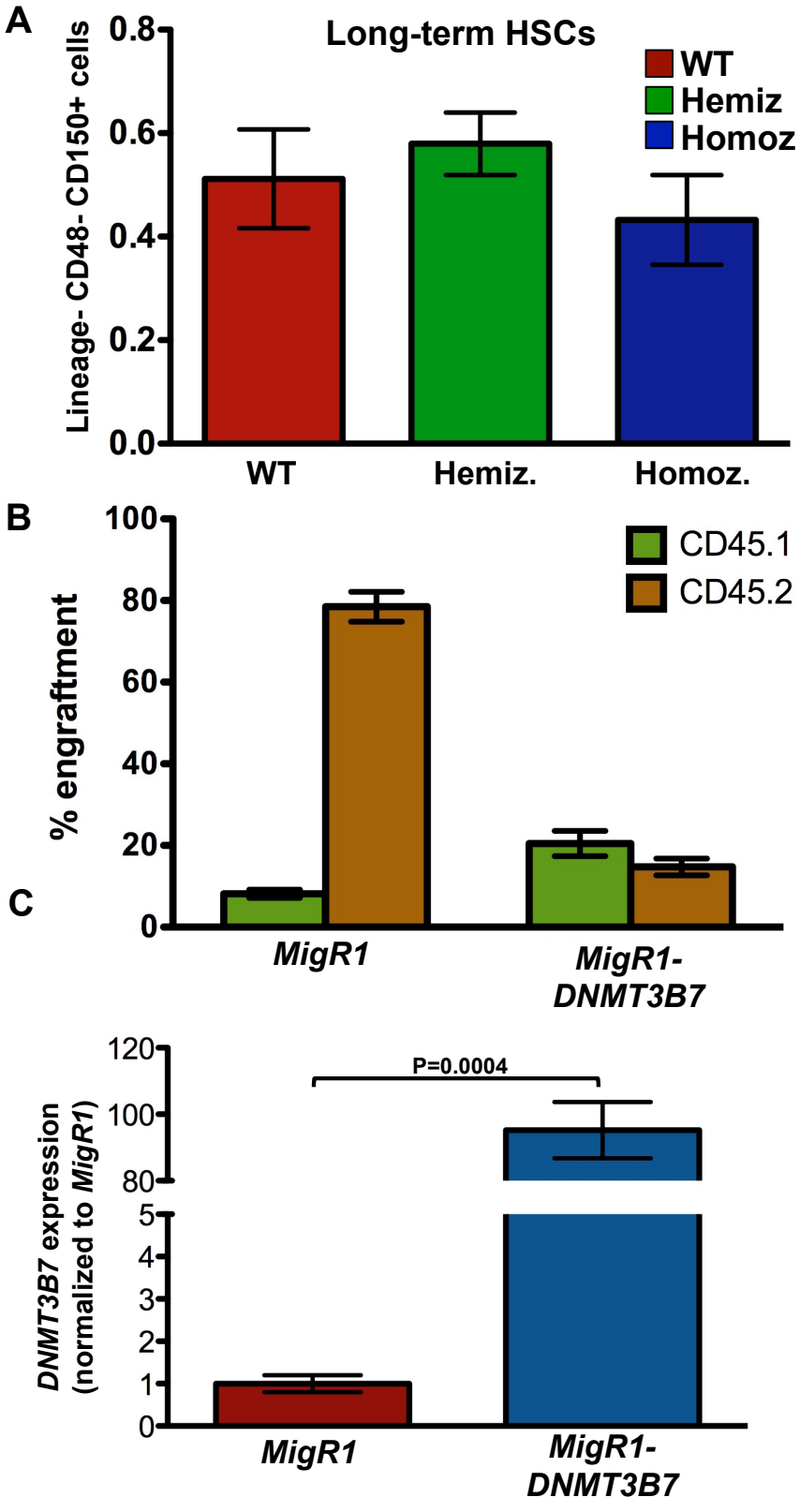
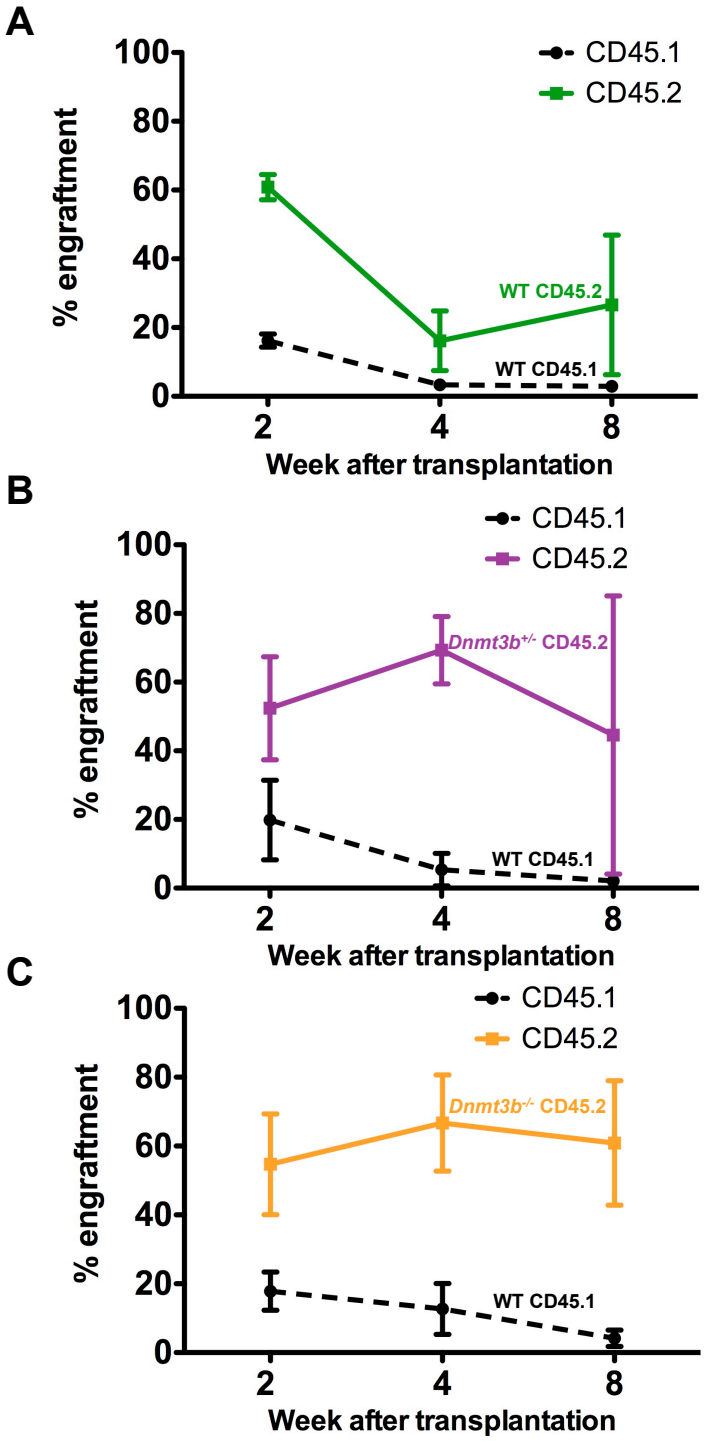
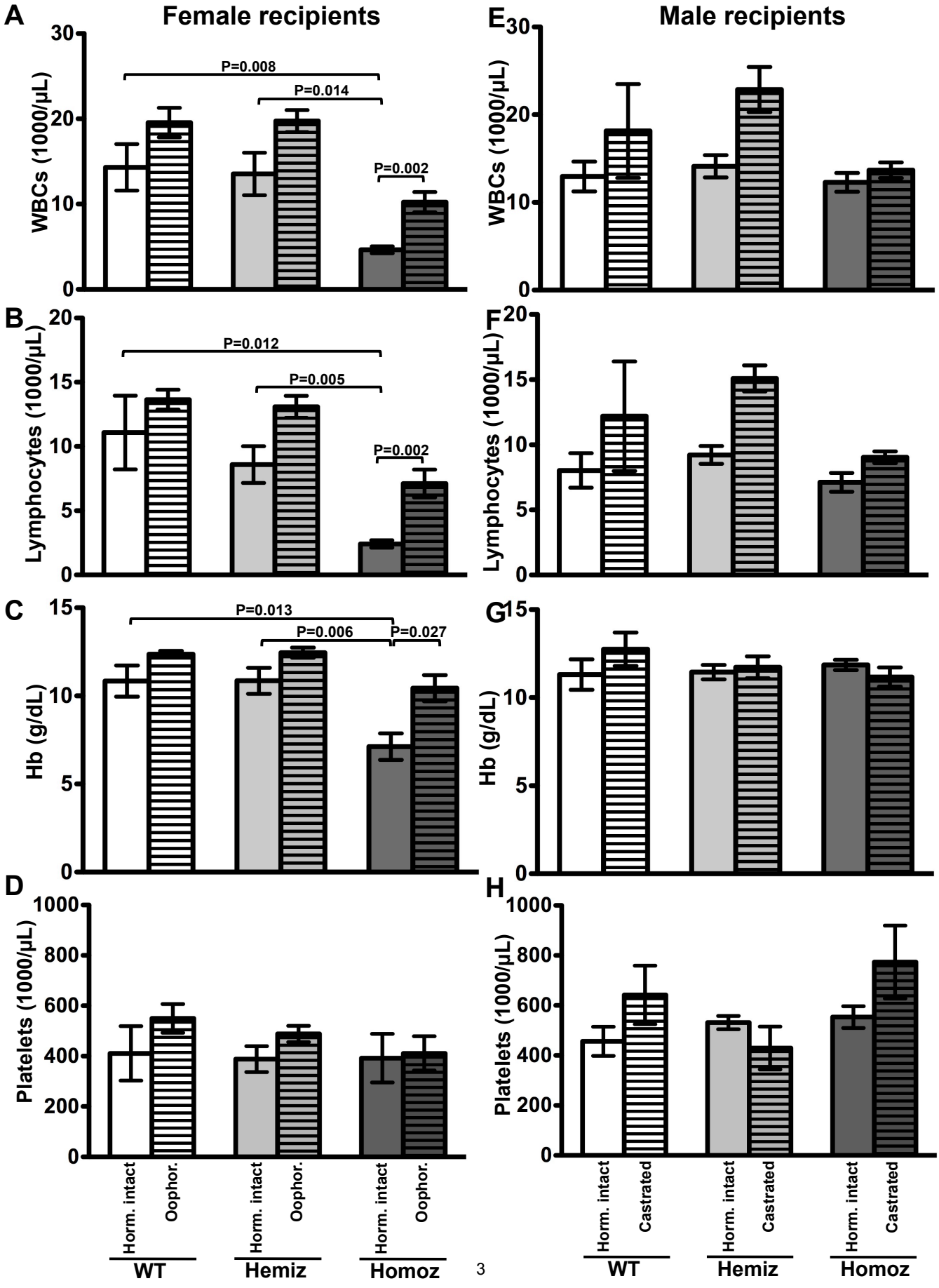
Supplemental Figure S1



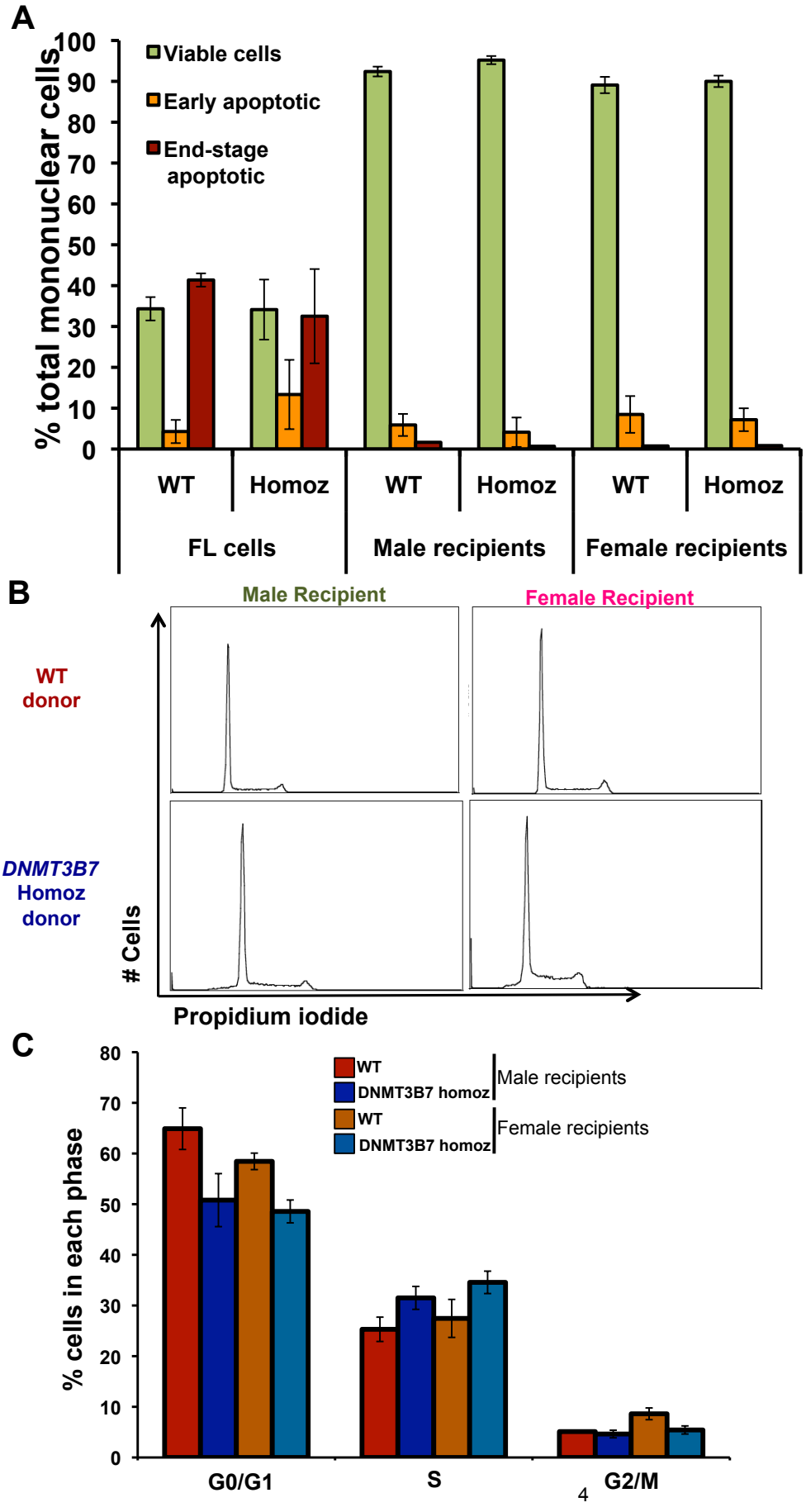
Supplemental Figure S2



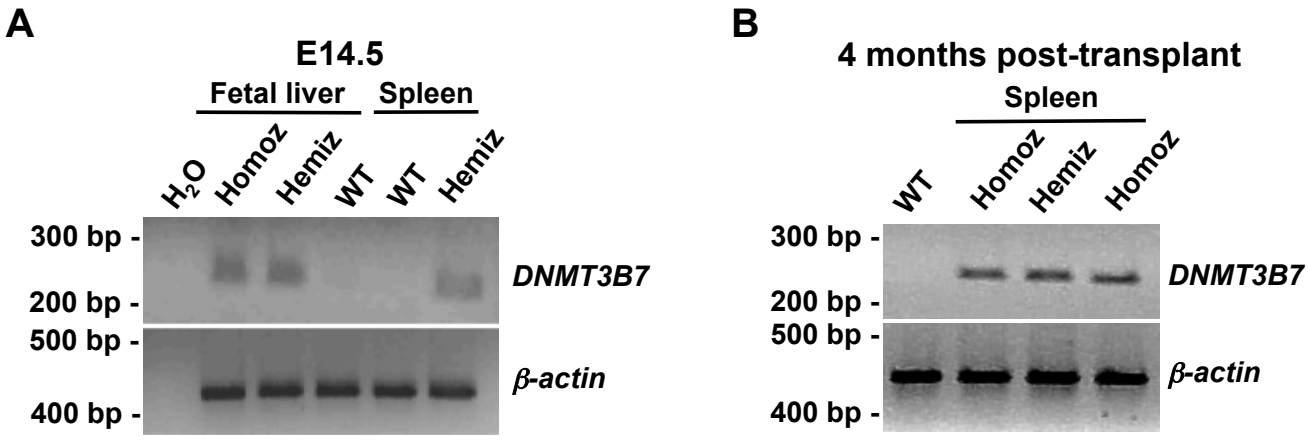
Supplemental Figure S3



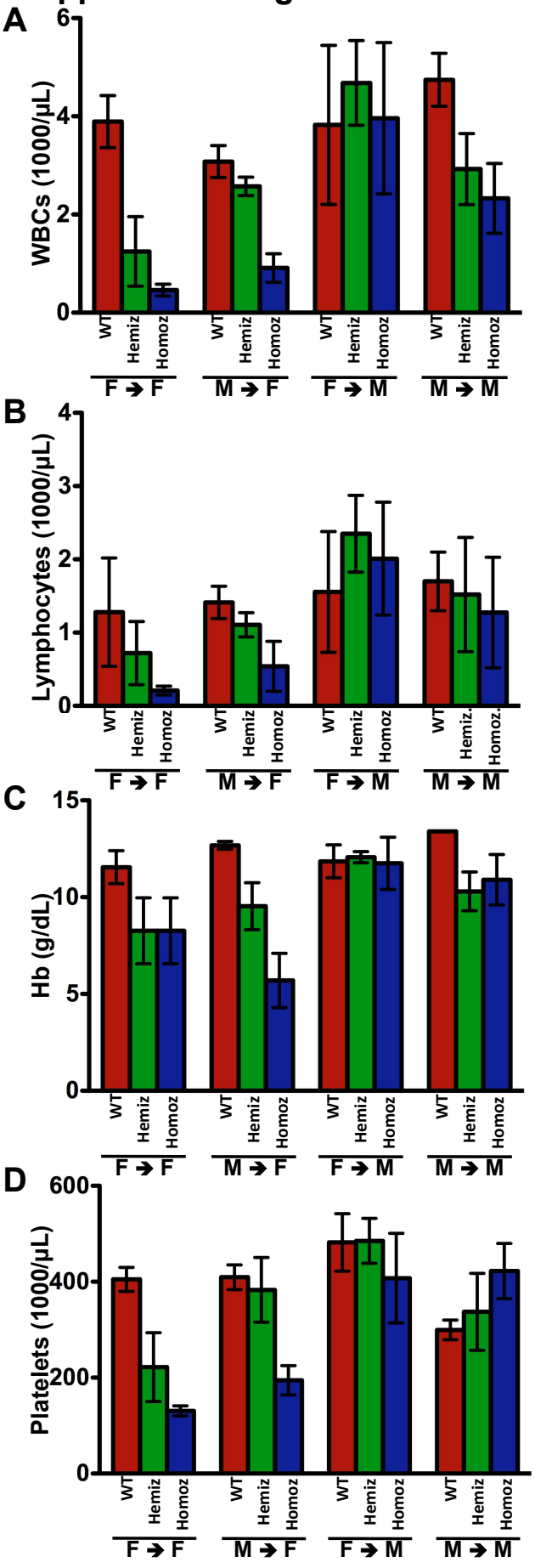
Supplemental Figure S4



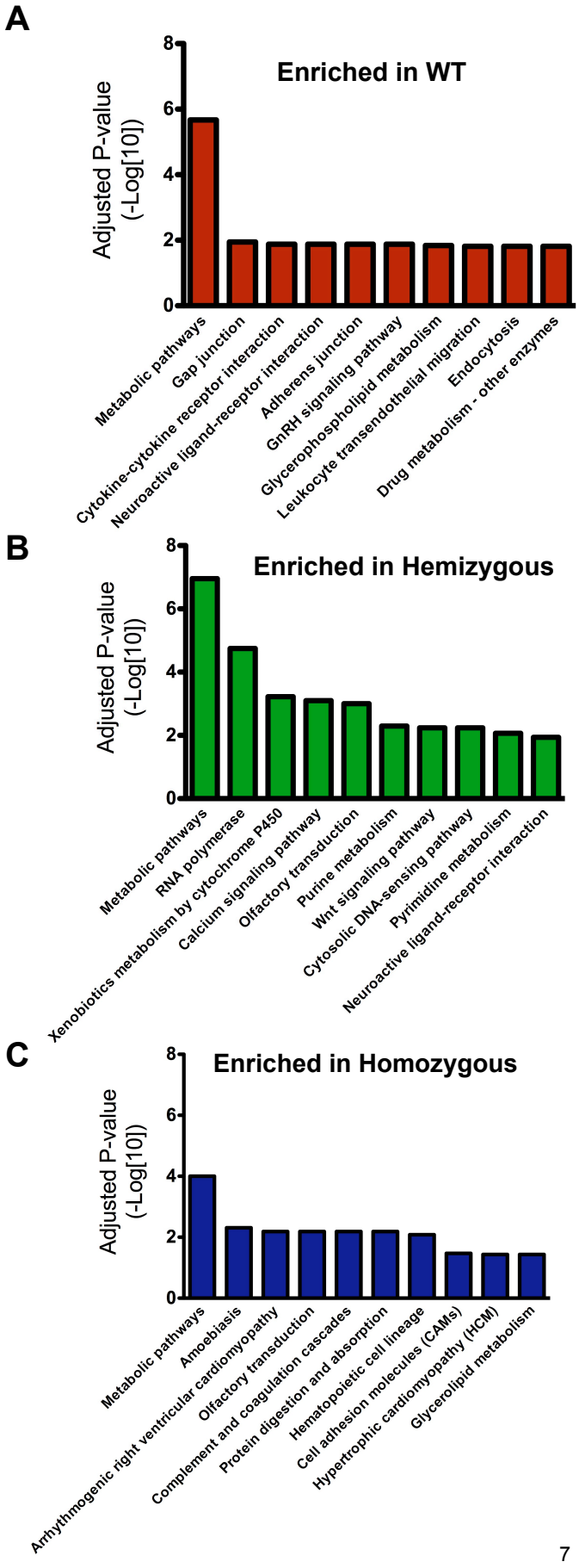
Supplemental Figure S5



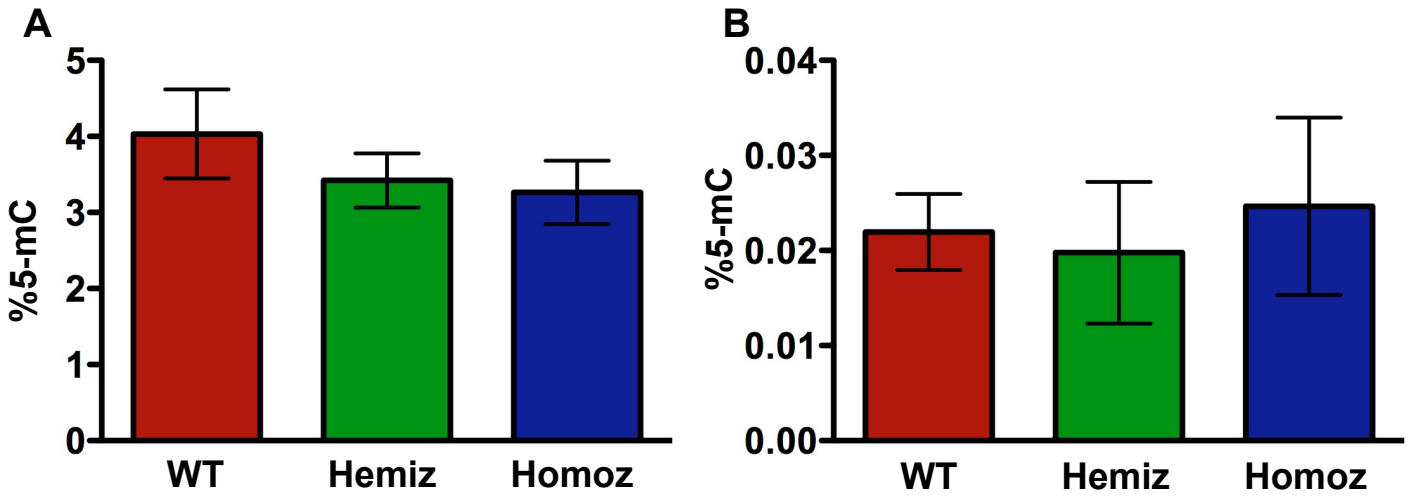
Supplemental Figure S6



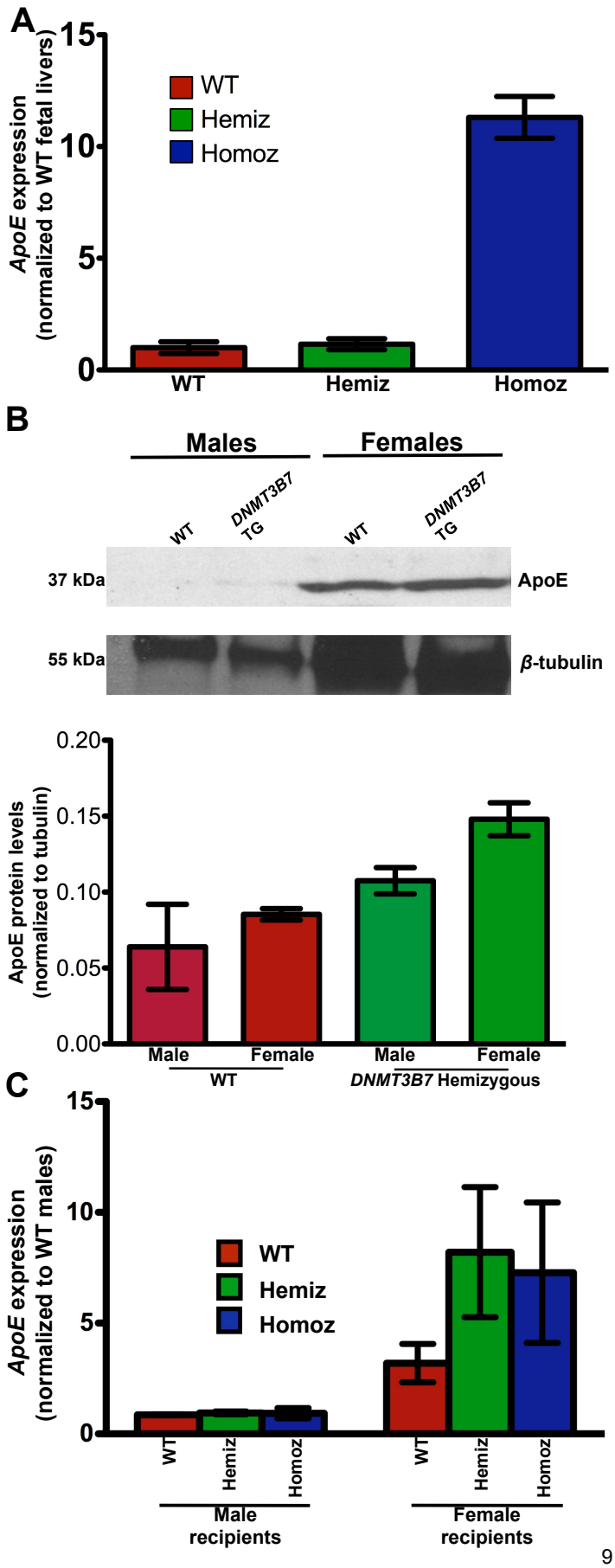
Supplemental Figure S7



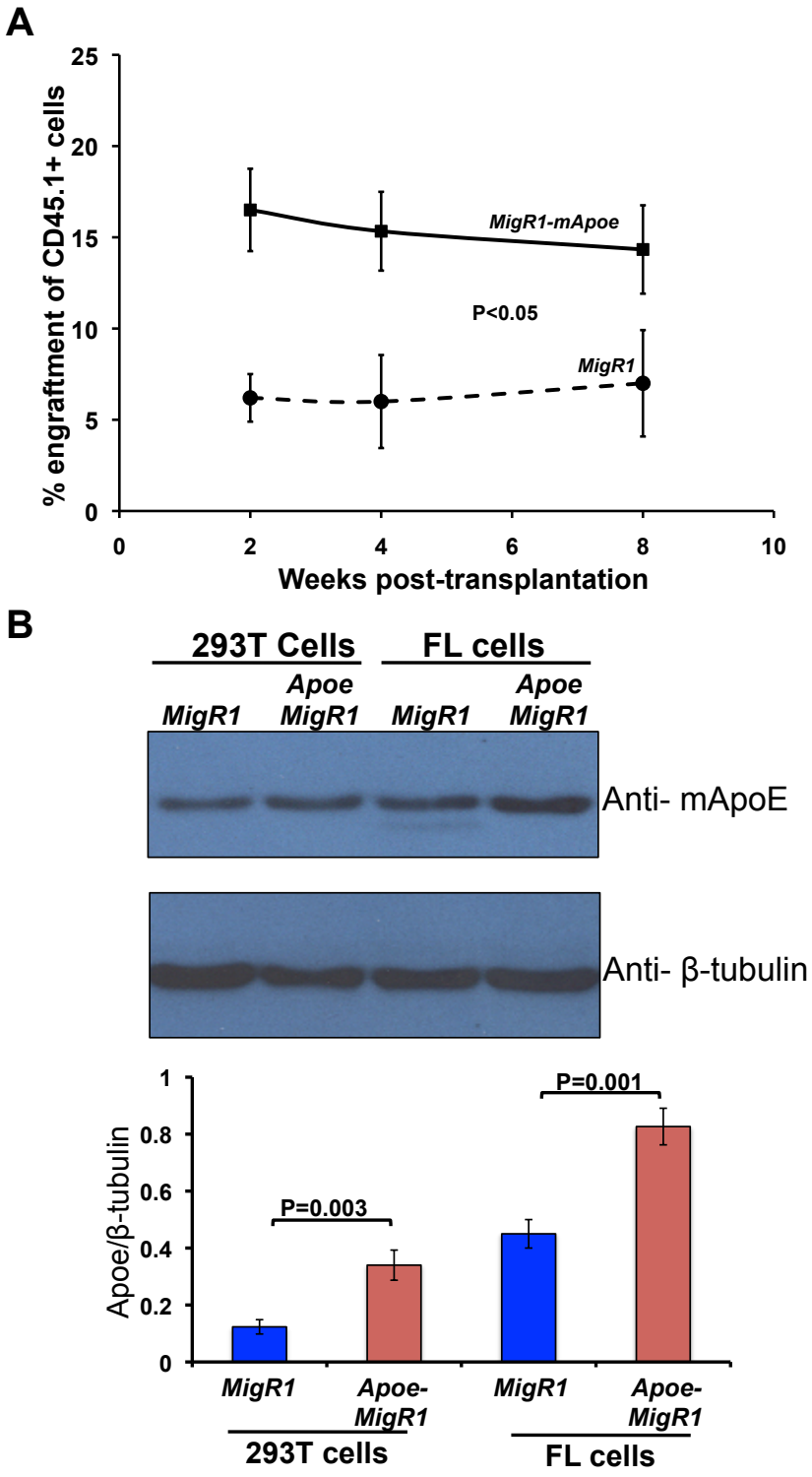
Supplemental Figure S8



Supplemental Figure S9

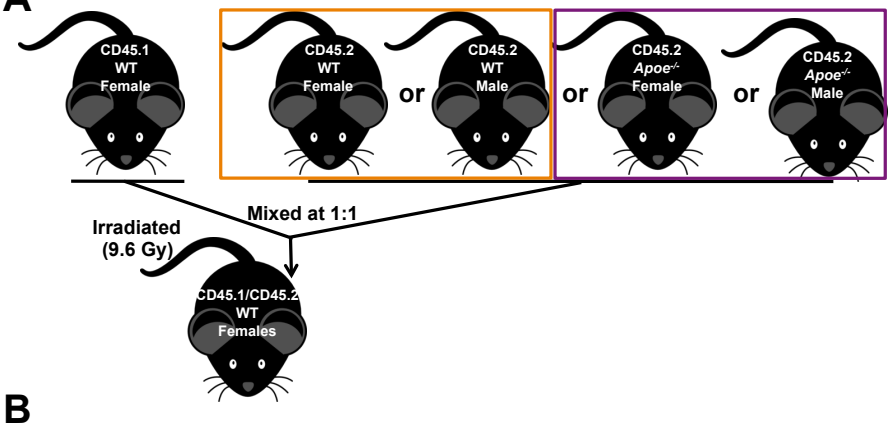


Supplemental Figure S10

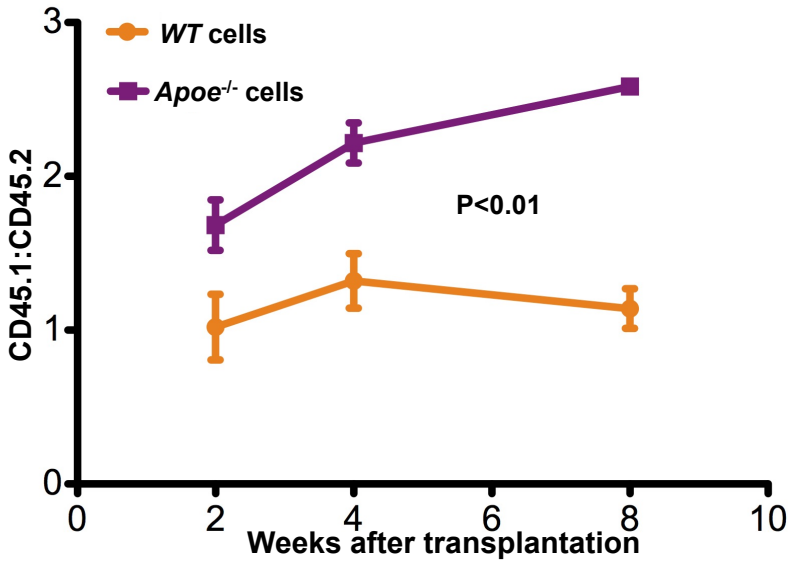


Supplemental Figure S11

A



B



Supplemental Figure Legends

Supplemental Figure S1. LT-HSC levels in *DNMT3B7* transgenic fetal livers and confirmation of DNMT3B7 effects by competitive transplantation using retrovirally transduced fetal liver cells.

(A) Lineage (FITC)-, CD48 (PE)-, CD150 (APC)+ cells were stained in fetal livers isolated from E14.5 embryos that were Wild-type (WT: red), hemizygous *DNMT3B7* transgenic (Hemiz.: green) and homozygous *DNMT3B7* transgenic (Homoz.: blue). Average percent total bone marrow mononuclear cells from n ≥ 4 per group ± SEM are plotted. Two-tailed Student's *t*-test was used to determine statistical significance.

(B) Percent engraftment of competitor CD45.1+ WT cells (green) and of experimental CD45.2+ fetal liver cells transduced with *MigR1* or *MigR1-DNMT3B7* (brown) two weeks post-transplant.

(C) Expression of *DNMT3B7* in spleens isolated from recipient female mice transplanted with fetal liver cells transduced with *MigR1* (left) or *MigR1-DNMT3B7* (right, orange).

Supplemental Figure S2. *Dnmt3b* knockout cells engraft optimally into normal female recipients.

(A) Percent engraftment of CD45.1-bearing WT competitor cells (dashed line) and CD45.2-bearing WT experimental cells (green solid line). (B) Percent engraftment of CD45.1-bearing WT competitor cells (dashed line) and CD45.2-bearing *Dnmt3b*$^{+/-}$ experimental cells (purple solid line). (C) Percent engraftment of CD45.1-bearing WT competitor cells (dashed line) and CD45.2-bearing *Dnmt3b*$^{-/-}$ experimental cells (orange solid line). Average percentages from n ≥ 4 per group ± SEM are plotted.

Supplemental Figure S3. Apoptosis and cell cycle profiles in *DNMT3B7* homozygous cells.

(A) Single cell suspensions of fetal livers (FL) of WT or *DNMT3B7* homozygous embryos (left), or of bone marrows from male (center) or female recipients of WT or *DNMT3B7* homozygous cells were assayed for apoptotic potential using the FITC-Annexin V Apoptosis Kit (BD biosciences). Viable cells are shown in green, FITC-Annexin V positive early apoptotic cells in orange, and FITC-Annexin V/PI double positive end-stage apoptotic cells are in brown. (B) Representative cell cycle profiles of bone marrow cells isolated from male (left) and female (right) recipients of WT (top) or DNMT3B7 homozygous (bottom) fetal liver cells. The Watson-Pragmatic model was used to fit the cell cycle data on FlowJo. (C) Quantitation of number of cells in the G1, S, G2/M phase of the cell cycle. Student's *t*-test was used to calculate biological significance.

Supplemental Figure S4. *DNMT3B7* transgenic cells do not engraft optimally into normal female recipients.

(A-D) Female and (E-H) Male recipients of wild-type (WT: light grey), hemizygous *DNMT3B7* transgenic (Hemiz: light grey) and homozygous *DNMT3B7* transgenic (Homoz: dark grey) cells at eight weeks post-transplantation. Ten million cells isolated from fetal livers of E14.5 embryos of the three genotypes were transplanted into hormonally intact (solid bars) or hormonally compromised (hatched bars) females (A-D) or males (E-H) that had been lethally irradiated at 960 rads. (A), (E) Average WBC number (x1000/μL); (B), (F) lymphocyte number (x1000/μL); (C), (G) Hb concentration (g/dL); and (D), (H) platelet number (x1000/μL) from n ≥

4 mice per genotype per gender are plotted ± SEM. Two-tailed Student's *t*-test was used to determine statistical significance.

Supplemental Figure S5. Expression of *DNMT3B7* in transgenic fetal livers and in spleens of transplanted mice 4 months post-transplant.

RNA isolated from (A) fetal livers from E14.5 embryos and (B) spleens from transplanted mice at 4 months post-transplant was used for semi-quantitative RT-PCR to detect the expression of *DNMT3B7* (top row), using *β-actin* (bottom row) as the loading control.

Supplemental Figure S6. Gender of donor cells do not influence the engraftment and reconstitution of HSPCs.

(A) WBC, (B), lymphocyte number, (C) Hb concentration and (D) Platelet number in female donor (left) and male donor (right) mice transplanted with cells obtained from WT (red), *DNMT3B7* hemizygous (green) and *DNMT3B7* homozygous (blue) fetal livers. Donor gender was determined by PCR for the sex determining region Y (*Sry*). *Sry*- cells are denoted as female and *Sry*+ cells as male. The gender of the donors and recipients are indicated at the bottom of each plot. n ≥ 3 mice per genotype per gender are plotted ± SEM

Supplemental Figure S7. Enrichment of genes in LSK cells isolated from WT, and *DNMT3B7* transgenic fetal livers from E14.5 embryos.

Gene lists generated from the analysis in Figure 4A were used for gene ontology analysis, and pathways enriched for in (A) WT, (B) *DNMT3B7* hemizygous, and (C) *DNMT3B7* homozygous cells are shown.

Supplemental Figure S8. Global DNA cytosine modifications in fetal livers and spleens isolated from male and female mice transplanted with WT or *DNMT3B7* transgenic cells. (A) %5-mC and (B) %5-hmC in WT (red), DNMT3B7 hemizygous (green) or DNMT3B7 homozygous (blue) fetal livers isolated at E14.5. Average percentages from n ≥ 3 per group ± SEM are plotted.

Supplemental FIgure S9. Expression of Apoe in pre-transplantation fetal livers and in post-transplantation hematopoietic organs (A) Expression of *Apoe* in WT, *DNMT3B7* hemizygous and *DNMT3B7* homozygous embryonic fetal livers at E14.5. Relative expression of *Apoe* was measured in all fetal livers by real-time PCR and plotted normalized to β-*Actin* levels. (B) Representative Western blot for Apoe (top) and β-tubulin (middle) and quantitation of the blots, normalized to intrinsic tubulin expression. (C) *Apoe* expression in bone marrow cells isolated from hormonally intact males (left), and hormonally intact females (right) receiving WT (red), *DNMT3B7* hemizygous (green) or *DNMT3B7* homozygous (blue) fetal liver cells. Relative expression was measured by real-time PCR and plotted normalized to β-*Actin* levels.

Supplemental Figure S10. ApoE overexpression compromises engraftment and increases the engraftment of competitor cells. (A) Percentage engraftment of CD45.1+ competitor cells in mice receiving *MigR1* (dashed line) or *MigR1-mApoe* (solid line) transduced cells. N=5 mice in *MigR1* group, N=6 mice in *MigR1-mApoe* group; ANOVA was used to calculate statistical significance. (B) Expression of mApoE in 293T cells (right) and FL cells

(left) transduced with *MigR1* or *MigR1-mApoe*. Quantification was performed relative to a loading control, β-actin.

Supplemental Figure S11. *Apoe* deficiency increases HSPC engraftment potential in female mice. (A) Experimental design for competitive transplantations of bone marrow isolated from WT or *Apoe*$^{-/-}$ mice. Bone marrows were isolated from male or female WT (light grey) or *Apoe*$^{-/-}$ (dark grey) mice (CD45.2) and mixed at a 1:1 ratio with competitor cells isolated from bone marrows of CD45.1 mice and transplanted into recipient female CD45.1/CD45.2 mice that had been irradiated with 9.6 Gy. (B) At two, four, and eight weeks post-transplantation, the mice were bled and donor repopulation measured by flow cytometry. CD45.2:CD45.1 ratios for mice receiving WT cells (orange) and *Apoe*$^{-/-}$ cells (purple) are plotted. ANOVA was used to assess statistical significance.

Supplemental materials and methods

Retroviral transduction and transplantation

The *DNMT3B7* construct was used as previously described [1]. Mouse cDNA was used to generate an *Apoe* construct (NCBI Ref#NM_009696.3) within the *MigR1* backbone. The above-described *MigR1* plasmids were co-transfected with *pcL-Eco* into 293T cells to generate the retroviral supernatant, which was then filtered through a 20 micron syringe filter, and spinoculation was used to infect wild-type CD45.2+ fetal liver cells in the presence of 8 μg/mL polybrene. After two rounds of spinoculation performed 24 hours apart, the cells were allowed to recover in complete media containing DMEM, IMDM, β-ME, 10 ng/mL each of mIL-3 and mIL-6 and 25 ng/mL of mSCF overnight. An aliquot of cells was used to measure engraftment of GFP+ cells. Two million total cells from the spinoculation were mixed with 0.5 million cells from CD45.1+ fetal liver cells before injection into recipient CD45.1/CD45.2 mice. Blood was collected from the tail vein at two, four, and eight weeks after transplantation, and percent engraftment measured at each time point by flow cytometry.

RNA-Sequencing (RNA-Seq).

Total RNA was isolated using Trizol Reagent (Invitrogen), and the integrity of the RNA was validated using the 2100 Bioanalyzer (Agilent). All samples had an RNA integrity number of at least 9 or greater. 1~10 μg of total RNA was used to isolate mRNA by polyA selection. The mRNA was then fragmented and randomly primed for reverse transcription, followed by second-strand synthesis to create double-stranded cDNA fragments. Ends of the cDNA fragments were repaired with a combination of fill-in reactions and exonuclease activity to produce blunt ends. An 'A'- base was added to the blunt ends followed by ligation to Illumina

sequencing adapters. cDNA fragments ranging from 300 to 500 bps were gel purified after the adaptor ligation step. Libraries were generated following the Illumina protocol for preparing samples for sequencing of mRNA. PCR amplified cDNA libraries were quantified on the Agilent 2100 Bioanalyzer and diluted to 10 pM for cluster generation and sequencing. Single end sequencing was performed for 50-cycles by using Single Read Cluster Generation Kit (TruSeq SR Cluster Kit v3 - cBot –HS, Cat# GD-401-3001) and Sequencing Kit (TruSeq SBS Kit v3– HS, Cat# FC-401-3002) on Illumina HiSeq 2000 machine. Sequence reads from RNA-Sequencing were aligned to genomic sequences.

RNA-Sequencing data mapping and analysis.

RNA-Sequencing analysis and quantitation was performed as described previously [2]. Briefly, we first obtained the Ensembl mouse transcript database, consisting of 23,148 genes from Ensembl (www.ensembl.org). For cases in which multiple transcript entries were available for a single gene, we retained the transcript with the highest homology to transcripts within the NCBI RefSeq database (downloaded from genome.ucsc.edu) as determined by a blastn search so that each gene possessed a single representative transcript. We next mapped these ‘cleared’ sequences to described database by using Tophat (version 1.4.1) [3] with default parameters to align the gene expression data for each genotype to the NCBI reference mouse genome sequence (Build 37). Assembly of the transcripts and gene expression analysis was performed using Cufflinks (version 1.3) [4]. BEDTools [5] was used to annotate the transcripts with the gene name according to the RefSeq gene annotations. BEDTools “intersectBed” finds the intersection between two BED files (e.g. the transcripts BED file generated by Cufflinks and the RefSeq genes BED file downloaded from the UCSC Genome Browser).

Analysis and Visualization of RNA-Sequencing data.

FPKM values for genes that are expressed in the fetal LSK cells of WT, *DNMT3B7* hemizygous, and *DNMT3B7* homozygous mice were generated by RNA-seq. The FPKM values for specific genes are analogous to the transcriptional expression levels of those genes in each genetic background. Thus, each genotype has an associated list of genes with corresponding FPKM values that reflect that gene's expression in the FL LSK cells of that genotype. To compare gene expression among the three genetic backgrounds, an adjusted list of genes was generated for each genotype that contains the genes that are uniquely expressed or 2-fold overexpressed in that genotype.

RNA-seq had generated lists for each genotype that contained all genes with detectable FPKMs. To generate lists that only contained genes that are uniquely expressed or overexpressed in a particular genotype:

1. Genes that are lowly expressed (genes with FPKM < 1) in a particular genotype were removed from that genotype's list.
2. Genes that were at least 2-fold overexpressed in a specific genotype compared to the other genotypes were removed from the lists of the other genotypes and retained in the specific genotype's list.
3. If two genetic backgrounds similarly overexpressed a specific gene by 2-fold compared to the third genotype, that gene was removed from the list of the third genotype.

In this manner, a list was generated for each genetic background that contains the genes that are expressed in that genotype and excludes the genes that are comparatively lowly expressed in that genotype. To visualize these lists and compare the genes that are

expressed in each genotype, a 3-way venn diagram was generated using BioVenn [6]. From the venn diagram, lists of genes that are uniquely expressed or overexpressed in a particular genotype were analyzed using GO.

Analysis of global DNA methylation and hydroxymethylation.

DNA hydrolysis was performed as previously described [7], with minor modifications. Briefly, one microgram of genomic DNA was first denatured by heating at 100°C. Five units of Nuclease P1 (Sigma) were added, and the mixture was incubated at 45°C for 1 hour. 1/10 volume of 1M ammonium bicarbonate and 0.002 units of venom phosphodiesterase 1 (Sigma) were added to the mixture, and the incubation continued for 2 hours at 37°C. Next, 0.5 units of alkaline phosphatase (Invitrogen) were added, and the mixture was incubated for 1 hour at 37°C. Before injection into the Zorbax XDB-C18 2.1 mm x 50 mm column (1.8 μm particle size) (Agilent 927700-902), the reactions were diluted 10-fold to decrease the concentrations of the salts and enzymes. An Agilent 1200 Series liquid chromatography machine in tandem with the Agilent 6410 Triple Quad Mass Spectrometer was employed to detect 5-mC and 5-hmC. LC separation was performed at a flow rate of 220 μL/min. Quantification was done using an LC-ESI-MS/MS system in the multiple reaction monitoring (MRM) mode.

Supplemental Table S1. Primers used for bisulfite sequencing and real-time PCR

Bisulfite Sequencing	**Primer Sequence**	**Tm**
mApoe	Fwd: 5'- GGA GTT GGA GGA ATA GTT GGG – 3'	60 C
	Rev: 5'-TAT CTC CTC TAT ACT CTA ACC CAA - 3'	
Real-time/ RT-PCR	**Primer Sequence**	
mApoe	Fwd: 5'-GGA CTT GTT TCG GAA GGA GC-3'	60 C
	Rev: 5'-ACA TTG CTG ACA GGA TGC CT-3'	
mβ-Actin	Fwd: 5'-GGC ATT GTT ACC AAC TGG GAC G-3'	54 C
	Rev: 5'-TTT GAT GTC ACG CAC GAT TTC C-3'	
hDNMT3B7	Fwd: 5'-AAA CCC AAC AAC ACG CAA C-3'	60 C
	Rev: 5'- GGT AGG AGG GTC CAG AGA G-3'	
mApoB100	Fwd: 5'-GCA CGT GGG CTC CAG-3'	60 C
	Rev: 5'-GTC ATT TCT GCC TTT GCG-3'	
mSRY	Fwd: 5'- AGATCTTGATTTTTAGTGTTC-3'	52 C
	Rev: 5'- GAGTACAGGTGTGCAGCTCTA-3'	
mDnmt3b	Common Fwd: 5'- TCACAGGGTACTTGGTGCTCAAGGA-3'	60 C
	WT Rev: 5'-TACCTCCAACTGCCCTGTTTGCACT-3'	
	KO Rev: 5'- ACACTCCAACCTCCGCAAACTCCTA-3'	

Supplemental Table S2. Categorization of HSPC percentage by gender and genotype

Gender	Genotype	Lineage-Sca1+c-Kit+ cell number (average±SD)	CD48+CD150- cell number (average±SD)	CD48-CD150+ cell number (average±SD)
Male	WT	1.45 ± 0.24	14.40 ± 2.40	0.30 ± 0.09
	DNMT3B7 hemiz.	1.23 ± 0.22	12.73 ± 4.00	0.54 ± 0.17
	DNMT3B7 homoz.	0.53 ± 0.11	4.92 ± 3.39	0.38 ± 0.16
Female	WT	1.39 ± 0.27	12.80 ± 1.57	0.57 ± 0.21
	DNMT3B7 hemiz.	1.17 ± 0.39	11.71 ± 2.95	0.65 ±0.29
	DNMT3B7 homoz.	0.49 ± 0.10	9.32 ± 0.69	0.60 ± 0.19

Supplemental References

1. Shah MY, Vasanthakumar A, Barnes NY, et al. DNMT3B7, a truncated DNMT3B isoform expressed in human tumors, disrupts embryonic development and accelerates lymphomagenesis. **CANCER RES.** 2010;70(14):5840–5850.

2. Madzo J, Liu H, Rodriguez A, et al. Hydroxymethylation at Gene Regulatory Regions Directs Stem/Early Progenitor Cell Commitment during Erythropoiesis. **CELL REP** 2014;6(1):231–244.

3. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. **NATURE BIOTECHNOLOGY** 2009.

4. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **NATURE BIOTECHNOLOGY** 2010;28(5):511–515.

5. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. **BIOINFORMATICS** 2010;26(6):841–842.

6. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. **BMC GENOMICS** 2008;9:488.

7. Song L, James SR, Kazim L, et al. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. **ANAL. CHEM.** 2005;77(2):504–510.