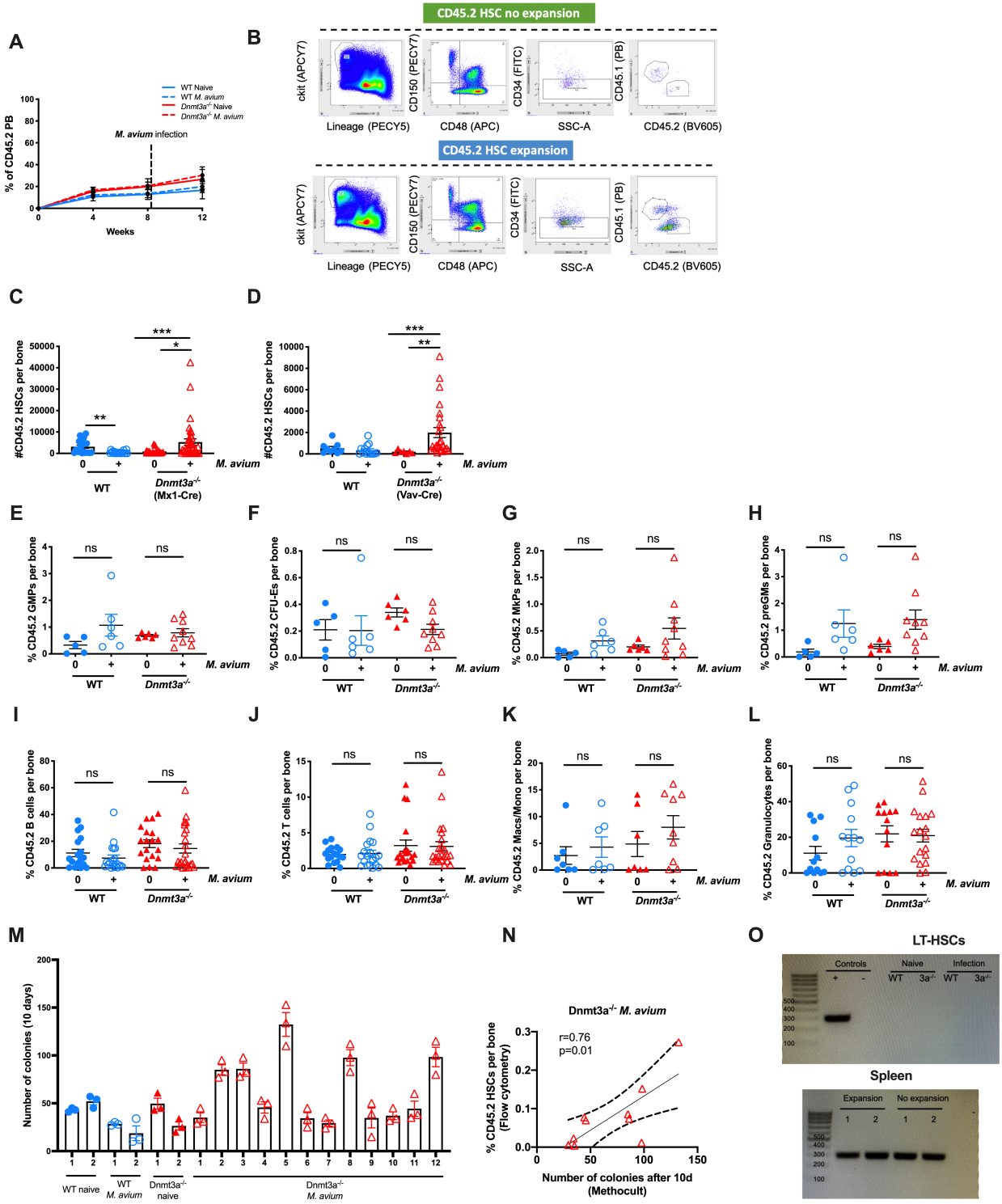
Supplementary Information

Table of Contents

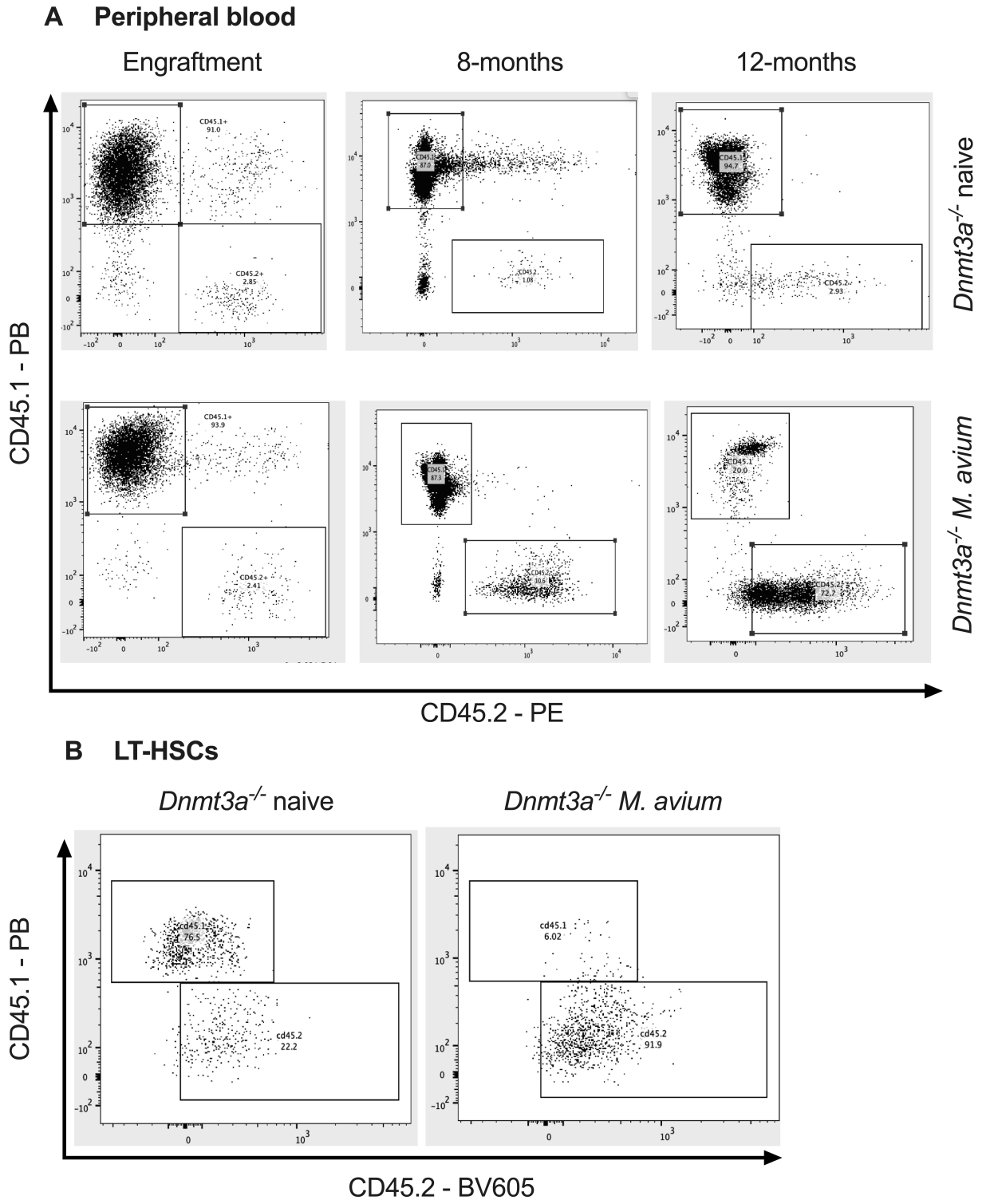
Supplemental Figure S1



Supplemental Figure S1: *M. avium* infection promotes *Dnmt3a*$^{-/-}$ clonal expansion in mice, related to Figure 1. (A) Analysis of percentage of CD45.2 blood cells in mosaic mice at different timepoints. Data shown are representative of others (n=3 independent experiments). (B) Representative flow plots showing the gating strategy using to identify CD45.2 and CD45.1 HSCs. (C-D) Absolute number of CD45.2 HSCs (KL CD150+ CD48- CD34-) per bone that were originally from Mx1-Cre *Dnmt3a*$^{fl/fl}$ donors (C) or Vav-Cre *Dnmt3a*$^{fl/fl}$ donors (D). n=5-20 per

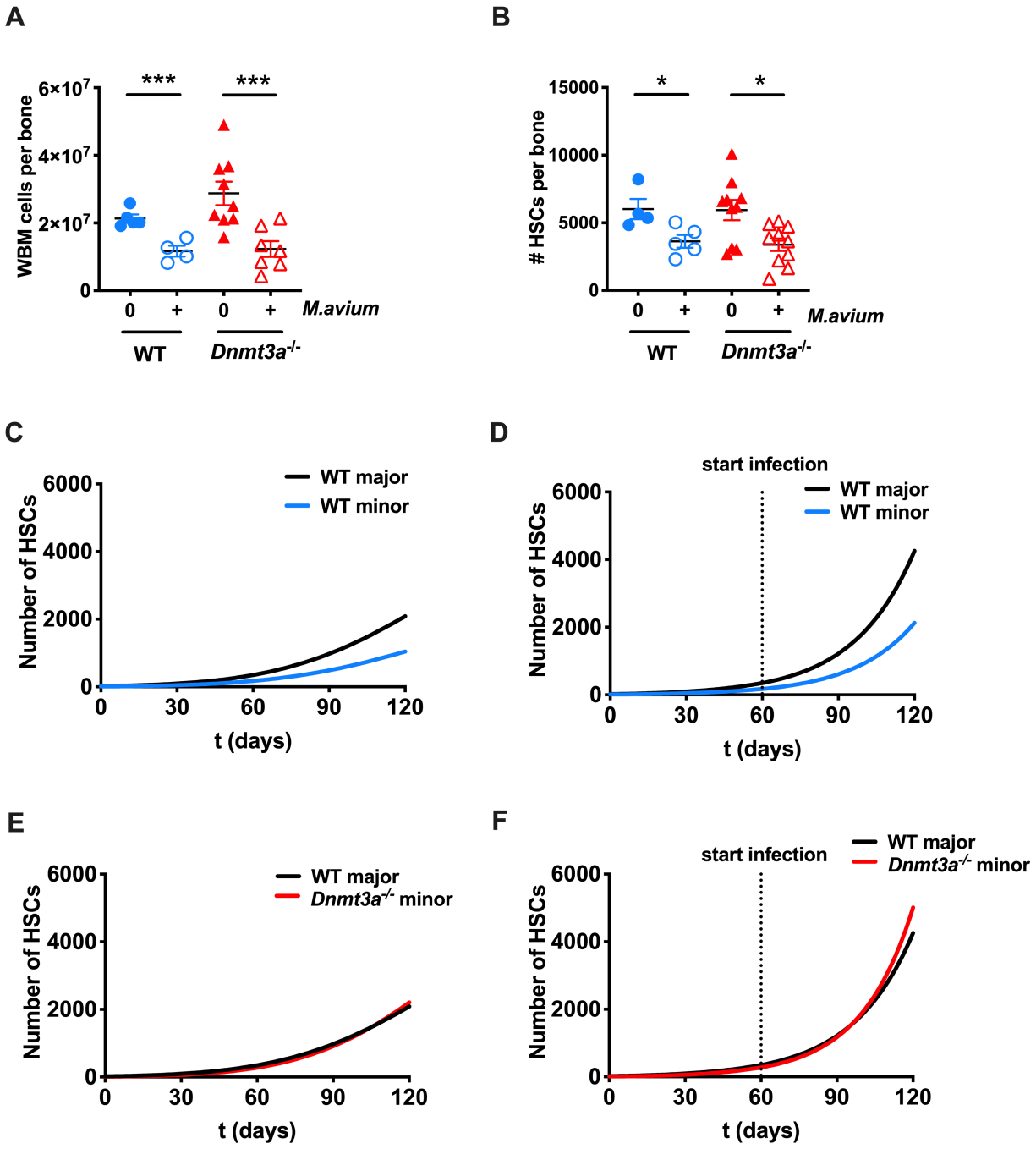
group; error bars, mean ± SEM. P values calculated by Kruskal-Wallis test. (E-H). Percentage of CD45.2 GMPs, CFU-Es, MKPs, and preGMs of WBM in mosaic mice. Error bars, mean ± SEM. *P* values calculated by Kruskal-Wallis test. (I-L) Percentage of CD45.2 B cells, T cells, Macs/Mono and Granulocytes of WBM in mosaic mice. Error bars, mean ± SEM. *P* values calculated by Kruskal-Wallis test. (M) CD45.2 WBM cells from a representative of each group were sorted at the end of the experiment and plated in Methocult. The number of colonies counted after 10 days of culture per individual mouse are presented in the graph. (N) Correlation between the number of CD45.2 *Dnmt3a*$^{-/-}$ WBM-derived colonies from infected mice after 10 days of culture and the percentage of CD45.2 HSCs per bone found by flow cytometry for same individual mouse. (O) PCR to detect 16S *M. avium* gene in LT-HSCs (top) or spleen (bottom) from 2 mice with *Dnmt3a*$^{-/-}$ HSC expansion after infection and 2 mice without expansion after infection. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Supplemental Figure S2



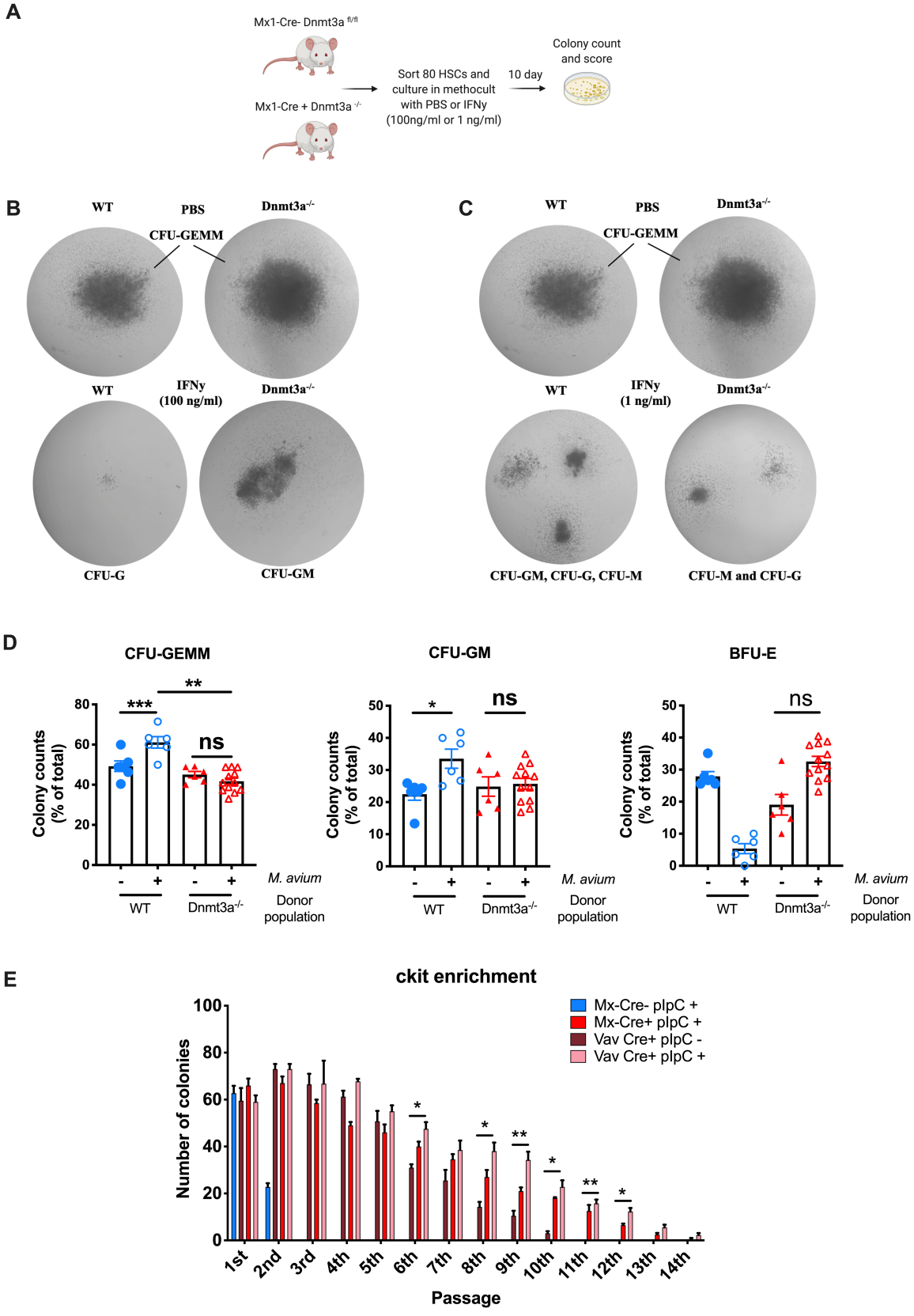
Supplemental Figure S2: *M. avium* infection promotes *Dnmt3a*$^{-/-}$ PB clonal expansion in mice, related to Figure 1 and 2. (A) Representative flow plots showing the gating strategy using to identify CD45.2 *Dnmt3a*$^{-/-}$ PB leukocytes and CD45.1 WT PB leukocytes in mosaic mice at engraftment and after 8-month and 12-month post-infection with *M. avium* or without infection (naïve). (B) Representative flow plots showing the gating strategy using to identify CD45.2 *Dnmt3a*$^{-/-}$ HSCs and CD45.1 WT HSCs in mosaic mice naïve or after 12-month post-infection with *M. avium*.

Supplemental Figure S3



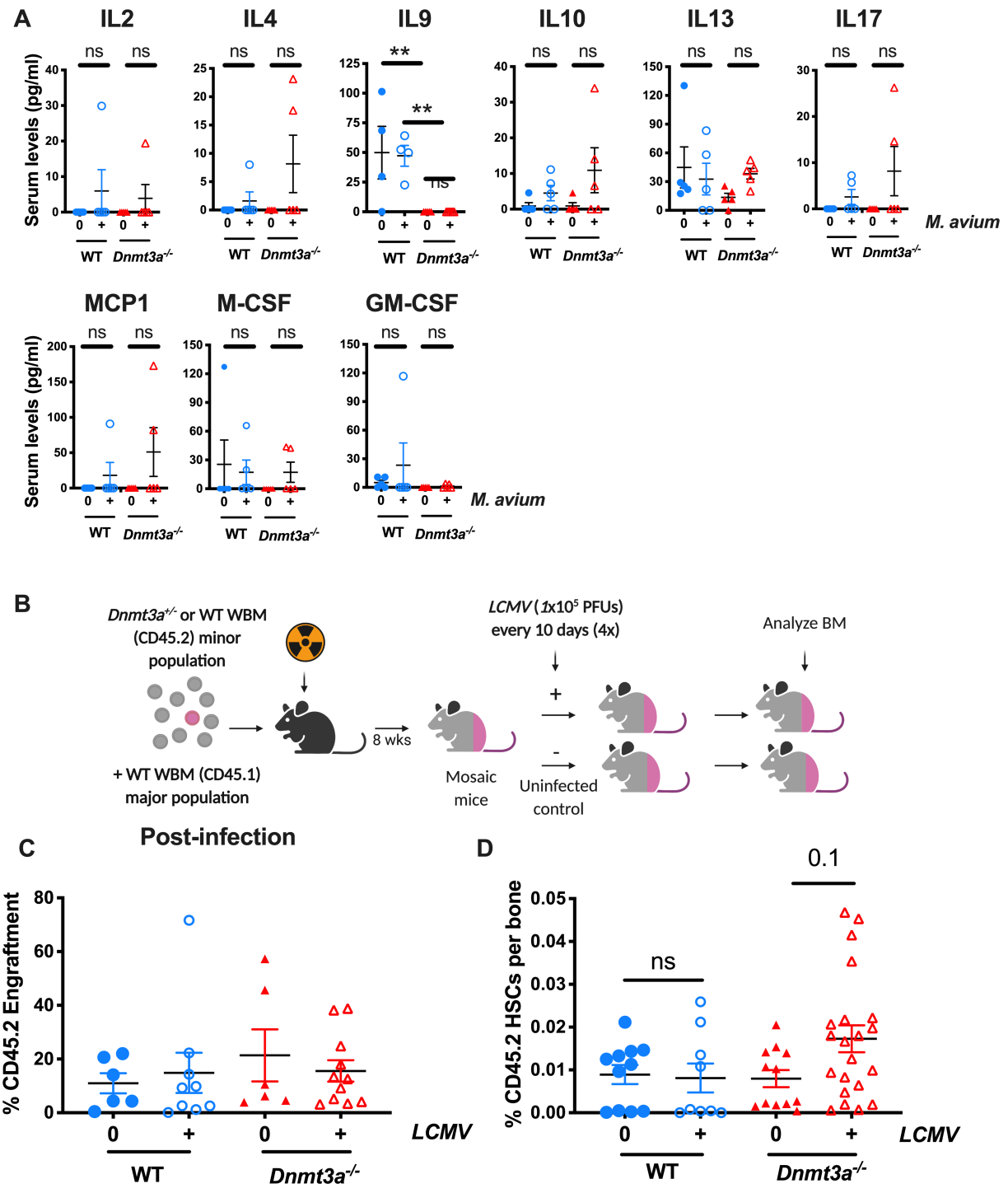
Supplemental Figure S3. Hematological responses of Dnmt3a-mutant mice to infection with *M. avium* and mathematical modeling, related to Figure 3. (A) WBM number per bone in WT and *Dnmt3a*$^{-/-}$ mice after one month of infection. (B) Absolute number of HSCs (KL CD150+ CD48- CD34-) per bone in WT and *Dnmt3a*$^{-/-}$ mice after one month of infection. n=3-5 per group; data are representative of 5 independent experiments. Error bars indicate mean ± SEM. P values calculated by two-sided t-test. (C-F) Predictive models of a major (WT) and minor (WT or *Dnmt3a*$^{-/-}$) HSC population after transplant in naïve mice (C, D) or mice infected at day 60 (E, F). For additional details on modeling, please see Part 2 of Supplemental Information. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Supplemental Figure S4



Supplemental Figure S4. **Mouse differentiation assay and serial replating assay to compare the differentiation and self-renew capacity of Dnmt3a$^{-/-}$ cells after inflammatory stress**, **related to Figure 4**. (A) Mouse differentiation assay scheme. 80 LT-HSCs (LSK CD150+ CD48-) were sorted from the pool of several *Dnmt3a*$^{+/+}$or *Dnmt3a*$^{-/-}$ mice and plated for 10 days in methocult medium in the presence or absence of IFNy. (B-C). Morphology of CFU-colonies after culture in methocult and treatment with high dose of IFNy (100 ng/ml) or low dose (1 ng/ml). Refer to Figure 4A-C. (D). CD45.2 WBM cells from individual mouse from the 4 different mosaic mice groups were sorted at the end of the experiment and plated in methocult; the morphology of colonies observed after 10 days of culture per each individual mouse was scored and is presented in the graph (E) Briefly, 500 ckit+ cells were plated on Methocult media completed with SCF, IL6, TPO and FLT3-L cytokines and the number of colonies were counted after 1 week of culture. Then colonies were harvested, and 10.000 cells were replated again for 1 week, counted and replated again until no colonies were observed. Error bars, mean ± SEM. *P* values calculated by two-way ANOVA. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

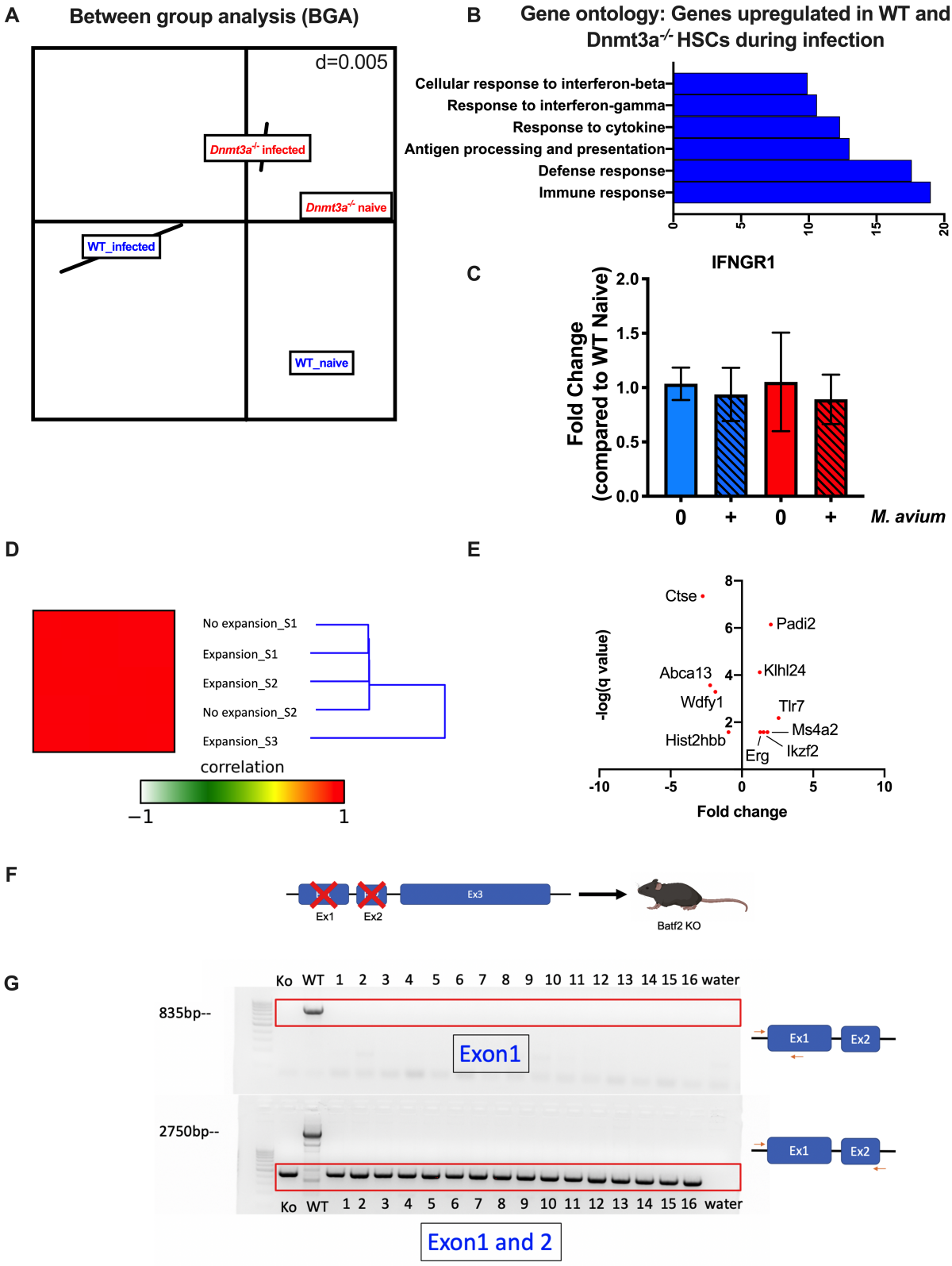
Supplemental Figure S5



Supplemental Figure S5. Cytokine profile and *LCMV* infections in mosaic mice, related to Figure 5. (A) Serum cytokine levels of IL2, IL4, IL9, IL10, IL13, IL17, MCP1, M-CSF and GM-CSF in WT and *Dnmt3a*$^{-/-}$ (Mx1-Cre) naïve and after one month of *M. avium* infection. n=5 per group. P values calculated by ordinary one-way ANOVA or Kruskal-Wallis test. (B) Mice were transplanted with a minor population or either Cre- *Dnmt3a*$^{fl/fl}$ CD45.2 WBM or Cre+ Dnmt3a$^{-/-}$ CD45.2 WBM + and a major population of competitive CD45.1 WBM. Two months after

transplant, mice were infected with 100000 PFUs of LCMV 4 times every 14 days. (C) Percentage of CD45.2 cells after LCMV infection. (D) CD45.2 HSCs (KL CD150+ CD48- CD34-) shown as a percentage of WBM in mosaic mice infected with LCMV. Data represent two independent experiments with n=20-40 per group. Error bars, mean ± SEM. *P* values calculated by Kruskal-Wallis test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

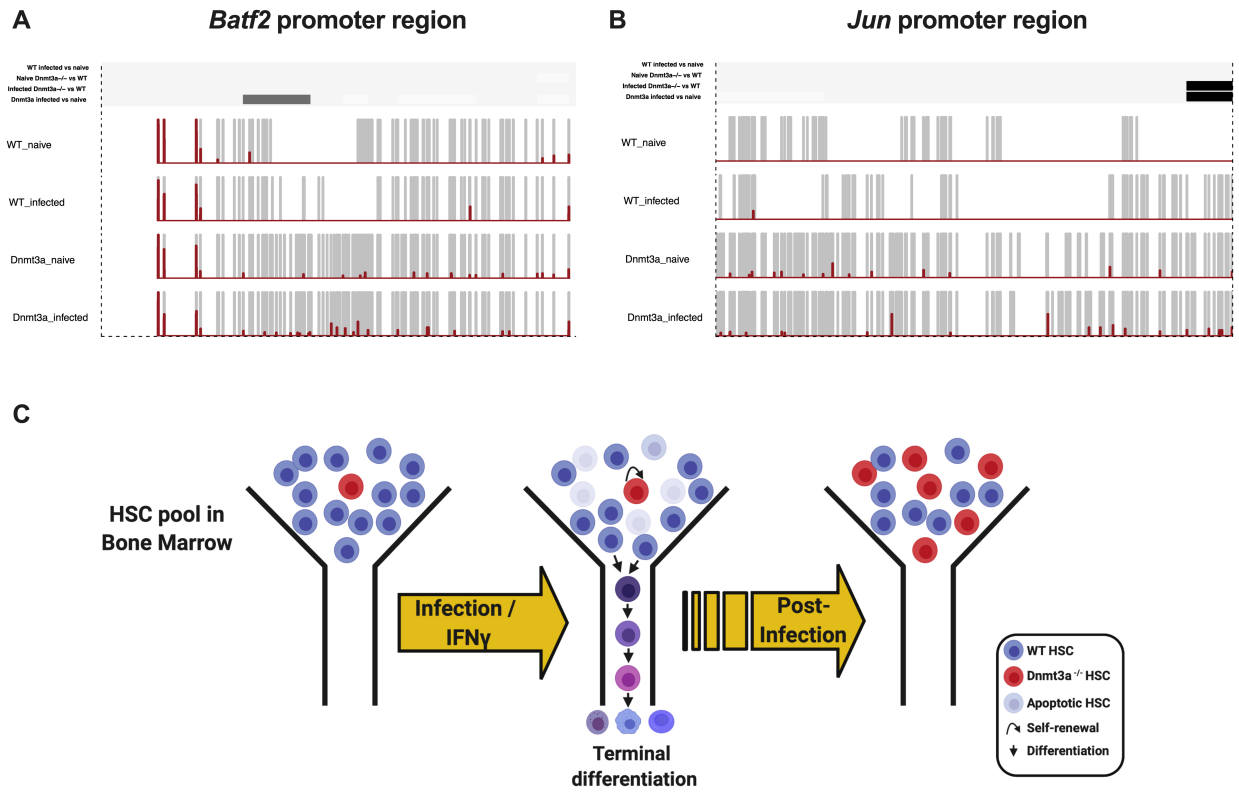
Supplemental Figure S6



Supplemental Figure S6: Transcriptome analysis of HSCs and progenitor cells from WT and *Dnmt3a*$^{-/-}$mice

after *M. avium* infection and generation of Batf2-KO mice, related to Figure 1 (Panels S6D and S6E) and Figure 6 (Panels S6A-C and S6E). (A) Between Group Analysis (BGA) was used to measure the relative distance between RNA samples obtained by projecting them into two-dimensional space using correspondence analysis (CoA). (B) Gene ontology analysis of pathways enriched from genes upregulated in both WT and *Dnmt3a*$^{-/-}$ HSCs after one month of infection. (C) *Ifngr1* expression on HSCs measured by qPCR. Error bars, mean ± SEM. *P* values calculated by Kruskal-Wallis test. (D). Clustering of RNA-seq samples in LSK cells from mosaic mice with or without *Dnmt3a*$^{-/-}$ HSC expansion after two months of infection with *M. avium*. **Related to Figure 1B**. (E) Volcano plot to show the genes significantly expressed in LSK cells between mosaic mice with or without *Dnmt3a*$^{-/-}$ *HSC* expansion after two months of infection with *M. avium.* **Related to Figure 1B**. (F) Batf2-KO mice were generated using CRISPR gene editing to remove exons 1 and 2 of Batf2 in a C57BL/6 background. Heterozygous mice from the initial transfer were outcrossed to WT mice and F1 pups were then crossed to generate homozygous mice. (G) Genotyping was determined by PCR, sequencing (not shown), and qPCR (not shown).

Supplemental Figure 7



Supplemental Figure S7: Promoters of pro-differentiation genes are hypermethylated in Dnmt3a$^{-/-}$ HSCs in infection, related to Figure 7. (A) DNA methylation profile of *Batf2* promoter region in WT or Dnmt3a $^{-/-}$ HSCs naïve or after 1-month of infection. Significant DMRs (p value < 0.05) between groups are indicated on the top of the graph. (B) DNA methylation profile of *Jun* promoter region in WT or Dnmt3a $^{-/-}$ HSCs naïve or after 1-month of infection. Significant DMRs (p value < 0.05) between groups are indicated on the top of the graph (C) Summary of results supporting that chronic infection promotes *Dnmt3a* loss of function clonal hematopoiesis. In the presence of *M. avium* infection, WT HSCs divide and differentiate or die as a result of stress-induced apoptosis, ultimately resulting in depletion of the HSC compartment. In contrast, *Dnmt3a*$^{-/-}$ HSCs are relatively agnostic to the differentiation and proapoptotic cues of infection, allowing them to survive better than the WT HSC population. Over time, the number of *Dnmt3a*-loss of function HSCs overtakes the number of WT HSCs.

Table S1: List of primers used in qPCR validations, related to Figure 6.

Gene name	Forward primer	Reverse primer
Jun (mouse)	TGGGTGCCAACTCATGCTAA	TCTGTCGCAACCAGTCAAGT
Junb (mouse)	TTGATCGTCCCCAACAGCAA	TGACAAAACCGTCCGCAAAG
Jund (mouse	TCGACATGGACACGCAAGAA	GAGGGTCTTGACTTTCTCCTCC
Fos (mouse)	CTCCAAGCGGAGACAGATCAA	GCAGACCTCCAGTCAAATCCA
Fosb (mouse)	GACATGCCAGGAACCAGCTA	TCTTCTTCTGGGGTAAGTGTCTC
Fosl (mouse)	GCTTTATCGCCTCAAGCCAAG	GTCTCAGGAGTGTCCAACCAG
Nr4a1 (mouse)	CTGCGAAAGTTGGGGGAGT	TGAGCTTGAATACAGGGCATCT
Egr1 (mouse)	CTGACCACAGAGTCCTTTTCT	AAGCGGCCAGTATAGGTGATG
Socs3 (mouse)	GGAGATTTCGCTTCGGGACT	CTCGCTTTTGGAGCTGAAGG
Batf2 (mouse)	GCAACTCTGTGGGAGTAGCG	CCACTCGGTTCTTCTGCTTCT
Stat1 (mouse)	,GGAAGGGGCCATCACATTCA	TTCTCGGCAGCCATGACTTT
Gbp2 (mouse)	CAAGACTCTGTGTGGTGGCA	GTCAGCACAGCACTCTCCAT;
Bst2 (mouse)	TCAGGAGTCCCTGGAGAAGAA	AGAAGTCTCCTTTTGGATCCTCAG
18s (mouse)	GTAACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG

Table S3: RNAseq *Dnmt3a*$^{-/-}$ infected vs *Dnmt3a*$^{-/-}$ naïve, related to Figure 6.

Gene.name	logFC	P.Value	adj.P.Val	Gene.description
H2-Ab1	2.86	2.18E-17	1.65E-13	histocompatibility 2, class II antigen A, beta 1
Serpina3g	2.01	3.38E-17	1.65E-13	serine (or cysteine) peptidase inhibitor, clade A, member 3G
Cd74	2.75	4.69E-17	1.65E-13	CD74 antigen
Iigp1	2.92	4.81E-17	1.65E-13	interferon inducible GTPase 1
Gbp2	2.82	5.31E-16	1.46E-12	guanylate binding protein 2
Gm1966	2.57	2.49E-15	5.69E-12	predicted gene 1966
H2-Eb1	3.06	4.73E-15	9.26E-12	histocompatibility 2, class II antigen E beta
Gbp3	2.67	2.34E-14	4.01E-11	guanylate binding protein 3
H2-Aa	2.34	8.63E-14	1.32E-10	histocompatibility 2, class II antigen A, alpha
Gm8995	1.40	1.49E-13	2.04E-10	predicted gene 8995
Gbp7	1.89	1.99E-13	2.48E-10	guanylate binding protein 7
Igtp	2.84	5.94E-13	6.79E-10	interferon gamma induced GTPase
Gbp5	2.30	4.72E-12	4.98E-09	guanylate binding protein 5
Gm12185	2.43	2.32E-11	2.27E-08	predicted gene 12185
Nr4a1	-1.18	2.56E-11	2.34E-08	nuclear receptor subfamily 4, group A, member 1
Irgm1	2.41	4.51E-11	3.86E-08	immunity-related GTPase family M member 1
Irf1	1.69	1.20E-10	9.71E-08	interferon regulatory factor 1
Stat1	2.42	5.40E-10	4.12E-07	signal transducer and activator of transcription 1
Gm12250	2.76	6.80E-10	4.91E-07	predicted gene 12250
Serpina3f	4.19	7.84E-10	5.38E-07	serine (or cysteine) peptidase inhibitor, clade A, member 3F
Tgtp2	1.56	2.47E-09	1.61E-06	T cell specific GTPase 2
Gm4951	2.83	3.24E-09	2.02E-06	predicted gene 4951
Samhd1	1.33	4.16E-09	2.48E-06	SAM domain and HD domain, 1
Thbs1	-3.44	7.29E-09	4.16E-06	thrombospondin 1
Irgm2	1.77	9.98E-09	5.47E-06	immunity-related GTPase family M member 2
Csf2rb	2.84	1.32E-08	6.98E-06	colony stimulating factor 2 receptor, beta, low-affinity
Gbp4	2.71	2.34E-08	1.19E-05	guanylate binding protein 4
Txnip	-0.91	3.38E-08	1.65E-05	thioredoxin interacting protein
Gm10925	-2.52	3.61E-08	1.70E-05	predicted gene 10925
Tap1	1.15	3.73E-08	1.70E-05	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)
Dtx3l	1.13	3.94E-08	1.74E-05	deltex 3-like, E3 ubiquitin ligase
Plac8	3.41	8.25E-08	3.54E-05	placenta-specific 8
Car1	-2.04	8.63E-08	3.58E-05	carbonic anhydrase 1
Itsn1	-1.04	1.62E-07	6.52E-05	intersectin 1 (SH3 domain protein 1A)
Tgtp1	2.77	7.49E-07	0.000293692	T cell specific GTPase 1
Csf2rb2	2.90	7.98E-07	0.000303915	colony stimulating factor 2 receptor, beta 2, low-affinity
Ciita	2.44	8.45E-07	0.000313103	class II transactivator
Parp9	1.78	1.56E-06	0.000561462	poly (ADP-ribose) polymerase family, member 9
Top2a	0.93	1.61E-06	0.00056695	topoisomerase (DNA) II alpha
F2r	0.95	1.67E-06	0.000573393	coagulation factor II (thrombin) receptor
Tapbp	0.95	1.87E-06	0.000625568	TAP binding protein
Prg2	-7.16	2.74E-06	0.000895038	proteoglycan 2, bone marrow
Zbp1	2.55	2.90E-06	0.000925633	Z-DNA binding protein 1
Ly6a	2.28	3.03E-06	0.000946052	lymphocyte antigen 6 complex, locus A
Ly6e	0.89	4.54E-06	0.001382378	lymphocyte antigen 6 complex, locus E
Gm15501	-1.53	5.35E-06	0.001595427	predicted pseudogene 15501
Oasl2	1.21	5.47E-06	0.001595427	2'-5' oligoadenylate synthetase-like 2
Psmb8	0.90	5.86E-06	0.001640052	proteasome (prosome, macropain) subunit, beta type 8

Wars	1.30	5.86E-06	0.001640052	tryptophanyl-tRNA synthetase
Mpo	-0.88	6.96E-06	0.001909998	myeloperoxidase
mt-Atp6	-2.42	7.33E-06	0.001970914	mitochondrially encoded ATP synthase 6
H2-T23	0.88	8.51E-06	0.002245169	histocompatibility 2, T region locus 23
Gm10801	-2.14	8.73E-06	0.002260538	predicted gene 10801
Dusp4	-1.29	1.05E-05	0.002659386	dual specificity phosphatase 4
Ifi47	1.49	1.29E-05	0.003204664	interferon gamma inducible protein 47
Cd274	1.43	1.90E-05	0.004646788	CD274 antigen
Hlf	-0.82	2.01E-05	0.004831274	hepatic leukemia factor
Trib2	-2.13	2.52E-05	0.00596244	tribbles pseudokinase 2
Parp14	1.06	2.77E-05	0.006442153	poly (ADP-ribose) polymerase family, member 14
Uba7	0.95	2.89E-05	0.006615639	ubiquitin-like modifier activating enzyme 7
Nr4a2	-1.14	3.14E-05	0.007070334	nuclear receptor subfamily 4, group A, member 2
Pygm	-1.59	3.34E-05	0.007391201	muscle glycogen phosphorylase
H2-T22	0.94	3.77E-05	0.008208401	histocompatibility 2, T region locus 22
Gm10800	-1.35	4.46E-05	0.009564986	predicted gene 10800
Elane	-1.33	4.61E-05	0.009732215	elastase, neutrophil expressed
Gbp8	2.52	4.69E-05	0.009752281	guanylate-binding protein 8
Gm11168	-1.56	5.13E-05	0.010501549	predicted gene 11168
Rnf213	0.63	5.25E-05	0.010584641	ring finger protein 213
Xist	-0.62	5.57E-05	0.011063578	inactive X specific transcripts
Aplp2	-0.76	5.75E-05	0.01126569	amyloid beta (A4) precursor-like protein 2
Gm15500	-1.30	6.04E-05	0.011662312	predicted pseudogene 15500
Nampt	1.15	6.61E-05	0.012591438	nicotinamide phosphoribosyltransferase
Ifi203	1.07	7.87E-05	0.014699307	interferon activated gene 203
Mndal	1.26	7.93E-05	0.014699307	myeloid nuclear differentiation antigen like
Serpinb1a	1.52	8.74E-05	0.015982047	serine (or cysteine) peptidase inhibitor, clade B, member 1a
Irf7	1.78	8.86E-05	0.015986803	interferon regulatory factor 7
Fgd5	-0.99	9.63E-05	0.016956587	FYVE, RhoGEF and PH domain containing 5
H2-DMa	1.31	9.70E-05	0.016956587	histocompatibility 2, class II, locus Dma
Cyp26b1	-2.64	9.77E-05	0.016956587	cytochrome P450, family 26, subfamily b, polypeptide 1
Rpl23a-ps3	-1.44	1.12E-04	0.019169913	ribosomal protein L23A, pseudogene 3
Psmb9	1.15	1.18E-04	0.019874224	proteasome (prosome, macropain) subunit, beta type 9
Ifitm3	0.99	1.19E-04	0.019874224	interferon induced transmembrane protein 3
Myom1	-0.95	1.34E-04	0.02211946	myomesin 1
Nlrc5	0.80	1.35E-04	0.02211946	NLR family, CARD domain containing 5 [
Oas3	1.47	1.38E-04	0.022228223	2'-5' oligoadenylate synthetase 3
H2-Q7	0.73	1.40E-04	0.022378784	histocompatibility 2, Q region locus 7
Clca3a1	1.30	1.89E-04	0.029848832	chloride channel accessory 3A1
Gm10275	-1.37	2.13E-04	0.033256064	predicted pseudogene 10275
Cxcl9	7.03	2.19E-04	0.033711588	chemokine (C-X-C motif) ligand 9
Rps23-ps1	-1.12	2.25E-04	0.03428905	ribosomal protein S23, pseudogene 1
Gm4070	1.27	2.34E-04	0.035333318	predicted gene 4070
Cdk6	0.56	2.52E-04	0.037550023	cyclin-dependent kinase 6
Rpl14-ps1	-1.20	2.55E-04	0.037550023	ribosomal protein L14, pseudogene 1
Rarg	-1.36	2.72E-04	0.039671525	retinoic acid receptor, gamma
Btg2	-0.71	3.11E-04	0.04490177	B cell translocation gene 2, anti-proliferative
Stat2	1.41	3.24E-04	0.04633537	signal transducer and activator of transcription 2
Slfn2	2.21	3.31E-04	0.046872047	schlafen 2
Tns2	-1.35	3.40E-04	0.047590322	tensin 2
Muc13	0.65	3.58E-04	0.049583497	mucin 13, epithelial transmembrane

Table S4: Parameters used in mosaic mouse predictions, related to STAR methods and Figure S3.

Genotype of minor population; condition:	Minor $N(0)$	Major $N(0)$	Minor λ_0 - λ_1	Major λ_0 - λ_1	Minor d	Major d	Minor $N(120)$	Major $N(120)$
WT; naive	10	20	.071 - .026	.071 - .026	.104	.104	1044	2088
WT; infection	10	20	.071 - .071	.071 - .071	.208	.208	2128	4256
***Dnmt3a*$^{-/-}$; naive**	10	20	.068 – .025	.071 - .026	.013	.104	2206	2088
***Dnmt3a*$^{-/-}$; infection**	10	20	.056 – .056	.071 - .071	.072	.208	5010	4256

1. The basis for predicting dynamics of two competing populations of HSCs is the differential equation of the form $\dot{N}(t) = -\lambda(t)N(t) + 2(1-d)\lambda(t)N(t)$, where $N(t)$ is the HSC count at time t, $\lambda(t)$ is the division rate of the HSC time-varying according to expression $\lambda(t) = \lambda_0 + (\lambda_1 - \lambda_0)t/120$ (time in days, λ's in day^{-1}), and d is the fraction of HSC progeny that are not HSC (i.e. that constitutes HSC "loss")[17]. We calculated λ_0 and λ_1 and $N(t)$ values from Ki67 (Figure 1C) and flow cytometric analysis of HSC counts in WT or *Dnmt3a*$^{-/-}$ mice with or without chronic *M. avium* infection (Figure 1G). We further calculated d for infected versus uninfected states (Figure 1H). The λ_0 - λ_1 and d values were then used to predict $N(t)$ from a starting set of 10 "minor population" HSCs and 20 "major population" (competitor) HSCs in post-transplant mice with or without infection introduced at 60 days post-transplant. Mathematical details and further illustrations are found in the Supplementary Information: Modeling and estimation of parameters for mosaic mice predictions.