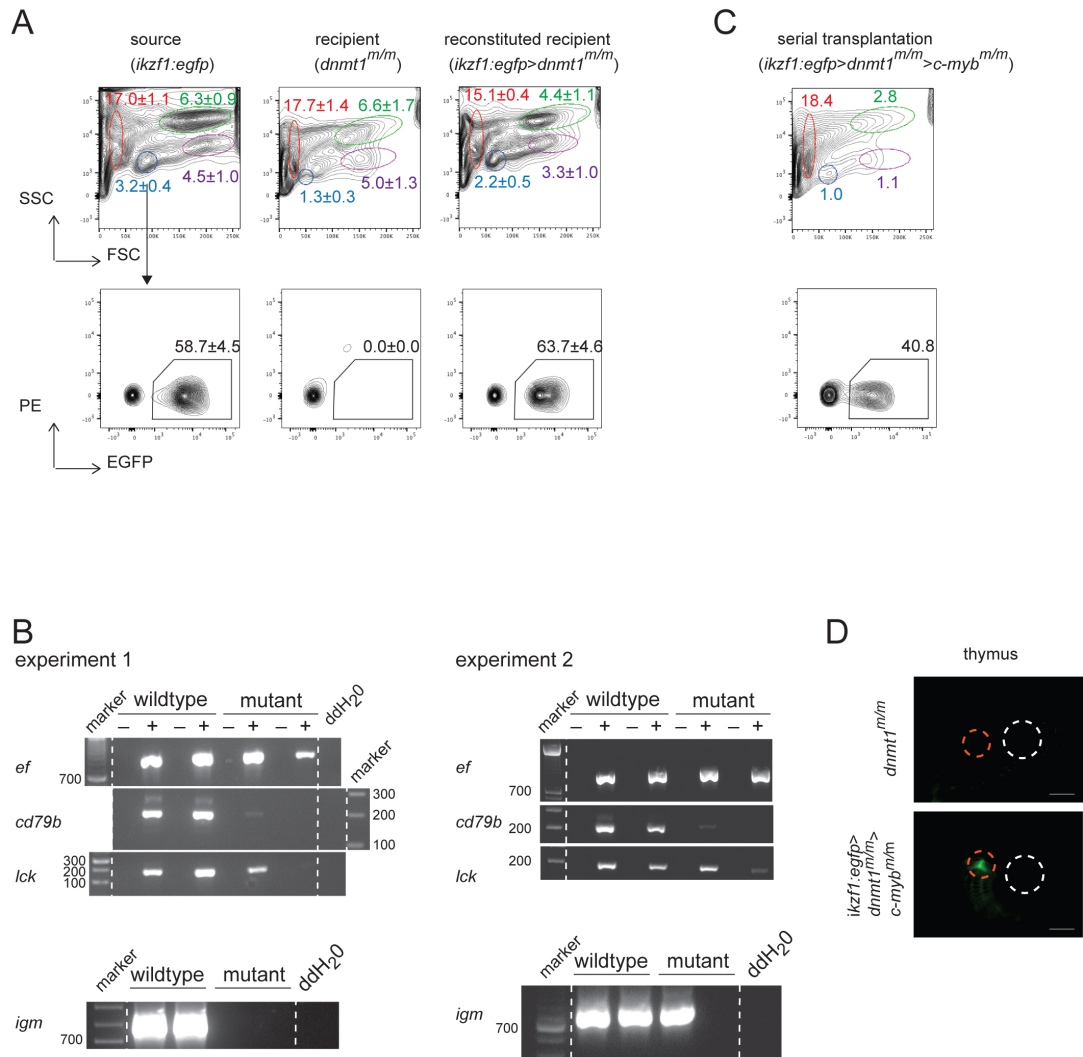


## Supplemental Information

### Epigenetic Protection of Vertebrate

### Lymphoid Progenitor Cells by *Dnmt1*

Norimasa Iwanami, Kohei Takeshita, Divine-Fondzenyuy Lawir, Isao Suetake, Shoji Tajima, Katarzyna Sikora, Inês Trancoso, Connor ÓMeara, Iliana Siamishi, Yousuke Takahama, Makoto Furutani-Seiki, Hisato Kondoh, Yasushige Yonezawa, Michael Schorpp, and Thomas Boehm



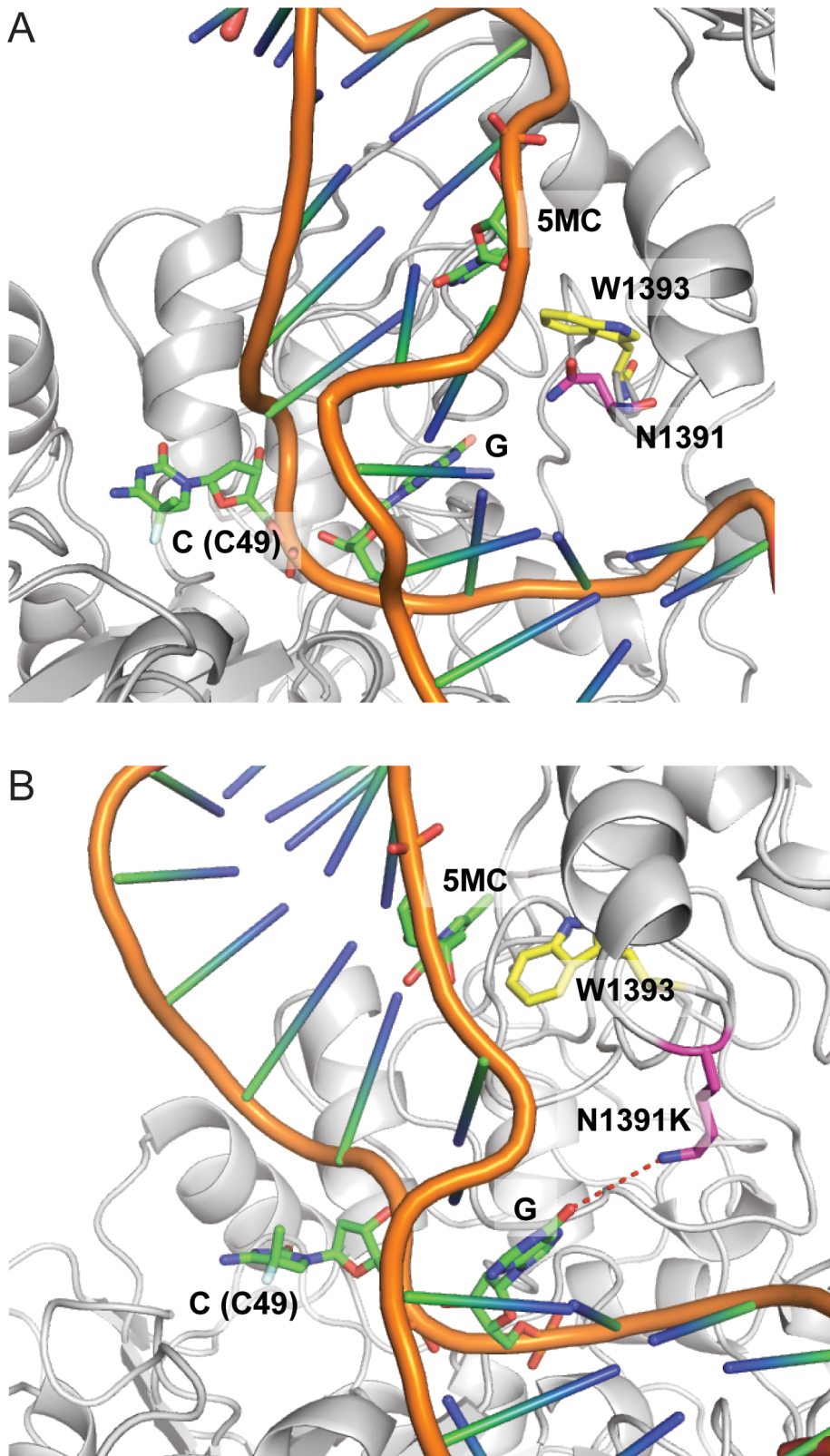
**Figure S1.** Long-term reconstitution of transplanted kidney cells in *dnmt1* mutant zebrafish. Related to Figures 1, 2.

(A) The top panels exemplify light scatter profiles of whole kidney marrow cells isolated from fish of the indicated genotypes and transplantation histories; the percentages of erythroid (red), lymphoid (light blue), myeloid (green), and precursor (purple) cell populations are indicated as mean±s.e.m.; the bottom panels exemplify EGFP expression patterns of cells in lymphoid populations, mean±s.e.m. is shown; n=7, 4, 6, respectively.

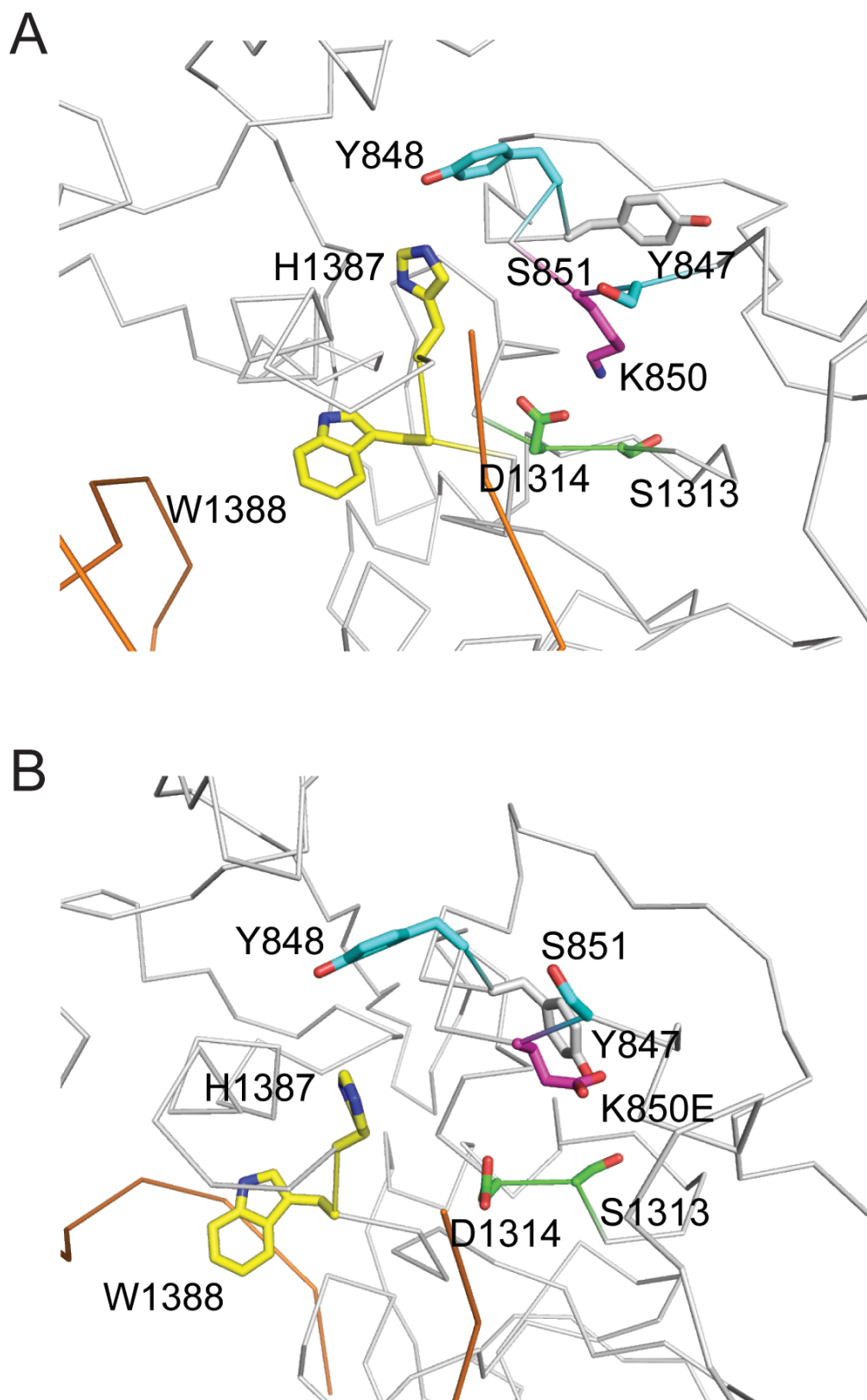
(B) Gel images of RT-PCR amplicons using kidney-derived RNA at 7 wpf. The results of 2 independent experiments are shown. Two *dnmt1<sup>+/+</sup>* (WT) fish and 2 *dnmt1<sup>m/m</sup>* (MT) fish were used for each experiment. The symbols – and + indicate cDNA templates generated without and with addition of reverse transcriptase to the reaction, respectively. *ef*, amplicon specific for elongation factor as a control for successful synthesis of cDNA; *cd79b*, amplicon specific for the B-cell marker CD79B factor; *lck*, amplicon specific for the T cell marker LCK; *igm*; amplicon specific for the B-cell marker immunoglobulin IGM; ddH<sub>2</sub>O denotes PCR reactions without template. Size markers are indicated in bp.

(C) Flow cytometric profiles of whole kidney marrow cells isolated from a *c-myb* mutant after receiving cells from a *dnmt1* mutant fish previously transplanted with wild-type whole kidney marrow cells; n=1.

(D) Fluorescent microscopic images of *dnmt1* mutant and the secondary *c-myb* mutant recipient shown in (B), 37 days after the secondary transplantation. Red circles indicate thymic region, and white circles indicate the location of the eye. Scale bars, 1mm.

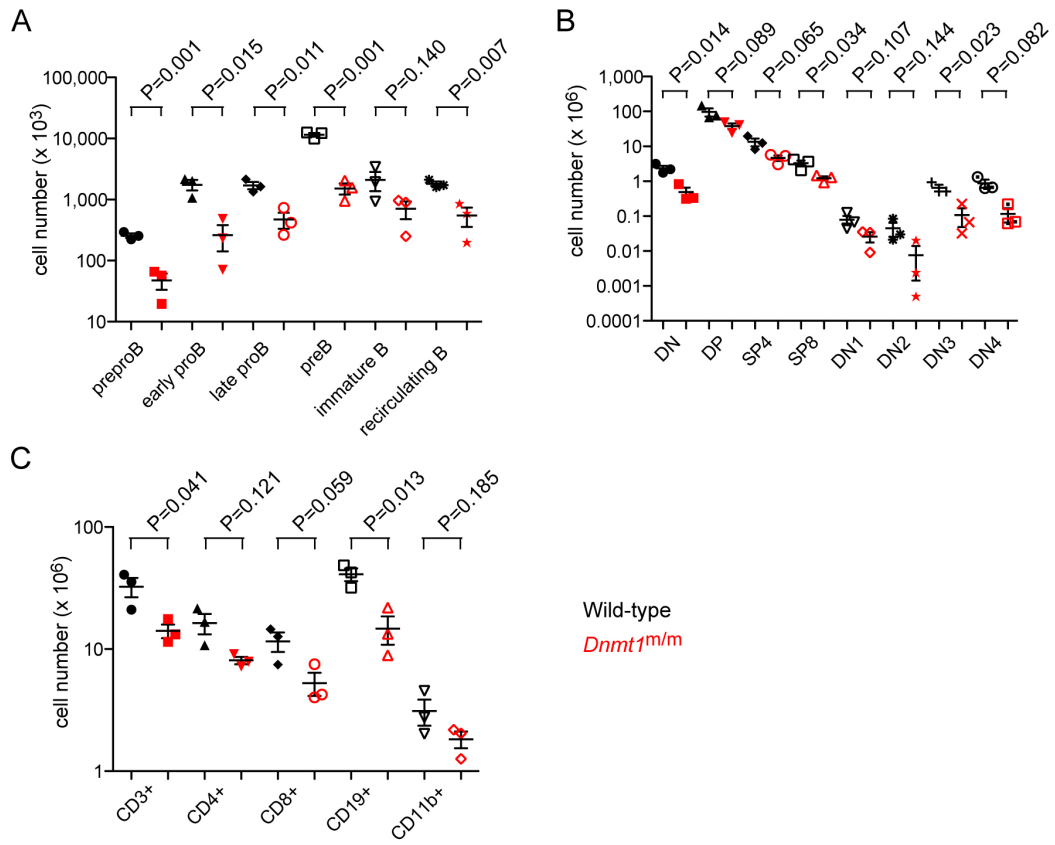


**Figure S2.** Structures of catalytic centers of zebrafish wild-type and mutant dnmt1 proteins as determined by molecular dynamics simulation. Related to Figures 1, 2, 4.  
 (A) Wild-type structure.  
 (B) N1391K mutant structure. Note the hydrogen bonds between the mutant lysine residue and the guanine base.



**Figure S3.** Structures of catalytic centers of medaka wild-type and mutant dnmt1 proteins as determined by molecular dynamics simulation. Related to Figures 1, 3, 4.  
(A) Wild-type structure.  
(B) K850E mutant structure. Note the distorted structure of the catalytic center.





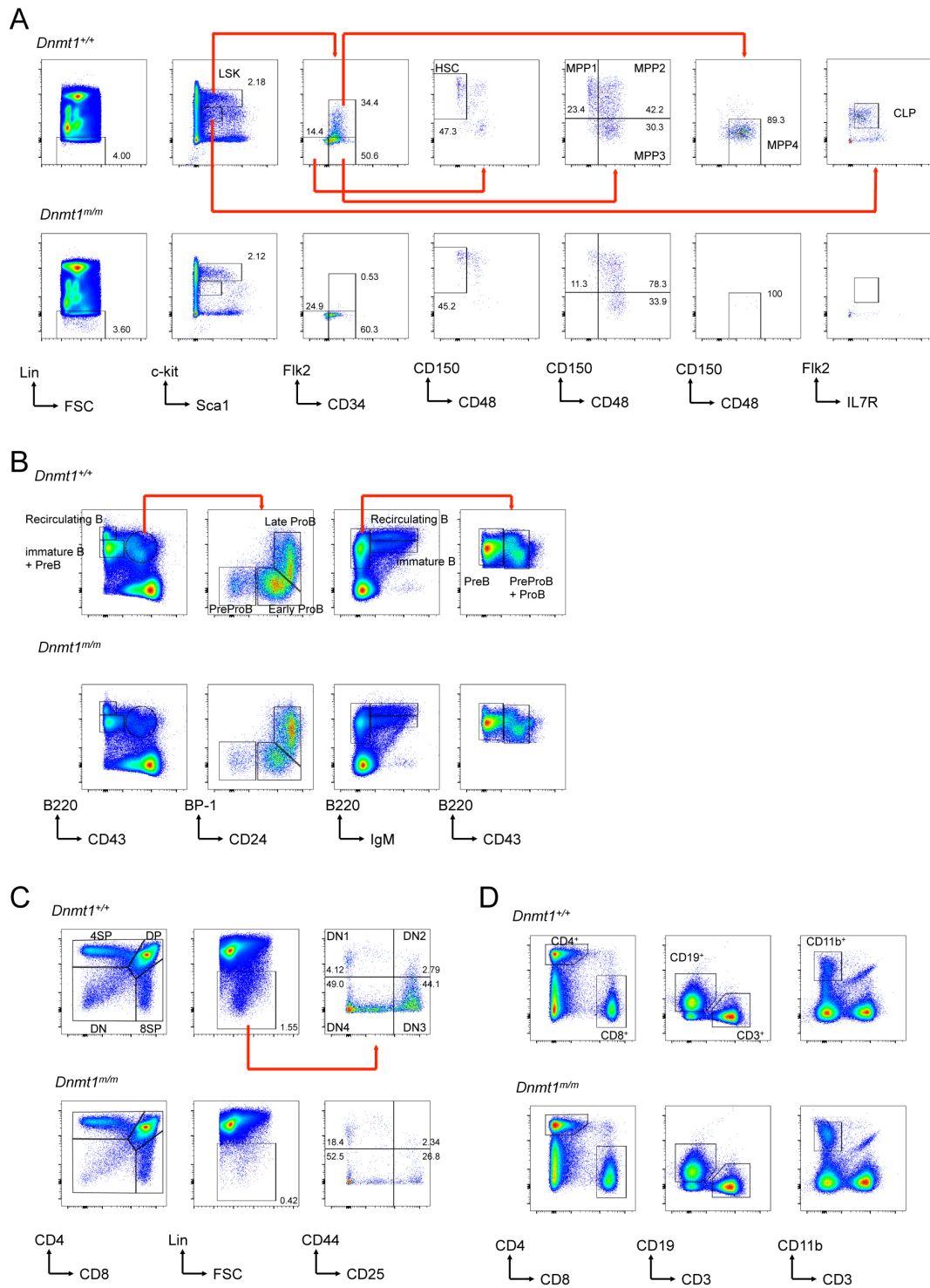
**Figure S4.** Impaired hematopoiesis in H1511D homozygous mice. Related to Figures 5, 6.

(A) B cell types in bone marrow.

(B) Cell types in thymus.

(C) Cell types in spleen.

Cell populations were measured by quantitative flow cytometry.



**Figure S5.** Gating strategies used for the characterization of the hematopoietic compartment of wildtype and H1511D mutant mice. Related to Figures 5, 6.

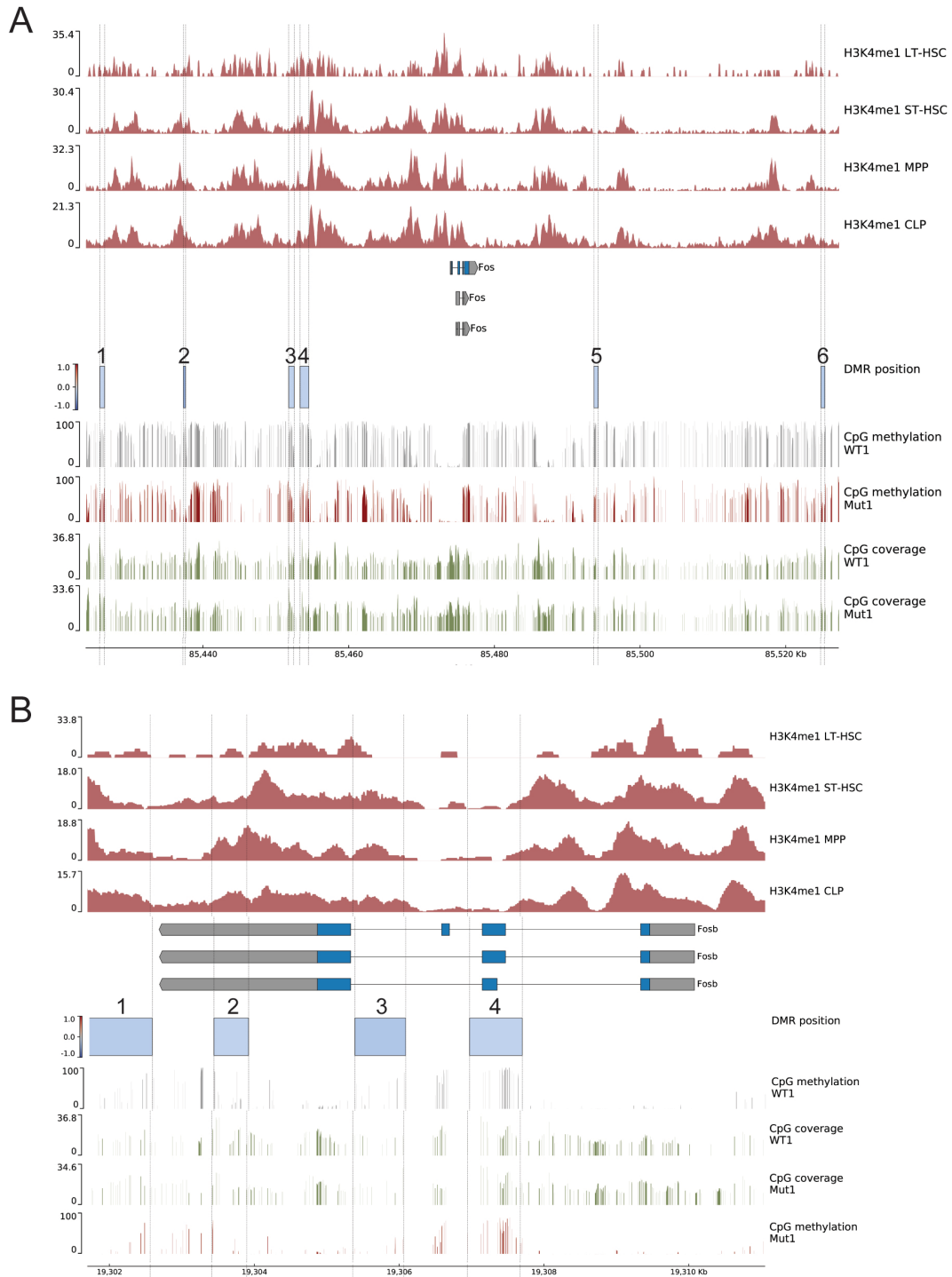
(A) Characterization of precursor populations in the bone marrow.

(B) Characterization of B cell precursor cells in the bone marrow.

(C) Characterization of T cell precursor cells in the thymus.

(D) Characterization of splenocytes.

Cell populations were measured by quantitative flow cytometry.

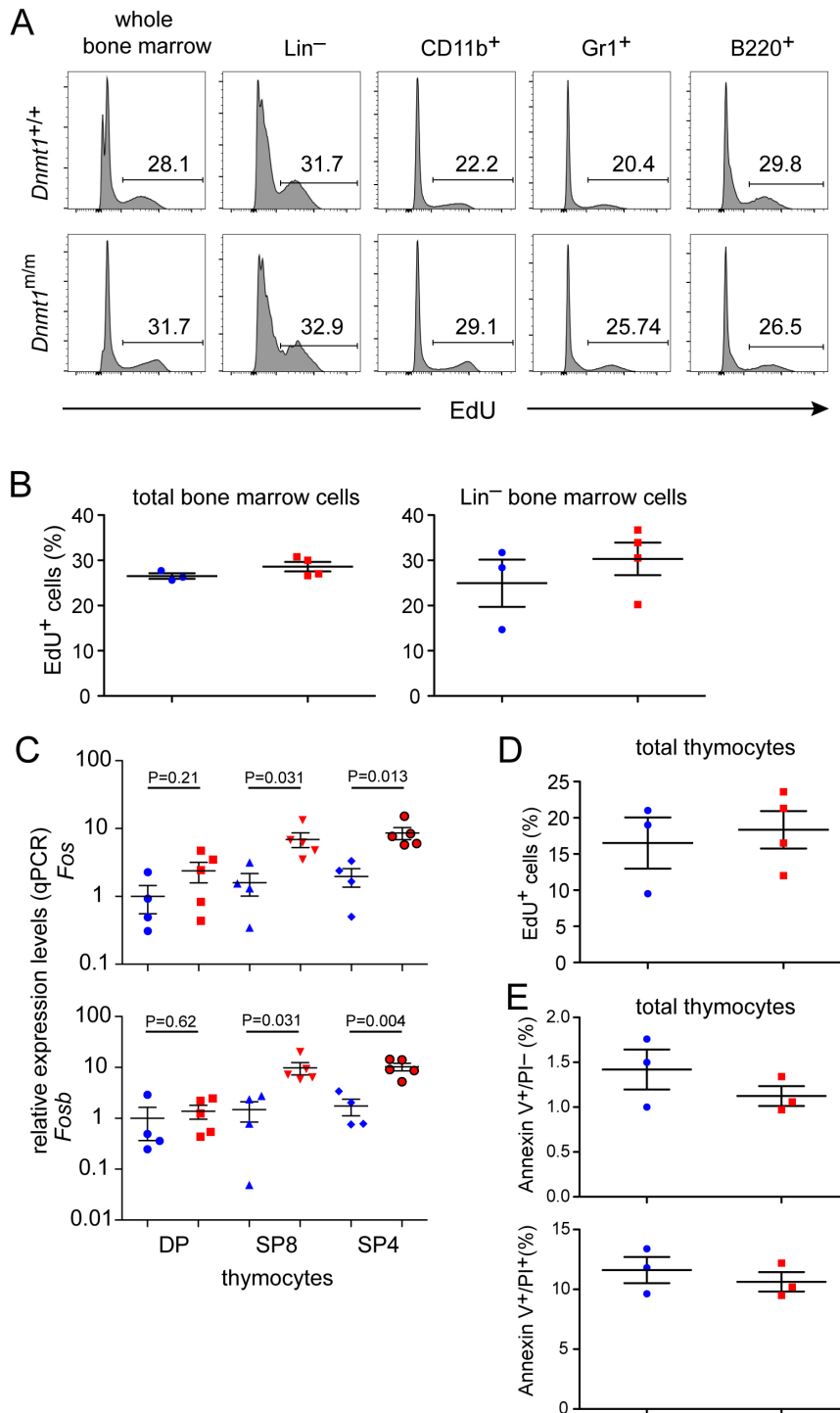


**Figure S6.** DMRs in the vicinity of *Fos* and *Fosb* genes. Related to Figure 7.

(A) Localization of 6 DMRs in the *Fos* locus; see Figure 7C.

(B) Localization of 4 DMRs in the *Fosb* locus; see Figure 7C.

Dashed vertical lines highlight the position of differentially methylated DMRs (hypomethylated) in wild-type and mutant mice.



**Figure S7.** Characterization of the hematopoietic compartment of H1511D mutant mice. Related to Figures 5, 6.

(A) Representative histograms showing EdU incorporation for the indicated bone marrow cell populations. Note the unperturbed proliferative potential of bone marrow cell populations in *Dnmt1*-mutants. (B) Unperturbed proliferative potential of bone marrow lin-negative precursor cell populations in *Dnmt1*-mutants.

(C) Elevated expression levels of *Fos* and *FosB* genes in thymocytes of *Dnmt1*-mutants as determined by qPCR.

(D) Unperturbed proliferative potential of thymocytes of and *Dnmt1*-mutant animals.

(E) Lack of increased apoptosis in thymocytes of *Dnmt1*-mutant animals.

B-E, wild-type (blue data points) and *Dnmt1*-mutant (red data points).

**Table S1.** Sequence polymorphisms in the critical genomic interval of the gkr medaka mutant. Related to Figure 3.

Row.names	POS	EFFECT	GENE	GENEID	HGVS_C	HGVS_P	Comments			
8:8390707_G/A	8390707	missense_variant	GB-ALPHA3	ENSORLG00000005267	c.208G>A	p.Ala70Thr	a) b)H,M			f)
8:8404841_G/A	8404841	missense_variant	zgc:92880	ENSORLG00000005283	c.373C>T	p.Leu125Phe			d)	f)
8:8420687_C/T	8420687	missense_variant	KIAA0172	ENSORLG00000005329	c.73G>A	p.Ala25Thr			d)	
8:8437467_G/A	8437467	missense_variant	DOCK6	ENSORLG00000005365	c.3554C>T	p.Pro1185Leu	a)		d)	
8:8499625_G/A	8499625	missense_variant	pde4a	ENSORLG00000005439	c.214C>T	p.Leu72Phe	a)	c)		f)
8:8538307_G/A	8538307	missense_variant	raver1	ENSORLG00000005458	c.1357C>T	p.Pro453Ser		c)		
8:8539847_T/A	8539847	missense_variant	raver1	ENSORLG00000005458	c.793A>T	p.Thr265Ser			d)	
8:8622951_A/C	8622951	missense_variant	dnmt1	ENSORLG00000005582	c.1743A>C	p.Glu581Asp			d)	
8:8624154_G/A	8624154	missense_variant	dnmt1	ENSORLG00000005582	c.2533A>G	p.Lys845Glu				
8:8633660_G/T	8633660	missense_variant	p2ry11	ENSORLG00000005606	c.406C>A	p.Leu136Ile	a) b)M	c)		
8:8634159_A/T	8634159	stop_gained	p2ry11	ENSORLG00000005606	c.14T>A	p.Leu5*	a) b)M		d)	e)
8:8638287_T/G	8638287	missense_variant	ppan	ENSORLG00000005634	c.575A>C	p.His192Pro			d)	
8:8641575_G/A	8641575	missense_variant	angptl6	ENSORLG00000005649	c.113G>A	p.Arg38His			d)	
8:8642057_G/T	8642057	missense_variant	angptl6	ENSORLG00000005649	c.515G>T	p.Gly172Val			d)	
8:8648875_G/A	8648875	missense_variant	ENSORLG00000005655	ENSORLG00000005655	c.297G>A	p.Met99Ile	a) b)H,M			f)
8:8649032_T/C	8649032	missense_variant	ENSORLG00000005655	ENSORLG00000005655	c.454T>C	p.Phe152Leu	a) b)H,M			f)
8:8649110_C/T	8649110	missense_variant	ENSORLG00000005655	ENSORLG00000005655	c.532C>T	p.Arg178Cys	a) b)H,M			f)
8:8666931_A/G	8666931	missense_variant	TRIP10__1_of_2_	ENSORLG00000005760	c.1616A>G	p.Asn539Ser			d)	f)
8:8776418_G/T	8776418	missense_variant	col5a3a	ENSORLG00000005829	c.3922C>A	p.Pro1308Thr	a)		d)	f)
8:8783104_G/T	8783104	missense_variant	col5a3a	ENSORLG00000005829	c.3259C>A	p.Leu1087Met	a)		d)	f)
8:8795210_C/A	8795210	missense_variant	col5a3a	ENSORLG00000005829	c.1278G>T	p.Met426Ile	a)		d)	f)
8:8804389_C/T	8804389	missense_variant	col5a3a	ENSORLG00000005829	c.481G>A	p.Val161Ile	a)		d)	f)
8:8831897_G/A	8831897	missense_variant	rdh8a	ENSORLG00000005853	c.920G>A	p.Arg307Gln		c)		f)
8:8905744_C/T	8905744	missense_variant	CAMSAP3__1_of_2_	ENSORLG00000006040	c.1081G>A	p.Ala361Thr	a)	c)		
8:8908131_G/C	8908131	missense_variant	CAMSAP3__1_of_2_	ENSORLG00000006040	c.437C>G	p.Ala146Gly	a)		d)	
8:8913101_T/A	8913101	missense_variant	CAMSAP3__1_of_2_	ENSORLG00000006040	c.299A>T	p.Lys100Ile	a)		d)	
8:9049797_G/T	9049797	missense_variant	MYH13__6_of_11_	ENSORLG00000006359	c.1903G>T	p.Ala635Ser			d)	f)
8:9104419_G/T	9104419	missense_variant	brd4	ENSORLG00000006490	c.2353G>T	p.Ala785Ser		c)		f)
8:9217392_C/G	9217392	missense_variant	ENSORLG00000006644	ENSORLG00000006644	c.98G>C	p.Arg33Thr			b)H,M,Z	
8:9217417_C/T	9217417	missense_variant	ENSORLG00000006644	ENSORLG00000006644	c.73G>A	p.Gly25Ser			b)H,M,Z	
8:9224420_C/T	9224420	missense_variant	trim35-13	ENSORLG00000006660	c.520C>T	p.His174Tyr	a)		b)H,M	
8:9224496_C/G	9224496	missense_variant	trim35-13	ENSORLG00000006660	c.596C>G	p.Ser199Cys	a)		b)H,M	
8:9229906_A/G	9229906	missense_variant	slc27a1a	ENSORLG00000006698	c.455T>C	p.Val152Ala			d)	f)
8:9233995_T/C	9233995	missense_variant	ptger1c	ENSORLG00000006707	c.25A>G	p.Lys9Glu			b)H,M	f)
8:9234003_C/T	9234003	missense_variant	ptger1c	ENSORLG00000006707	c.17G>A	p.Ser6Asn			b)H,M	f)
8:9460925_G/C	9460925	missense_variant	dnm2a	ENSORLG00000006841	c.2427C>G	p.Phe809Leu				f)

a) Ensembl transcript does not contain start codon.

b) No ortholog in human (H), mouse (M), or zebrafish (Z).

c) Amino acid residues are not conserved among 4 species at around mutation. Alignment is difficult.

d) Mutation residue is not conserved among 4 species.

e) One of two variants (exon1) contains the mutation nucleotide. Neither exon1 prediction might be wrong.

f) At least one more paralogue in medaka.



**Table S2.** Primers and oligonucleotides used in this study. Related to Figures 1 and 3.

Primer name	Primer sequence	Note
<b>medaka mapping primers</b>		
1345-F	TGCAATGACCGAGGATCTGTAAAG	
1345-R	CTTCTGGGCTAAAATCTCAGGCAA	
Scaf31_1470_F	CACAATCGTCCGCCCTCTTC	
Scaf31_1470_R	CCTCGCATCTGAAGACATTG	
<b>medaka dnmt1 genotyping</b>		
ol_dnmt1_L1	TTCTGTATGCGCTGGCTTGC	
ol_dnmt1_R1	CAGTTTGACGTCTGCAGTGTC	ol_dnmt1_R1 also for sequencing
<b>medaka bisulfite primers</b>		
ol_cpn1_BSF1	TGTTYGGAATGTAAGAGATTAGATAGAGATTG	Annealing temperature: 56°C
ol_cpn1_BSR1	TCRCTAACTCCAACACATAAAAAATAAC	
P5_ol_cpn1_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTTTTTAGTTTTYGTGTGTTTATTTGAGTTGAG	Annealing temperature: 55°C
P7_ol_cpn1_R1	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTTCRCTAACTCCAACACATAAAAAATAAC	
ol_irx3a_BSF1	GATTTTTTAAATGGGATAAGAAGAGATGAGTTGTGGG	Annealing temperature: 58°C
ol_irx3a_BSR1	TTAAACRACAAAATAAECTACATTCAAAACAAC	
P5_ol_irx3a_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTGGGGTYGTGGTTGTTTTGTTAAAG	Annealing temperature: 55°C
P7_ol_irx3a_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAACRACTATAAAAAATTTACAAAAACCAAAAC	
ol_krt4_BSF1	ATGTGGGTGTATGAGATAATATGGTTTATAG	Annealing temperature: 55°C
ol_krt4_BSR1	CAAATTTTTTAAAAAATCCTCCCTCTTCTAAAC	
P5_ol_krt4_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTAAAGYGGTTGTTGAGGGATTGATTTG	Annealing temperature: 57°C
P7_ol_krt4_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTTCTATACCATAACCAATCAACTCTCCAC	
ol_krt8_BSF1	GGYGGTATGGGGATTTGGGAGATG	Annealing temperature: 63°C
ol_krt8_BSR1	AACCCTCCCTCRCTACTACTTTAAAACTCTCC	
P5_ol_krt8_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTAAGTTTTYGAATTTTTATGGAAGAGTTAAG	Annealing temperature: 55°C
P7_ol_krt8_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAACRAACCAATTACTCTAACCAACCCAAAAATCC	
ol_ntla_BSF1	AAYGTTTTGAGTTGAATTTGATTTAGTGTGAAAAAG	Annealing temperature: 58°C
ol_ntla_BSR1	AACCAACRCCTAAAAATCTCTCTAAAC	
P5_ol_ntla_F1	ACACTCTTCCCTACACGACGCTCTCCGATCTAAYGTTTTGAGTTGAATTTGATTAGTGTGAAAAAG	Annealing temperature: 57°C
P7_ol_ntla_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAAAAACRAAATCTATAAAACAAATTTAAAAAATTAACC	
ol_rarga_BSF1	AGTTAGGTAGGAATTAGATTGTATTTAGTTG	Annealing temperature: 55°C
ol_rarga_BSR1	TACTACTCTACAATCAATAACAAAAACCCCTTAAC	
P5_ol_rarga_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTTATTAATTTAGATAGTTGTTGTTGTTGTTGATTAG	Annealing temperature: 57°C
P7_ol_rarga_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAATATATCRACAAAAAACAACAAAAACTCAC	
ol_pou5f1_BSF1	ATTTTAGGTYGTAGGGTAGAAGATTTTATGGGGTATGG	Annealing temperature: 56°C
ol_pou5f1_BSR1	CTTCAACCRRAACTTACAACCAATTTTAAATC	
P5_ol_pou5f1_F2	ACACTCTTCCCTACACGACGCTCTCCGATCTGATTGTTTTYGGGTTTGGGTTATTGGGGGGTTG	Annealing temperature: 59°C
P7_ol_pou5f1_R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTAACCTACTTACAACAAACCACTTATAAC	
ol_rgc101640_BSF1	AGGAATTTGTGATAAATTGAAAGTAGTGGAGTG	Annealing temperature: 55°C
ol_rgc101640_BSR1	ACCTCTCCACAATCAATAACATACATATAC	
P5_ol_rgc101640F2	ACACTCTTCCCTACACGACGCTCTCCGATCTTTAYGGTAGTTAAGGAAGGAAGAAGATAG	Annealing temperature: 57°C
P7_ol_rgc101640R2	GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTTATCCCACTATCTTACTTACATAATAAACTCTCTC	

**Table S4.** Differentially expressed genes in mouse hematopoietic progenitor populations.  
Related to Figure 7.

**HSCs**

	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
Fos__chr12	224.086878	1.57818887	0.27528112	5.73300799	9.87E-09	6.04E-05
Slc25a31__chr3	13.6636173	2.16486176	0.37323785	5.80022027	6.62E-09	6.04E-05
Hspa1b__chr17	15.7211853	2.00249014	0.37231931	5.3784214	7.51E-08	0.00030668
n-R5-8s1__chr18	33.9826963	1.7617384	0.34726974	5.07311232	3.91E-07	0.00119796
3830403N18Rik__chrX	10.6482505	1.85223151	0.37108556	4.99138666	5.99E-07	0.00146799
Jun__chr4	376.401288	1.30609389	0.2673006	4.88623635	1.03E-06	0.00209743
Fosb__chr7	115.444473	1.53445223	0.31932398	4.80531477	1.55E-06	0.00236475
Xlr3a__chrX	11.7733374	1.80175627	0.37298699	4.8306143	1.36E-06	0.00236475

**MPP1**

	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
Fos__chr12	181.652816	2.13959281	0.28994635	7.3792713	1.59E-13	3.24E-09
Gm10801__chr2	12712.2057	1.60943243	0.23855998	6.74644774	1.52E-11	1.54E-07
Gm26870__chr9	3632.56799	1.50105305	0.24793635	6.05418714	1.41E-09	9.58E-06
Gm17535__chr9	416.695834	1.62142188	0.29697577	5.45977835	4.77E-08	0.00024283
Fosb__chr7	122.053456	1.61859893	0.30050091	5.38633628	7.19E-08	0.00029303
Slc17a8__chr10	21.0513276	1.63427191	0.37906329	4.31134318	1.62E-05	0.04723095

**MPP4**

	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
Fos__chr12	218.987584	1.8925346	0.26778623	7.06733351	1.58E-12	2.31E-08
Slc25a31__chr3	11.7190425	2.42489572	0.41258584	5.87731203	4.17E-09	3.05E-05
3830403N18Rik__chrX	12.7715231	2.37805736	0.41040171	5.79446266	6.85E-09	3.34E-05
Hspa1a__chr17	17.5242296	2.27939166	0.41160531	5.53780923	3.06E-08	0.00011203
Tifab__chr13	55.6836923	1.67305569	0.31646211	5.28674882	1.25E-07	0.00036434
Cryab__chr9	10.250812	1.98444916	0.41161383	4.82114301	1.43E-06	0.00348067
Fosb__chr7	82.3970523	1.88705981	0.39593392	4.76609784	1.88E-06	0.00392588
Jun__chr4	283.010291	1.44605207	0.30679563	4.71340496	2.44E-06	0.00445535
Egr1__chr18	76.7558932	1.71870469	0.38567116	4.45639929	8.33E-06	0.01354956
Ube2c__chr2	219.965014	1.24885469	0.28639192	4.36064915	1.30E-05	0.01897306
Clec10a__chr11	7.4482292	1.75459929	0.40887703	4.29126401	1.78E-05	0.02363025
Stag3__chr5	27.1886181	1.43762347	0.35052816	4.10130666	4.11E-05	0.04995012
Xlr4a__chrX	7.67843443	1.68392317	0.41238281	4.08339809	4.44E-05	0.04995012

**Table S5.** Antibodies used in this study. Related to Figures 5, 6, S4, S5, and S7.

<b>Antigen/Staining Reagent</b>	<b>Clone</b>	<b>Conjugate</b>	<b>Source</b>	<b>Identifier</b>
<b>Thymocyte stage analysis</b>				
CD4	GK1.5	APC Cy7	Biolegend	100414
CD8a	53-6.7	PE	eBioscience	12-0081-85
CD44	IM7	APC	eBioscience	17-0441-81
CD25	PC61	PE Cy7	BD Biosciences	552880
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
TCR $\gamma\delta$	eBioGL3	FITC	eBioscience	11-5711-82
NK1.1	PK136	FITC	Biolegend	108706
CD11c	HL3	FITC	BD Biosciences	557400
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
<b>Annexin V detection in thymocytes</b>				
CD4	GK1.5	PE Dazzle	Biolegend	100455
CD8a	53-6.7	BV421	Biolegend	100738
CD44	IM7	APC	eBioscience	17-0441-81
CD25	PC61	PE Cy7	BD Biosciences	552880
TCR $\gamma\delta$	GL3	PerCP Cy5.5	Biolegend	118118
B220 (CD45R)	RA3-6B2	PE	Biolegend	103208
NK1.1	PK136	PE	eBioscience	12-5941-83
CD11c	N418	PE	eBioscience	12-0114-82
CD11b (Mac1)	M1/70	PE	BD Biosciences	553311
<b>Thymocyte stages TCR-<math>\beta</math></b>				
CD4	GK1.5	APC Cy7	Biolegend	100414
CD8a	53-6.7	BV421	Biolegend	100738
CD44	IM7	APC	eBioscience	17-0441-81
CD25	PC61	PE Cy7	BD Biosciences	552880
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
NK1.1	PK136	FITC	Biolegend	108706
CD11c	HL3	FITC	BD Biosciences	557400
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
TCR $\gamma\delta$	GL3	PerCP Cy5.5	Biolegend	118118
TCR- $\beta$	H57-597	PE	eBioscience	12-5961-83
<b>Thymocyte stages sort</b>				
CD4	GK1.5	APC Cy7	Biolegend	100414
CD8a	53-6.7	BV421	Biolegend	100738
CD44	IM7	PE	BD Biosciences	553134
CD25	PC61	Alexa Fluor 647	Biolegend	102020
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
TCR $\gamma\delta$	eBioGL3	FITC	eBioscience	11-5711-82
NK1.1	PK136	FITC	Biolegend	108706
CD11c	HL3	FITC	BD Biosciences	557400
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
<b>B cell stage analysis</b>				
IgM	II/41	FITC	BD Biosciences	553437
BP-1 (Ly51)	6C3	PE	Thermo Fischer	12-5891-83
CD43	S7	APC	BD Biosciences	560663
B220 (CD45R)	RA3-6B2	PE Cy7	eBioscience	25-0452-82
CD24	M1/69	eF450	eBioscience	48-0242-82
<b>Annexin V detection in B cells in the bone marrow</b>				
IgM	II/41	PE	eBioscience	12-5790-81
CD43	S7	APC	BD Biosciences	560663
B220 (CD45R)	RA3-6B2	PE Cy7	eBioscience	25-0452-82

CD24	M1/69	eF450	eBioscience	48-0242-82
CD3	145-2C11	PerCP Cy5.5	Biolegend	100328
<b>Peripheral lymphocyte analysis</b>				
CD4	GK1.5	APC Cy7	Biolegend	100414
CD8a	53-6.7	BV421	Biolegend	100738
TCR $\gamma\delta$	GL3	PerCP Cy5.5	Biolegend	118118
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
CD11c	HL3	FITC	BD Biosciences	557400
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
CD3e	145-2C11	APC	Thermo Fischer	17-0031-82
$\alpha$ -GalCer Loaded CD1 tetramer		PE	ProImmune	E001-2X
<b>EdU incorporation in Thymus</b>				
CD4	GK1.5	APC Cy7	Biolegend	100414
CD8a	53-6.7	BV421	Biolegend	100738
CD44	IM7	APC	eBioscience	17-0441-81
CD25	PC61	BV605	Biolegend	102035
<b>Sorting of CFSE labelled thymocytes</b>				
CD8a	53-6.7	PE Cy7	eBioscience	25-0081-82
CD4	GK1.5	PE Dazzle	Biolegend	100455
TCR $\beta$	H57-597	BV421	Biolegend	109229
CD62L	MEL-14	Alexa Fluor 700	Biolegend	104426
CD44	IM7	APC	eBioscience	17-0441-81
CD19	eBio1D3	PE	eBioscience	12-0193-83
<b>HSC-MPP phenotyping</b>				
CD3e	145-2C11	FITC	Biolegend	100306
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
Gr1 (Ly-6G/Ly-6C)	RB6-8C5	FITC	Biolegend	108406
TER-119	TER-119	FITC	eBioscience	11-5921-81
Sca1 (Ly-6A/E)	D7	APC	eBioscience	17-5981-81
Ckit (CD117)	2B8	BV421	Biolegend	105827
Flk2 (CD135)	A2F10.1	PE	BD Biosciences	561068
CD34	RAM34	Alexa Fluor 700	eBioscience	56-0341-82
CD150 (SLAMF7)	TC15-12F12.2	BV605	Biolegend	115927
CD48	HM48-1	APC Cy7	Biolegend	103431
<b>CLP phenotyping</b>				
CD3e	145-2C11	FITC	Biolegend	100306
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
Gr1 (Ly-6G/Ly-6C)	RB6-8C5	FITC	Biolegend	108406
TER-119	TER-119	FITC	eBioscience	11-5921-81
Sca1 (Ly-6A/E)	D7	APC	eBioscience	17-5981-81
Ckit (CD117)	2B8	BV421	Biolegend	105827
IL-7Ra	A7R34	PE Dazzle	Biolegend	135031
Flk2 (CD135)	A2F10	PerCP efluor710	eBioscience	46-1351-80
<b>Peripheral blood reconstitution</b>				
CD45.1	A20	PE	eBioscience	12-0453-83
CD45.2	104	BV421	Biolegend	109831
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
Gr1 (Ly-6G/Ly-6C)	RB6-8C5	FITC	Biolegend	108406
CD3e	145-2C11	APC	Thermo Fischer	17-0031-82
B220 (CD45R)	RA3-6B2	PE Cy7	eBioscience	25-0452-82
<b>HSC engraftment analysis</b>				
CD3e	145-2C11	FITC	Biolegend	100306
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206

CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
Gr1 (Ly-6G/Ly-6C)	RB6-8C5	FITC	Biolegend	108406
Ter119	TER-119	FITC	eBioscience	11-5921-81
Sca1 (Ly-6A/E)	D7	PE Cy7	Biolegend	108113
CD48	HM48-1	PerCP efluor710	eBioscience	46-0481-82
CD150 (SLAMF)	TC15-12F12.2	PE Dazzle	Biolegend	115935
Flk2 (CD135)	A2F10.1	PE	BD Biosciences	561068
CD45.2	104	Alexa Fluor 700	Biolegend	109822
CD45.1	A20	APC Cy7	Biolegend	110716

#### **Thymocyte engraftment analysis**

CD4	GK1.5	PE Cy7	Biolegend	100422
CD8	53-6.7	BV421	Biolegend	100738
CD44	IM7	PE	BD Biosciences	553134
CD25	PC61	BV605	Biolegend	102035
TCR $\gamma\delta$	GL3	PerCP Cy5.5	Biolegend	118118
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
NK1.1	PK136	FITC	Biolegend	108706
CD11c	HL3	FITC	BD Biosciences	557400
CD11b (Mac1)	M1/70	FITC	BD Biosciences	557396
CD45.2	104	Alexa Fluor 700	Biolegend	109822
CD45.1	A20	APC Cy7	Biolegend	110716

#### **Analysis of B cell engraftment in the bone marrow**

IgM	II/41	FITC	BD Biosciences	553437
CD43	S7	PE	BD Biosciences	553271
B220 (CD45R)	RA3-6B2	PE Cy7	eBioscience	25-0452-82
CD24	M1/69	eF450	eBioscience	48-0242-82
CD45.2	104	Alexa Fluor 700	Biolegend	109822
CD45.1	A20	APC Cy7	Biolegend	110716

#### **Analysis of engraftment of peripheral leukocytes**

CD4	GK1.5	PE Cy7	Biolegend	100422
CD8	53-6.7	BV421	Biolegend	100738
B220 (CD45R)	RA3-6B2	FITC	Biolegend	103206
TCR $\gamma\delta$	GL3	PerCP Cy5.5	Biolegend	118118
CD11b (Mac1)	M1/70	PE	BD Biosciences	553311
Gr1 (Ly-6G/Ly-6C)	RB6-8C5	PE	Biolegend	108408
CD45.2	104	Alexa Fluor 700	Biolegend	109822
CD45.1	A20	APC Cy7	Biolegend	110716



## Transparent Methods

**Fish lines.** The zebrafish (*D. rerio*) wild-type strain TLEK (Tüpfel long fin/Ekkwill), the *ikaros:eGFP* transgenic zebrafish (Bajoghli et al., 2009), the medaka (*O. latipes*) strain cab and the *rag1:eGFP* transgenic medaka (Li et al., 2007) are maintained in the animal facility of the Max Planck Institute of Immunobiology and Epigenetics. The medaka Kaga (strain ID: IB833) and the *gyokuro* (strain ID: MT369) strains (Furutani-Seiki et al., 2004; Iwanami et al., 2008; Iwanami et al., 2004) were supplied by the National Bioresource Project (NBRP) Medaka (<https://shigen.nig.ac.jp/medaka/>). All animal experiments involving medaka and zebrafish were approved by the institute's review committee and conducted under licenses from the local government (Regierungspräsidium Freiburg; license 35-9185.81/G-15/115). The developmental stages of medaka were designated as described (Iwamatsu, 2004).

**ENU mutagenesis of zebrafish and recovery of *dnmt1* mutant.** The IY071 (allele  $t^{25501}$ ) mutant (Iwanami et al., 2016) was established in collaboration with Tübingen 2000 Screen consortium (Boehm et al., 2003).

**ENU mutagenesis of medaka and recovery of *dnmt1* mutant.** For meiotic recombination mapping of the responsible mutation in the *gyokuro* line (*j48-12B*), the panels of PCR length polymorphism (PLP) markers in M-marker 2009 (Kimura and Naruse, 2010) were used for bulk segment analysis using DNA isolated from 21 *gkr* embryos and 38 wild-type siblings derived from an *gkr/cab* x kaga mapping cross. 103 *gkr* embryos were analyzed using markers described in MLBase ([http://mbase.nig.ac.jp/mbase/ml\\_base.html](http://mbase.nig.ac.jp/mbase/ml_base.html)) and custom-made PLP primers on chromosome 8. Primers used are listed in Table S2. In order to identify the mutation, genomic DNA was extracted from 87 *gkr* and 82 siblings derived from a *gkr/cab* x kaga mapping cross at stage 35 after *in situ* hybridization using a *rag1*-specific probe. After whole genome sequencing (Iwanami et al., 2016), sequencing reads were mapped to the medaka reference genome (MEDAKA1; Ensembl release 72) and any polymorphisms in the coding regions of the genes in the critical chromosomal interval between *gkr* mutants and wild-type siblings were recorded. Of the identified variants, those found in mutant DNA but identical to the reference sequence were not considered further, because the reference HdrR strain is closely related to the cab strain. To safeguard against the possibility of misphenotyped embryos included in the pools used for preparation of DNA, more than 80% of reads from the mutant pool and less than 50% of reads from wild-type pool needed to differ from the reference sequence, with an additional requirement that the coverage was at least 10 reads from each pool. No mutation was found in splice donor and acceptor sites; however, one nonsense and 36 missense mutations were identified (Table S1). The derived protein sequences of human, mouse, and zebrafish orthologs of candidate genes were aligned with medaka sequences to examine the degree of conservation of the mutated residues among these species (Table S1). Using this criterion, only two missense mutations (K850E in *dnmt1* and F809L in *dnm2a*) were deemed to be interesting candidates. Since there is a paralog of *dnm2a*, *dnm2*, located on chromosome 1 with possible associated functional redundancy, we focused on the *dnmt1* K850E missense mutation as the one responsible for the *gkr* phenotype.

**RNA *in situ* hybridization.** For hybridization using whole embryos or tissue slides, the relevant zebrafish probes were taken from (Schorpp et al., 2006; Soza-Ried et al., 2010). For the medaka *mpx* probe, nucleotides 645-1453 in Genbank accession number XM\_004074804 sequence were cloned into pGEM-T easy vector (Promega). After linearization of the plasmid with *SacI*, *in vitro* transcription was carried out with T7 polymerase using DIG RNA labeling kit (Roche). The other medaka probes were taken from (Bajoghli et al., 2009; Li et al., 2007).

**Histological analysis.** Histological analysis of fish specimens followed (Schorpp et al., 2006).

**RNA extraction and cDNA synthesis.** Total RNA was extracted using TRI Reagent (Sigma) following the manufacturer's instructions. After treatment with DNaseI (Promega), RNA extraction using TRI Reagent was repeated. Superscript III Reverse Transcriptase (Invitrogen) and random hexamers were used for cDNA synthesis from total RNA.

**RT-PCR.** For RT-PCR to examine expression of zebrafish genes, primer sequences were taken from (Hess and Boehm, 2016; Hess et al., 2013; Schorpp et al., 2006). Cycle numbers were as follows; *ef* (*elongation factor*): 30 cycles, *cd79b* and *lck*: 35 cycles, *igm* (igV<sub>H1</sub>-Cm): 30 cycles plus nested 30 cycles.

**Quantitative PCR.** qPCR was carried out (Boehm et al., 2003); (Iwanami et al., 2011); (Iwanami et al., 2016) using SYBR Premix Ex Taq (Takara) and 7500 fast real-time PCR system (Applied Biosystems). Zebrafish *actb1* and mouse *Hprt* were used as reference genes. The primer sets for zebrafish and mouse genes were purchased from BioRad.

**Expression of Dnmt1 protein variants.** The cloning and expression of Dnmt1 variants (aa 291-1620) followed (Vilkaitis et al., 2005; Berkyurek et al., 2014).

***in vitro* methylation assay.** The enzymatic activities of Dnmt1 variants were assayed *in vitro* following (Sugiyama et al., 2010; Garvilles et al., 2016), using unmethylated and hemi-methylated substrates (Suetake et al., 2006). Briefly, an annealed oligonucleotide (100 nM) was methylated with recombinant Dnmt1 (2 nM) in the presence of 2.2 μM [3H]-S-adenosyl-methionine (PerkinElmer Life Sciences) in 25 μl of buffer (5 mM EDTA, 20% glycerol, 0.2 mM DTT, 0.2 mM phenylmethylsulfonyl fluoride, and 20 mM Tris-HCl, 7.4) at 37°C for 1 h. The radioactivity incorporated into the DNA was measured using a liquid scintillation counter.

**Flow cytometry of zebrafish cells.** Flow cytometric analysis of light-scatter characteristics of WKM cells followed (Traver et al., 2003); staining with hydroxystilbamidine (Enzo Life Sciences; final concentration 1 μg/mL) was used to exclude dead cells.

**Zebrafish transplantation experiments.** Transplantation of whole kidney marrow cells of zebrafish followed (Hess et al., 2013).

**Imaging of zebrafish and medaka specimens.** Embryos and larvae were anesthetized and immobilized in 3% methylcellulose. Fluorescence microscopy was performed with Zeiss Imager.Z1.

**Mouse lines.** *Dnmt1* mutants (*Dnmt1* K979E; *Dnmt1* N1510K; *Dnmt1* H1511D) were generated by CRISPR-Cas9 technology and locus-specific sgRNAs (Table S2). Sequence-specific single-stranded repair oligonucleotides (Table S2) contained additional mismatches to avoid cleavage by the pre-assembled sgRNA/Cas9 RNP. The targeting sequences for guide RNAs were designed according to (Hwang et al., 2013) and cloned into the pDR274 vector (Addgene plasmid #42250). After digestion with *DraI* restriction enzyme (New England Biolabs), sgRNA was generated by *in vitro* transcription using MAXIscript T7 Transcription Kit (Thermo). For injection into fertilized eggs, two sgRNAs per target site were combined (final concentration 25 ng/μL), recombinant Cas9 protein from *Streptococcus pyogenes* (PNA Bio; final concentration 50ng/μL), and repair oligonucleotide (final concentration 5 μM) were mixed on ice in 10mM Tris, pH 7.5; 0.15mM EDTA; approximately 1–2 nL of the solution were injected per fertilized egg. When introduced into Balb/c, C57BL/6 and FVB genetic backgrounds, homozygosity of K979E or N1510K led to embryonic lethality. For the K979E mutant, the following results were obtained. Balb/c background: (a) Newborn mice: 19 mice of +/+ genotype; 17 mice of +/- genotype; 0 mice of m/m genotype. (b) E12.5 mouse embryos: 10 mice of +/+ genotype; 9 mice of +/- genotype; 0 mice of m/m genotype. C57BL/6 background: (a) Newborn mice: 18 mice of +/+ genotype; 25 mice of +/- genotype; 0 mice of m/m genotype. FVB background: No m/m mice were ever observed, but precise numbers of mice of +/+ genotype and +/- genotype were not determined. For the N1510K mutant, the following results were obtained. Balb/c background: (a) E8.5 mouse embryos: 2 mice of +/+ genotype; 13 mice of +/- genotype; 2 mice of m/m genotype. (b) E10.5 mouse embryos: 5 mice of +/+ genotype; 9 mice of +/- genotype; 0 mice of m/m genotype. C57BL/6 background: (a) E10.5 mouse embryos: 2 mice of +/+ genotype; 9 mice of +/- genotype; 2 mice of m/m genotype. (b) E13.5-E15.5 mouse embryos: 8 mice of +/+ genotype; 24 mice of +/- genotype; 0 mice of m/m genotype. Balb/c, C57BL/6, and FVB background: Newborn mice: No m/m mice were ever observed, but precise numbers of mice of +/+ genotype and +/- genotype were not determined. By contrast, the Balb/c background supported the survival of mice homozygous for the H1511D mutation. All mice were kept in the animal facility of the Max Planck Institute of Immunobiology and Epigenetics under specific pathogen-free conditions. All animal experiments were performed in

accordance with the relevant guidelines and regulations, approved by the review committee of the Max Planck Institute of Immunobiology and Epigenetics and the Regierungspräsidium Freiburg, Germany (license AZ 35-9185.81/G-15/35).

**Flow cytometry and cell sorting of mouse cells.** Phenotyping of lymphocyte populations was performed by flow cytometry after preparation of single cell suspensions from lymphoid organs and staining using antibodies listed in Table S5. Single cell suspension of thymus and spleen were prepared in FACS buffer (2% FBS, 1 mM ethylenediaminetetraacetic acid (EDTA), 1% penicillin- streptomycin, in PBS) by tissue homogenization with a syringe plunger against a 40  $\mu\text{m}$  cell strainer. For preparation of cell suspensions from bone marrow femur, tibia and pelvis were flushed with FACS buffer using a 10 ml syringe and a 26 gauge-needle and then passed through 40  $\mu\text{m}$  cell strainer to obtain single cell suspensions. For red blood cell lysis, the cell suspensions were treated with ACK lysis buffer (0.15M  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , 0.1mM EDTA in  $\text{H}_2\text{O}$ , pH 7.2-7.4), washed and resuspended in FACS buffer. Data were collected on a LSRFortessa and/or LSRII apparatus (BD Biosciences) and were analysed with FlowJo software version 10; cell sorting was done using a FACSAria instrument (BD Biosciences). In some experiments, the MojoSort mouse hematopoietic progenitor cell isolation kit (BioLegend) was used in order to analyze or sort lineage-negative bone marrow cells.

**EdU staining and cell cycle analysis of mouse cells.** For *in vivo* cell cycle analysis of thymocytes, mice received a single intra-peritoneal injection of EdU (5-ethynyl-2'-deoxyuridine) diluted in PBS at a dose of 50 mg/kg body weight. After 16h of EdU exposure, the mice were sacrificed and single cell suspensions of bone marrow and thymus were prepared. Single cell suspensions of EdU labeled cells were processed using the Click-iT EdU Flow Cytometry Assay (Thermo Fisher) according to the manufacturers protocol. Briefly, the cells were washed with 1% BSA in PBS and stained with surface antibodies (listed in Table S5) at 4 °C. The cells were washed again and fixed in Click-iT fixative for 15 min, washed once and then permeabilized in saponin-based permeabilization and wash reagent for 15 min. The Click-iT reaction cocktail was added to the cells for additional 30 min incubation. The cells were finally washed and analyzed by flow cytometry. All steps after the cell surface antibody staining were carried out at room temperature.

**Annexin V apoptosis detection in developing mouse lymphocytes.** The Annexin V (FITC) apoptosis detection Kit (eBioscience) was used for detection of apoptotic cells during development of T and B lymphocytes. Cell suspensions prepared from thymus and bone marrow were stained with surface antibodies (listed in Table S5) in FACS buffer (2% FBS, 1 mM EDTA, 1% penicillin- streptomycin, in PBS). After washing with 1 X diluted binding buffer (in  $\text{dH}_2\text{O}$ ) they were resuspended again in 1X binding buffer and incubated with FITC-conjugated Annexin V for 15 minutes at room temperature. The cells were finally washed with 1X binding buffer, resuspended in 1X binding buffer containing propidium iodide, and analyzed by flow cytometry.

**Competitive bone marrow reconstitution in mice.** Bone marrow (BM) cell suspensions from CD45.2 *Dnmt1<sup>m/m</sup>* and CD45.1/CD45.2 wild-type mice on the Balb/c background (7-12 weeks of age) were prepared and an aliquot was stained with surface antibodies for flow cytometric determination of HSC numbers. A bone marrow cell suspension of the desired donors, each containing 100 HSCs, were combined from the required sources before transplantation (wild-type donor [Ly5.1/Ly5.2];  $1.4 \times 10^6$  cells; H1511D/H1511D mutant donor [Ly5.2/Ly5.2];  $2.7 \times 10^6$  cells). The mixtures of CD45.1/CD45.2 wild-type and CD45.2 *Dnmt1<sup>m/m</sup>* bone marrow cells were resuspended in 100  $\mu\text{L}$  PBS and transplanted into lethally irradiated (a total of 9.5 Gy delivered in two doses of 5 Gy and 4.5 Gy, separated by a 3h interval) CD45.2 recipient mice (7-12 weeks of age) by tail vein injection. Peripheral blood was obtained from the recipient mice at 4 and 8 weeks after transplantation. Following red blood cell lysis, the contributions of CD45.2<sup>+</sup> donor derived cells to the T cell (CD3<sup>+</sup>), B cell (B220<sup>+</sup>) and myeloid (CD11b/Gr1<sup>+</sup>) lineages were assessed by flow cytometry. At 12 weeks after transplantation, the mice were sacrificed and BM, thymus and spleen were collected for flow cytometric analysis using the antibodies listed in Table S5.

**RNA\_seq analysis of lineage-negative mouse bone marrow cells.** For RNA-seq analyses, HSCs (n=3 biological replicates), MPP1 (n=2), and MPP4 (n=4) cells were isolated from bone marrow cells of *Dnmt1<sup>+/+</sup>* and *Dnmt1<sup>m/m</sup>* animals. Raw sequencing reads were processed with the snakePipes (Bhardwaj et al., 2019) version 0.5 scRNAseq-mapcount workflow to produce a count table per sample. Read trimming

was enabled, and annotation was filtered to remove entries with keyword "decay" or "pseudogene" in biotype description. Postprocessing and differential gene expression analysis were performed in R version 3.4.0. Counts were normalized by downscaling to 200,000 transcripts per cell with RaceID3 (Herman et al., 2018) and rounded to integer counts. Outlier samples were removed. Differentially expressed genes were called between *Dnmt1* mutant and wildtype replicates in each cell population separately, using R packages zingeR [<https://github.com/statOmics/zingeR>] version 0.1.0 and DESeq2 (Love et al., 2014) version 1.18.1. Sequencing batch was not included in the model. Gene lists were filtered for FDR<5% and absolute log<sub>2</sub> fold change of at least 1.

**Whole genome bisulfite sequencing.** Genomic DNA was extracted from three *dnmt1*<sup>+/+</sup> and three *dnmt1*<sup>m/m</sup> zebrafish at 18 dpf using the DNeasy blood and tissue kit (Qiagen). 1µg and 0.5µg of DNA was used for bisulfite reactions and library construction using the TruSeq DNA PCR-free library preparation kit (Illumina) and the EpiGnome Methyl-Seq kit (Epicentre), respectively. The fragments were sequenced in paired-end 100bp mode on 1 lane of Illumina HiSeq 2500 instrument.

**Zebrafish whole genome methylation analysis.** Raw sequencing reads were trimmed with cutadapt version 1.9.1 (Martin, 2011) as follows: First, 2 (TruSeq) or 6 (EpiGnome) 5'-most nucleotides were hard-trimmed and Illumina adapter sequences removed. Bisulfite-specific operations on reads and reference genome were performed with methylCtools version 0.9.4 (Hovestadt et al., 2014). Bisulfite-converted reads were mapped to bisulfite-converted GRCz10 zebrafish genome with bwa-mem version 0.7.12 separately for the two library types. Back-converted bam files were sorted with samtools version 1.3.1, PCR duplicates removed and read group information added with Picard tools v1.136. The two resulting bam files per sample were merged with samtools and methylation bias profiled with MethylDackel v0.1.7 [<https://github.com/dpryan79/MethylDackel>]. Methylated and unmethylated read counts per CpG position were extracted with methylCtools v0.9.4 with mapping quality threshold of 10, SNP detection, counting only 1 of two overlapping paired end reads, skipping 5 nucleotides from each read length and zero-padding of uncovered positions.

**Medaka amplicon methylation analysis.** The medaka CpG islands orthologous to the eight zebrafish CpG islands described by Potok et al. (Potok et al., 2013) were identified as described below. Primers for medaka amplicons were designed using Bisulfite Primer Seeker 12S (Zymo Research) and are listed in Supplementary Table 2. Genomic DNA was extracted from three *dnmt1*<sup>+/+</sup> and three *dnmt1*<sup>m/m</sup> medaka at 18 dpf using the DNeasy blood and tissue kit (Qiagen). 150ng of DNA was used for the bisulfite reaction using the EZ DNA methylation kit (Zymo Research) and the bisulfite-treated DNA was recovered in 15µl of elution buffer. 1µl of bisulfite-treated DNA solution was used for 30 cycles of PCR amplification with the indicated bisulfite primers in a reaction volume of 20µl; 1µl of the first PCR reaction was used for a nested PCR reaction with P5 and P7 adaptor-attached bisulfite primers. After the nested PCR reactions, amplicons were purified using PCR purification kit (Qiagen) and qualified and quantified by gel electrophoresis, the High Sensitivity DNA kit (Agilent) for the Agilent Bioanalyzer and the Qubit dsDNA HS assay kit (Invitrogen). The eight amplicons derived from the same original source were mixed at 1ng/µl each and 15µl of the mixture were amplified using P5 and P7 primers attached to barcodes using NEBNext DNA library prep kit (New England Biolabs), following the manufacturer's protocol. The amplicons were purified using AMPure XP (Beckman Coulter) and qualified and quantified using Bioanalyzer and Qubit dsDNA HS assay kit. Finally, all amplicons were mixed at 1ng/µl each and 5 µl of mixed amplicons were used for sequencing in paired-end 150bp mode on 1 lane of MiSeq Sequencing System (Illumina). Raw sequencing reads were trimmed with cutadapt version 1.9.1 (Martin, 2011), removing the first two 5'-most nucleotides and Illumina adapter sequences. Base quality trimming was performed using prinseq lite version 0.20.4. Bisulfite-specific mapping to bisulfite-converted MEDAKA1.72 genome was performed with Bismark v0.14.6 and Bowtie2 v2.2.8 in non-directional mode allowing for dovetail alignments. Samtools sorted and indexed bams were used for methylation bias calculation (MethylDackel v0.1.7) and read count extraction in target regions with methylCtools version 0.9.4 (Hovestadt et al., 2014), using MAPQ threshold 10, SNP detection, counting one of two overlapping mates, skipping 5 nucleotides from read ends, and zero-padding.

### **Post-processing and statistical analysis of methylation data in selected target regions of fish genomes.**

All postprocessing and statistical analysis were performed in R version 3.2.3. Read counts for CpG positions with illegal nucleotide frequency of at least 0.25 were set to NA. Detailed methylation analysis was performed on CpG sites from main chromosomes only. Methylated as well as unmethylated read counts from the two strands were summed per CpG position; coverage was calculated as sum of methylated and unmethylated reads per position on both strands. Methylation ratio (beta value) was calculated as ratio of methylated reads to coverage per site. Beta values for sites with coverage less than 10 reads were set to NA. A methylation matrix was obtained by merging methylation values for single samples with CpG positions in rows and samples in columns.

Zebrafish CpG methylation values obtained through whole genome bisulfite sequencing were intersected using bedtools version 2.23.0 with genomic intervals selected from Supplementary Table S4 of Potok et al. (Potok et al., 2013) lifted over to GRCz10 on the UCSC webpage. Single CpG methylation values were aggregated to mean values per interval allowing for maximum 1 NA in the total of six samples, and at least 20% of non-NA CpGs per interval. Mean beta values per interval were plotted for every replicate in each genotype group with ggplot2 version 2.1.0.

Medaka CpG methylation values were obtained through amplicon sequencing of the selected zebrafish genomic regions indicated above lifted over to MEDAKA1 using UCSC tools web service and corrected for primer placement. Single CpG methylation values were aggregated to mean values per interval allowing for a maximum of 1 NA in a total of 6 samples, and at least 20% of non-NA CpGs per interval. Mean beta values per interval were plotted for every replicate in each genotype group with ggplot2 version 2.1.0.

Between-group statistical testing of methylation ratio differences in zebrafish and medaka was performed on logit-transformed interval mean beta values (adjust offset 0.025, R package cars version 2.1-1). P values from two-sided two-sample t-test (with 3 replicates per genotype group) are reported.

**DNA methylation analysis of mouse Lin<sup>-</sup> cells.** Raw sequencing reads were processed with the snakePipes (Bhardwaj et al., 2019) version 1.2.0 WGBS workflow providing mm10 genome to produce bam files filtered for PCR duplicates. This was done separately for reads obtained with each library preparation kit. The resulting bam files were merged per sample. Merged bams from the previous step were processed with the snakePipes (Bhardwaj et al., 2019) version 1.2.0 WGBS workflow using the `-fromBam` argument and providing a sample sheet. Gencode version m9 gene models were used. This produced a.o. a list of differentially methylated regions between the mutant and the wildtype sample groups filtered for FDR<2% and absolute methylation difference between groups of at least 20%. Bigwig tracks with WGBS methylation and coverage were generated with snakePipes (Bhardwaj et al., 2019).

**Reanalysis of publicly available mouse hematopoietic progenitor cell ChIP-seq data.** ChIP reads for mouse HSC H3k27me3 (Sun et al., 2014) were downloaded from ENA (GSE47765), as were ChIP reads for mouse LT/ST-HSC and MPP H3K4me3, H3K4me1 and H3K27ac (Lara-Astiaso et al, 2014; GSE59636), as well as ATAC-seq reads for mouse LT/ST-HSC, MPP3/4 and proB.CLP (Yoshida et al., 2019; GSE100738). Fastq files were merged for multiple runs of the same sample.

Reads were mapped to mouse genome GRCm38 and processed with snakePipes (Bhardwaj et al., 2019) version 2.1.1 DNA-mapping and ATAC-seq workflows to produce bigwig files with normalized coverage. Genomic tracks plots for specific genes were obtained with pyGenomeTracks (Ramirez et al., 2018) version 3.2. To align CHIP and ATAC\_seq reads to DMRs, normalized coverage bigwig files were obtained for H3K27me3, H3K4me3, H3K4me1 and H3K27ac marks and for open chromatin, as described above, gencode version m9 gene model gtf for GRCm38, bed files with DMR positions, bigwig files with filtered CpG coverage for WT and Mut replicates, bigwig files with methylation value (0-100%) for wild-type and mutant replicates.

The plot for *Fos* was generated for genomic interval chr12:85423890-85527273 (gene locus + 50kb flanks); for *Fosb*, chr7:19301696-19311051 (1 kb flanks); for *ikzf1*, chr11:11684003-11773926 (1kb flanks); for *Ebfl*, chr11:44616317-45009091 (1kb flanks).

**Statistical analysis of animal phenotypes.** No randomization of animals was done in the present studies; phenotypes were recorded by a blinded observer before genotyping. In zebrafish and medaka, the sex of animals can only be determined in late adult stage; therefore, in the present studies, no consideration was given to sex. No animals were excluded from analyses. Samples size was estimated from the degrees of variability in previous analyses (Boehm et al., 2003); (Iwanami et al., 2016); (Schorpp et al., 2006) in order



to be able to detect biologically meaningful differences in examined parameters, usually 20% difference from control values. t-tests were performed for samples with equal variance; otherwise, F-tests were used.

**Molecular dynamics simulation.** Homology models were constructed using the structure of the complex of mouse Dnmt1 with a hemi-methylated CpG DNA substrate (PDB ID: 4da4) (Song et al., 2012) using SWISS-MODEL (Waterhouse et al., 2018). Missing residues in the protein were modeled using the Modellar program (Sali and Blundell, 1993), and missing hydrogen atoms were added under conditions of pH 7. For stabilization and equilibration of homology models, structures were set at the center of a rectangular box and were solvated in a rectangular box filled with TIP3P (Jorgensen et al., 1983) model waters; appropriate numbers of sodium and chloride ions were added to achieve charge neutral systems with an ionic concentration of 150 mM. The number of water molecules was defined to be about 17,500. The closest distance between the protein and the rectangular box was set to 17 Å. The force field of proteins and ions were calculated using amber03 force field (Duan et al., 2003); the water molecules were rigidified using the SETTLE algorithm (Miyamoto and Kollman, 1992); electrostatic interactions were calculated using the particle-mesh-Ewald method (Essmann et al., 1995) using a cutoff radius of van der Waals interactions of 8 Å. Adequate restrained potentials were allied to the heavy atoms at the original positions. The total potential energy was minimized using the steepest decent method. The system was then simulated using a constant number of atoms at a constant temperature (310K) with the Berendsen algorithm (Berendsen et al., 1984). Chemical bonds in the protein were treated as rigid using the linear constraint solver (LINCS) algorithm (Hess, 2008). The time step was set to 2 fs, and the simulation with a constant number of atoms, constant pressure (1 atm), and constant temperature (310K) was conducted using the Parrinello-Rahman barostat (Parrinello and Rahman, 1980) and Bussi–Donadio–Parrinello thermostat (Bussi et al., 2007). The restrained potential was then released, and a production run was initiated. Coordinates of the whole system were recorded every 2 ps for the analysis; however, the first 5-ns trajectories were used to equilibrate the system and were then discarded. Molecular dynamic simulations were performed using the Gromacs-4.5.5 program package (Pronk et al., 2013). For the modeling of the medaka protein, the dnmt1 variant X1 was used (accession number XM\_023957476.1; XP\_023813244.1) comprising a total of 1,496 amino acids; mouse residue K979 corresponds to medaka residue K850, and zebrafish N1391 corresponds to medaka N1386.

**Data availability.** The R code necessary to reproduce statistical analyses and results is reported in Supplemental Code available at [https://github.com/katsikora/Iwanami2019\\_SupplementaryCodeAndData\\_B](https://github.com/katsikora/Iwanami2019_SupplementaryCodeAndData_B). The methylation and RNA\_seq data have been deposited in the GEO database under accession number GSE98648.

## Supplemental references

Berendsen, H.J.C., Postma, J.P.M., Gunsteren van, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* *81*, 3684-3690.

Berkyurek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M., and Tajima, S. (2014) The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemi-methylated DNA. *J Biol Chem* *289*, 379-386.

Bhardwaj, V., Heyne, S., Sikora, K., Rabbani, L., Rauer, M., Kilpert, F., Richter, A.S., Ryan, D.P., and Manke, T. (2019). snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics* *35*, 4757-4759.

Boehm, T., Bleul, C.C., and Schorpp, M. (2003). Genetic dissection of thymus development in mouse and zebrafish. *Immunol Rev* *195*, 15-27.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J Chem Phys* *126*, 014101.

Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., *et al.* (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* *24*, 1999-2012.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen L. G. (1995) A smooth particle mesh Ewald method. *J Chem Phys* *101*, 8577-8593.

Garvilles, R.G., Hasegawa, T., Kimura, H., Sharif, J., Muto, M., Koseki, H., Takahashi, S., Suetake, I., and Tajima, S. (2015) Dual Functions of the RFTS Domain of Dnmt1 in Replication-Coupled DNA Methylation and in Protection of the Genome from Aberrant Methylation. *PLoS One* *10*, e0137509.

Herman, J.S., Sagar, and Grun, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat Methods* *15*, 379-386.

Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* *4*, 116-122.

Hovestadt, V., Jones, D.T.W., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., *et al.* (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* *510*, 537-541.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* *31*, 227-229.

Iwanami, N., Mateos, F., Hess, I., Riffel, N., Soza-Ried, C., Schorpp, M., and Boehm, T. (2011). Genetic evidence for an evolutionarily conserved role of IL-7 signaling in T cell development of zebrafish. *J. Immunol.* *186*, 7060-7066.

Jorgensen, W.L., Chandrasekhar, J., and Madura, J.D. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* *79*, 926-935.

Kimura, T., and Naruse, K. (2010). M-marker 2009, a marker set for mapping medaka mutants using PCR length polymorphisms with an automated microchip gel electrophoresis system. *Biotechniques* *49*, 582-583.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I. *et al.* (2014) Chromatin state dynamics

during blood formation. *Science* 345, 943-949.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10-12.

Miyamoto, S., and Kollman, P.A. (1992). SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry* 13, 952-962.

Parrinello, M., and Rahman, A. (1980). Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters* 45, 1196-1199.

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., *et al.* (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845-854.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., *et al.* (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun* 15, 189. doi:10.1038/s41467-017-02525-w

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.

Soza-Ried, C., Hess, I., Netuschil, N., Schorpp, M., and Boehm, T. (2010). Essential role of *c-myb* in definitive hematopoiesis is evolutionarily conserved. *Proc. Natl. Acad. Sci. USA* 107, 17304-17308.

Suetake, I., Hayata, D., and Tajima, S. (2006). The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. *J Biochem* 140, 763-776.

Sugiyama, Y., Hatano, N., Sueyoshi, N., Suetake, I., Tajima, S., Kinoshita, E., Kinoshita-Kikuta, E., Koike, T., and Kameshita, I. (2010). The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. *Biochem J* 427, 489-497.

Sun, D., Luo, M., Jeong, M., Rodriguez, B. *et al.* (2014) Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* 14, 673-688.

Traver, D., Paw, B.H., Poss, K.D., Penberthy, W.T., Lin, S., and Zon, L.I. (2003). Transplantation and in vivo imaging of multilineage engraftment in zebrafish bloodless mutants. *Nat Immunol* 4, 1238-1246.

Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005). Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem* 280, 64-72.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46, W296-W303.

Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A. *et al.* (2019) The cis-Regulatory Atlas of the Mouse Immune System. *Cell* 176, 897-912.