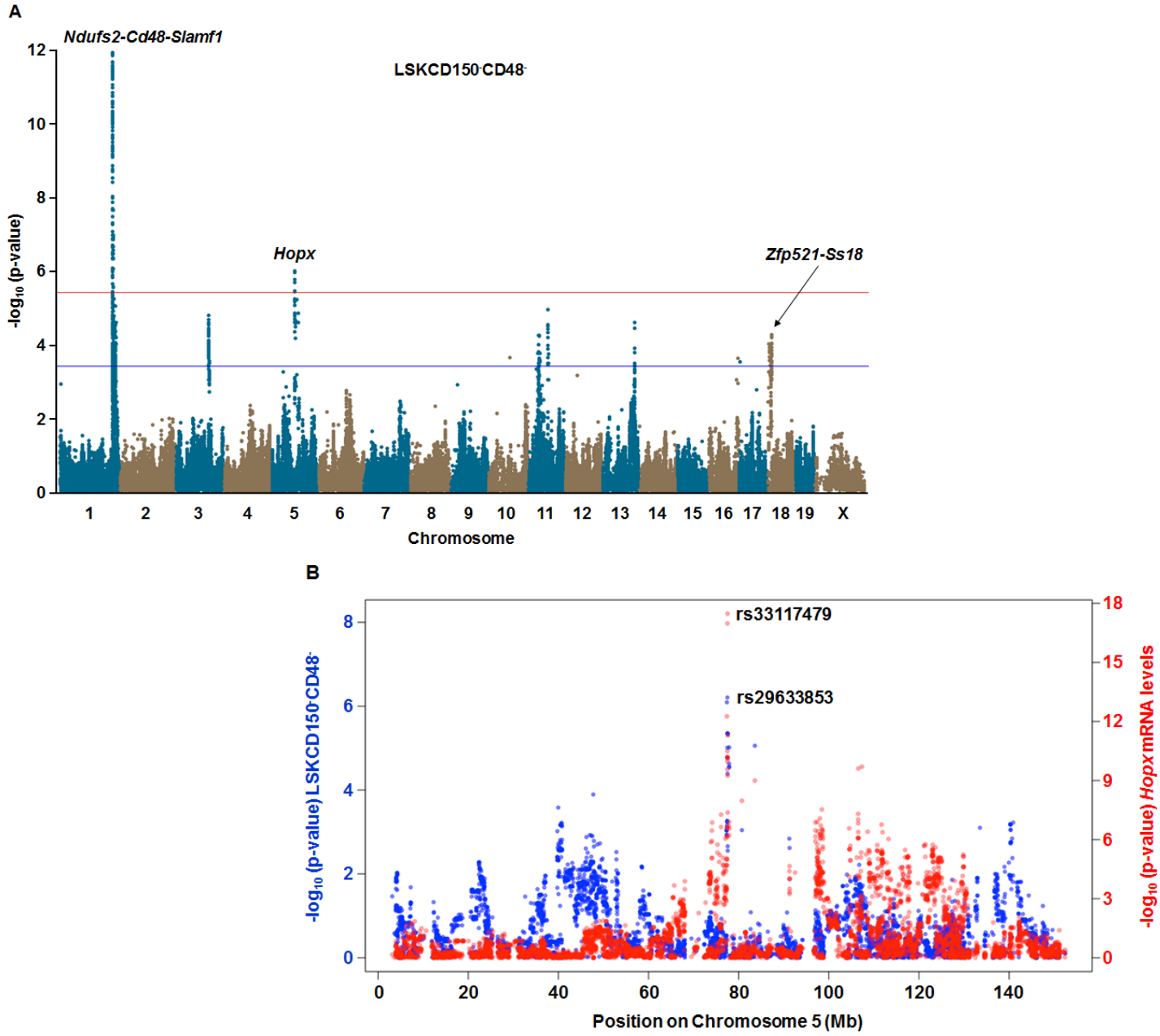
Supplemental Figure 1. Gating strategy for flow cytometric analyses of HSPCs in the HMDP, Related to Figure 1. Representative sequential gating strategy for the Lin$^-$Sca-1$^+$c-Kit$^+$CD150$^+$CD48$^-$ and Lin$^-$Sca-1$^+$c-Kit$^+$CD150$^-$CD48$^-$ cells are demonstrated. Five mouse strains are shown 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.

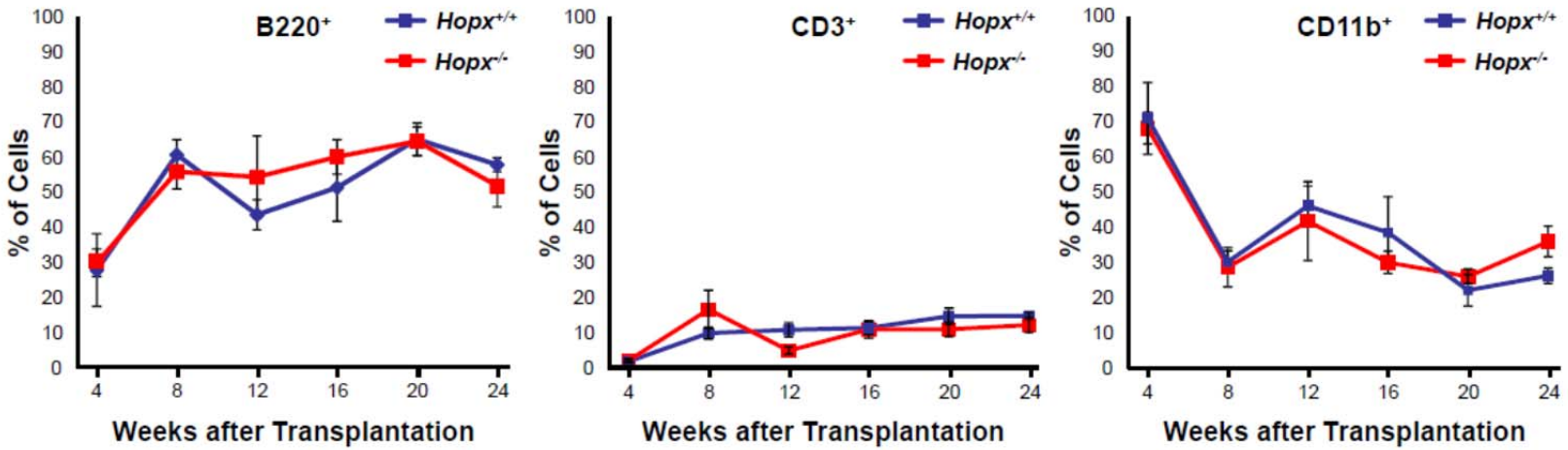


Supplemental Figure 3. GWAS results for LSK frequency after exclusion of strains carrying the uninformative Sca-1-expressing haplotype, Related to Table 1 and Figure 2. **(A)** A Manhattan plot shows no significantly associated locus but a suggestive peak on chromosome 18 harboring *Zfp521* and *Ss18* 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.