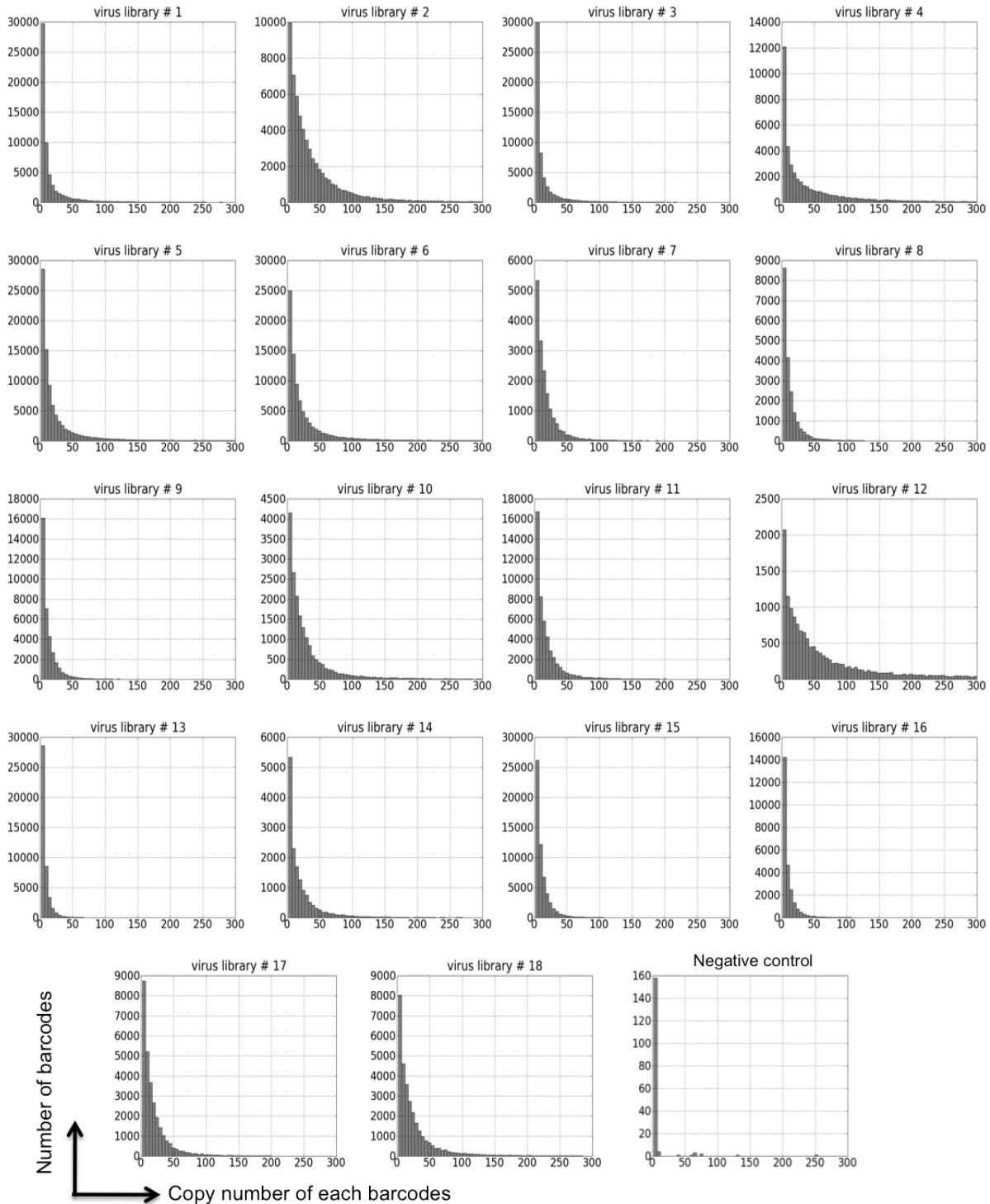
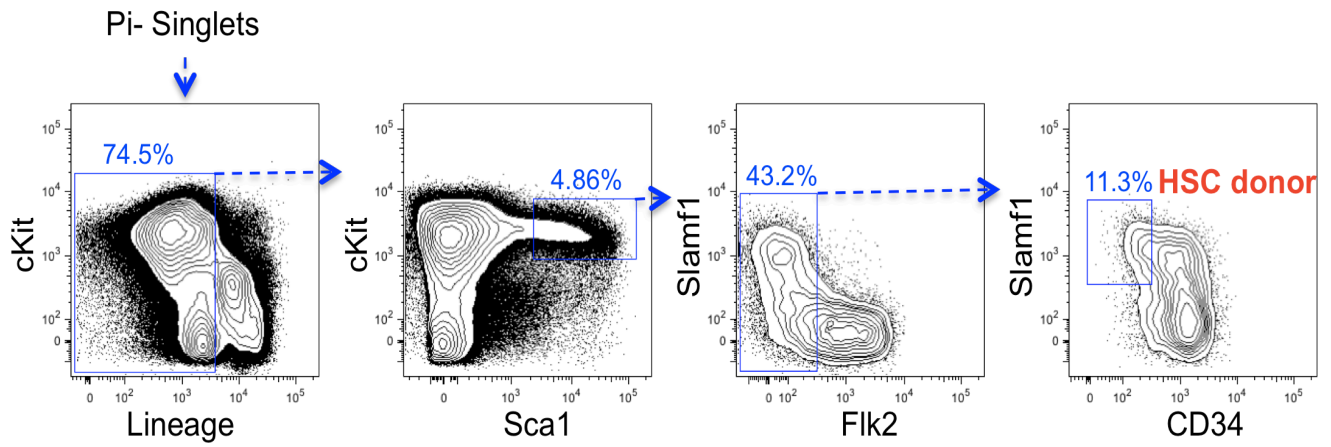


SUPPLEMENTARY FIGURES

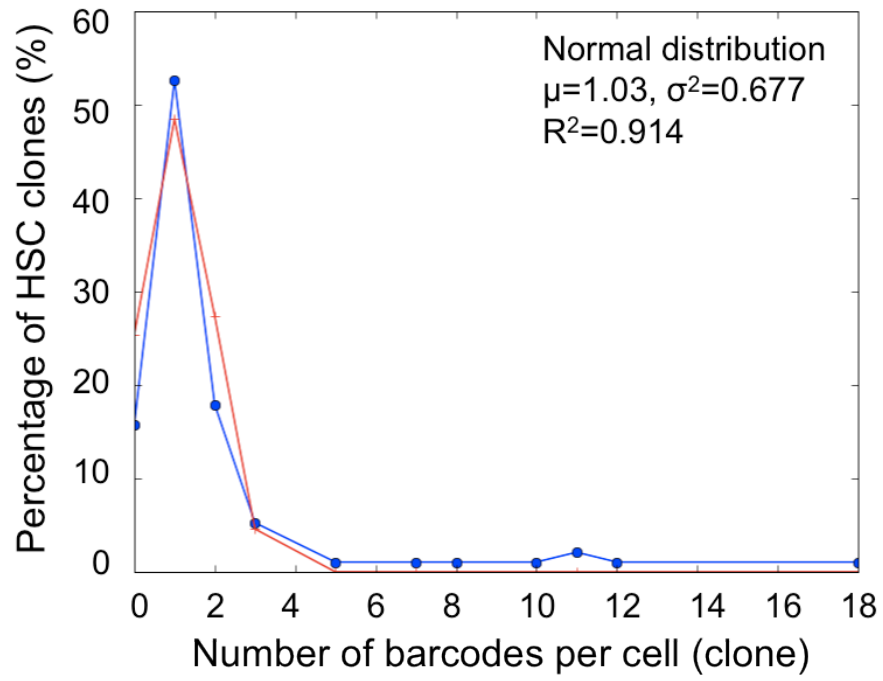
Supplementary Figure 1 Histogram displaying the distribution of DNA barcode copy numbers for each virus library. 100,000 BB88 cells were infected by lentivirus that carried the corresponding barcode libraries. Cells were cultured for one week before harvesting. Barcodes were recovered from the cells and analyzed using high throughput sequencing. The sequencing result did not comprehensively cover the entire barcode libraries as most barcodes were sequenced with low copy numbers.



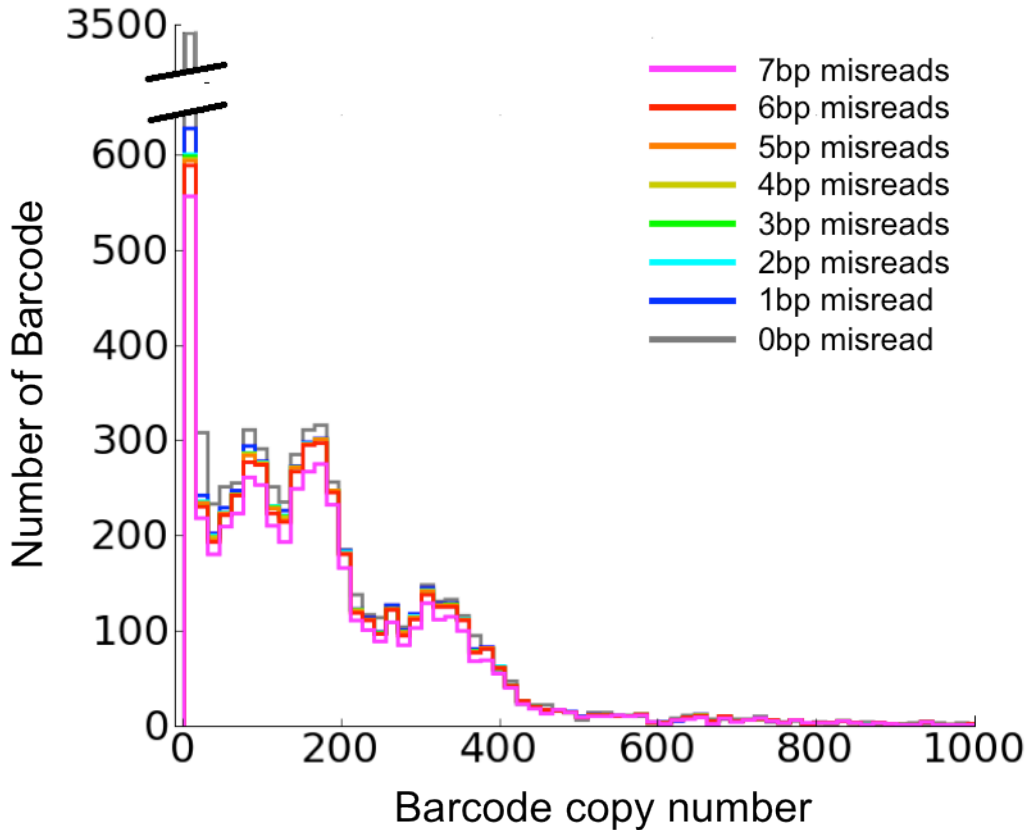
Supplementary Figure 2 Gating for donor hematopoietic stem cells (lineage (CD3, CD4, CD8, B220, Gr1, Mac1, Ter119)-/cKit+/Sca1+/Flk2-/CD34-/CD150+) from cKit enriched bone marrow cells. Details for the cell preparation are described in Methods. Antibodies used are shown in Supplementary Table 3.



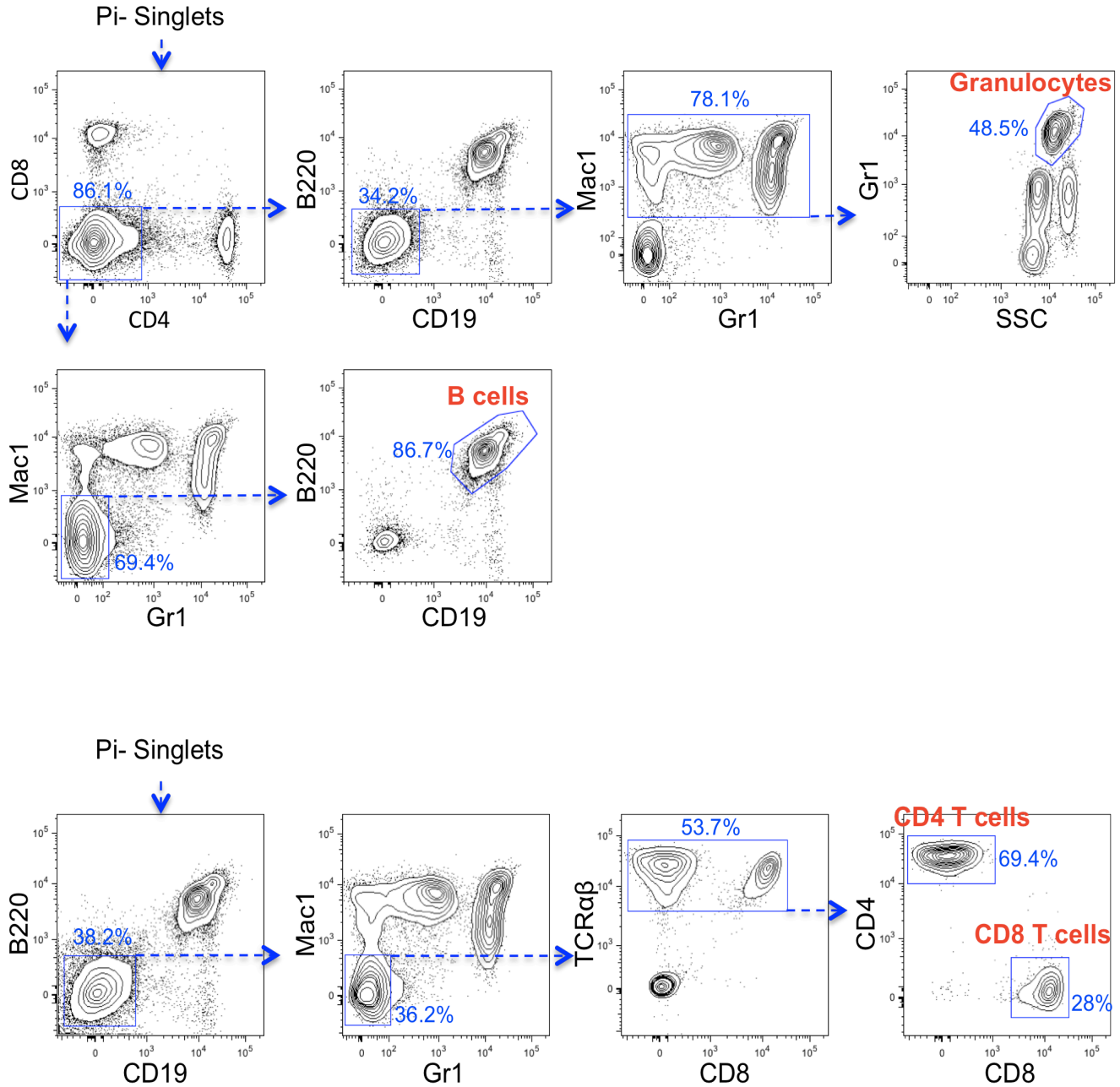
Supplementary Figure 3 The barcode copy numbers per HSC clones (Fig. 2b) are fitted by a normal distribution. Experimental data is shown in blue and the fitted curve is shown in red.



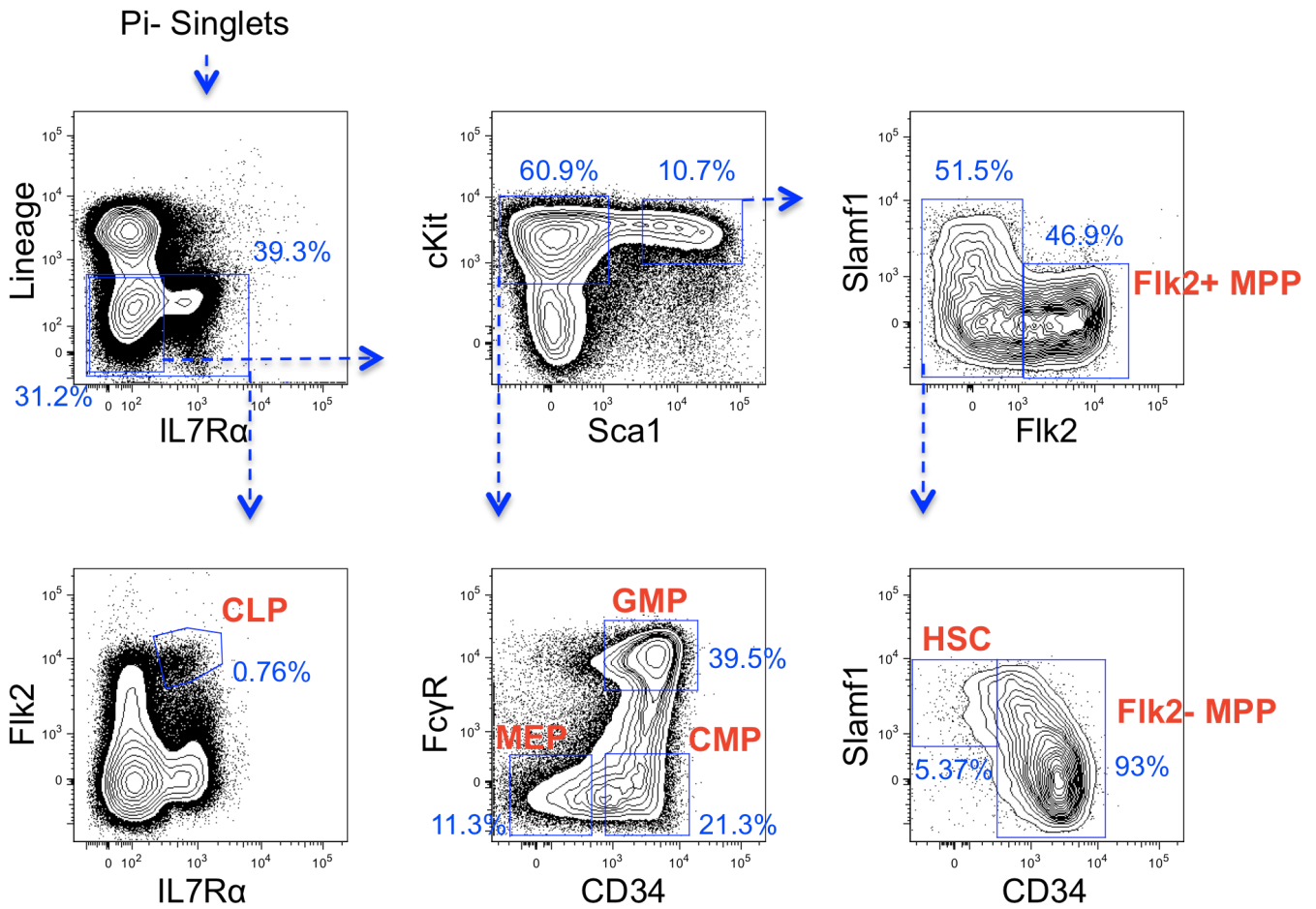
Supplementary Figure 4 Histogram displaying the copy number distribution of background noise sequences when different numbers of misreads are allowed. All sequences with a common library ID are combined allowing for different numbers of misreads. Shown are the combined sequences with total copy number less than 1000, which represent the background noise sequences (Fig. 3). The copy number distribution of the background noise sequences changes when the number of allowed misreads increases from zero to one or more. This effect is particularly prominent at low copy numbers close to 0. However, the overall copy number distribution of the background noise remains similar even when more misreads are allowed. This suggests that while some of the background noise sequences arise from sequencing error, most of the noise may be attributed to other sources.



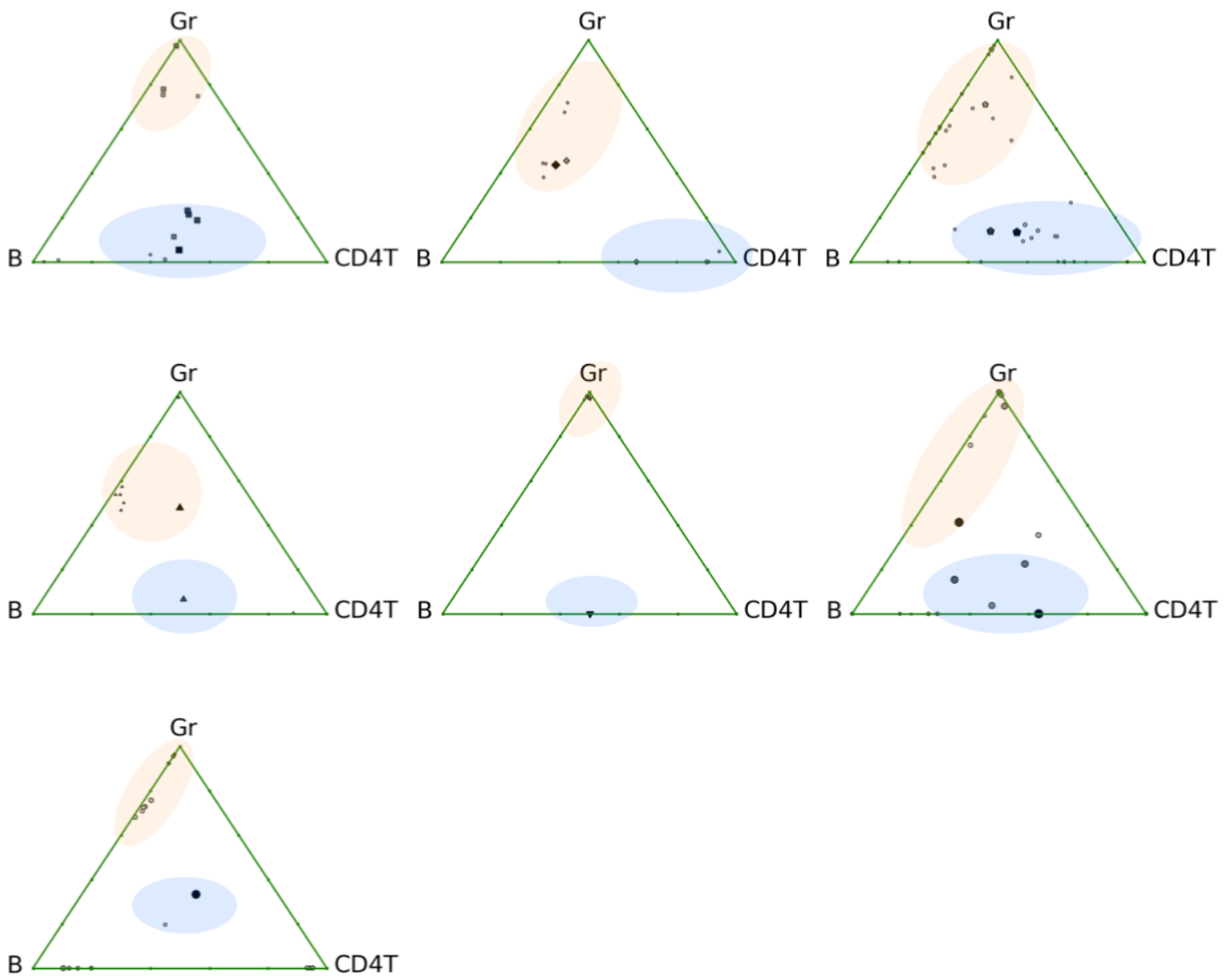
Supplementary Figure 5 Gating for harvesting of mature blood cells. The four harvested blood cell populations are labeled in red. Red blood cells were lysed before sorting (details described in Methods). A summary list of the cell surface markers for each harvested cell population is included in Methods. The antibodies used are shown in Supplementary Table 3.



Supplementary Figure 6 Gating for hematopoietic progenitor populations harvested from bone marrow. The harvested HSCs and six additional progenitor cells are labeled in red. Bone marrow cells were enriched by ckit and IL7R α (details described in Methods). A summary list of the cell surface markers for each harvested cell population is included in Methods. The antibodies used for sorting are shown in Supplementary Table 3.



Supplementary Figure 7 Lineage bias of HSC differentiation after irradiation. Triangle plots show the relative proportion of barcodes in granulocytes (Gr), B cells (B) and CD4+ T cells (CD4T) (Fig. 4). Each dot within the triangle represents a distinct barcode. Bigger and darker dots represent more abundant barcodes. The distance of a dot to the three vertices of the triangle is inversely correlated with the relative abundance of the barcode in the corresponding cell populations. For example, if a barcode is only found in one cell population, the dot is plotted at the corresponding vertex; if a barcode appears equally in all three populations, the dot is plotted in the middle of the triangle. Each triangle depicts a single mouse transplanted with 9000 donor HSCs after lethal irradiation. Blood cells were harvested 22 weeks after transplantation. Distinct barcode groups are highlighted by blue and orange ellipses.



SUPPLEMENTARY TABLES

Supplementary Table 1 Library ID sequences. Every 33bp barcode begins with a 6bp library ID as shown below. These library IDs help identify the barcode library of origin. They also allow several experiments using different barcode libraries to be combined in a single sequencing run.

Virus #	Library ID
1	GATCTG
2	TCAAGT
3	CTGATC
4	AAGCTA
5	GTAGCC
6	TACAAG
7	ATGACA
8	AGCGGT
9	ACTCAG
10	TAACGT
11	TGTTAC
12	TCCGTA
13	GAGTTC
14	GTCGAG
15	GCAACT
16	CAGTGC
17	CTTATG
18	CGACCT

Supplementary Table 2 Representation of single cells. Monte Carlo simulations were performed using experimentally derived data (Fig. 2a and Fig. 2b). We calculated the null hypothesis P values for different target cell population sizes such that more than 95% of the barcodes represent single cells (Fig. 2c). Shown are the maximum numbers of cells that can be used to ensure that with greater than 95% probability more than 95% of the barcodes represent single cells. Since the experimentally observed barcode library size is a lower bound of the actual size, the calculated cell population size is also a lower bound on the true size that can be used to ensure single cell representation.

Virus #	Total sequences	Number of barcodes	Maximum cell number	Probability of non-single-cell representation
virus # 1	589622	57581	810	4.84%
virus # 2	1905941	60083	1270	4.42%
virus # 3	533253	52377	750	4.70%
virus # 4	651316	40807	540	4.92%
virus # 5	1218693	83089	1470	4.76%
virus # 6	1398942	82694	1500	4.59%
virus # 7	265727	16953	350	4.68%
virus # 8	236955	20102	400	4.36%
virus # 9	438632	35794	830	4.88%
virus # 10	444384	18344	370	4.82%
virus # 11	800268	48077	1130	4.62%
virus # 12	888489	16161	320	4.90%
virus # 13	306101	44297	1000	4.57%
virus # 14	293547	15562	280	3.92%
virus # 15	612777	57002	1340	4.76%
virus # 16	242610	25301	530	4.08%
virus # 17	493350	28952	640	4.16%
virus # 18	682103	31032	660	4.86%

Supplementary Table 3 Monoclonal antibodies applied in this study.

(a) The following monoclonal antibodies were used for sorting donor HSCs:

Antigen	Conjugation	Vendor	Catalog #	Clone	Lot #
B220	Pacific Blue	home conjugated		6B2	
CD3	Pacific Blue	home conjugated		KT31.1	
CD34	FITC	ebioscience	11-0341-85	RAM34	E00265-195
CD4	Pacific Blue	home conjugated		GK1.5	
CD8	Pacific Blue	home conjugated		53-6.7	
Ckit	APC-A 750	ebioscience	27-1171-82	2B8	E033660
Flk2	PE	ebioscience	12-1351-83	A2F10	E026170
Gr1	Pacific Blue	home conjugated		8C5	
Mac1	Pacific Blue	home conjugated		M1/70	
Sca1	PE-Cy7	ebioscience	25-5981-82	D7	E030156
Slamf1	Alexa fluor 647	biolegend	115918	TC15-12F12.2	B111222
Ter119	Pacific Blue	home conjugated		Ter119	

(b) The following monoclonal antibodies were used to harvest blood cells:

Antigen	Conjugation	Vendor	Catalog #	Clone	Lot #
B220	Alexa Fluor 647	ebioscience	51-0452-82	RA3-6B2	E08573-290
CD19	PE CY5-5	ebioscience	35-0193-82	eBio1D3	E031334
CD4	Biotin	home conjugated		GK1.5	
CD8	PE-Cy5	ebioscience	15-0081-83	53-6.7	E026286
Gr1	e450	ebioscience	48-5931-82	RB6-8C5	E10103-1207
CD45.2	PE-Cy7	biolegend	109830	104	B128067
CD45.1	Alexa Fluor 680	home conjugated		A20.1.7	
Mac1	APC-e780	ebioscience	47-0112-82	M1/70	E10339-1357
Streptavidin	Qdot 605	invitrogen	Q10101MP	n.a.	729527
TCR $\alpha\beta$	PE	ebioscience	12-5961-83	H57-597	E01951-1628

(c) The following monoclonal antibodies were used to harvest progenitor cells in the bone marrow:

Antigen	Conjugation	Vendor	Catalog #	Clone	Lot #
B220	Pacific Orange	home conjugated		6B2	
CD3	Pacific Orange	home conjugated		KT31.1	
CD34	A647	ebioscience	51-0341-82	RAM34	E08567-165
CD4	Pacific Orange	home conjugated		GK1.5	
CD8	Pacific Orange	home conjugated		53-6.7	
Ckit	APC-A750	ebioscience	27-1171-82	2B8	E033660
Fc γ R	Biotin	ebioscience	13-0161-82	93	E02430-161

FIk2	PE-Cy5	ebioscience	15-1351-82	A2F10	E026568
Gr1	Pacific Orange	home conjugated		8C5	
IL7R α	e450	ebioscience	48-1271-82	A7R34	E10929-1095
CD45.2	A680	home conjugated		AL1-4A2	
CD45.1	PE Cy5-5	ebioscience	35-0453-82	A20	
Mac1	Pacific Orange	home conjugated		M1/70	
Sca1	PE-Cy7	ebioscience	25-5981-82	D7	E030156
Slanf1	PE	ebioscience	12-1501-80	9D1	E020036
Streptavidin	Qdot 605	invitrogen	Q10101MP	n.a.	729527
Ter119	Pacific Orange	home conjugated		Ter119	

Supplementary Table 4 Pearson correlation coefficients of barcode representations (copy numbers) between every two hematopoietic populations. M1-M7 refer to mouse #1-mouse #7. Cells were harvested at 22 weeks after lethal irradiation mediated transplantation. Abbreviations: hematopoietic stem cell (HSC), Flk2- multipotent progenitor (MPP-), Flk2+ multipotent progenitor (MPP+), granulocyte/monocytic progenitor (GMP), megakaryotic/erythroid progenitor (MEP), common lymphocyte progenitor (CLP), granulocyte (Gr), B cell (B), CD4 T cells (CD4T) and CD8 T cells (CD8T).

		B	Gr	CD4T	MPP+	CLP	CD8T	HSC	MEP	GMP
MPP-	M 1	-0.04358	0.204141	-0.15721	0.4429	0.500613	-0.16483	0.816008	0.415896	0.261053
MPP-	M 2	0.130588	0.574309	0.201966	0.497514	0.504934	0.161223	0.431082	0.725771	0.71611
MPP-	M 3	0.037741	0.13374	-0.02558	0.007092	0.083607	-0.03973	0.94736	0.160448	0.184216
MPP-	M 4	-0.30521	0.951969	-0.30611	0.4391	-0.29883	-0.31283	0.91094	-0.12119	0.199636
MPP-	M 5	0.839805	0.915203	0.430158	0.898093	0.876153	0.238985	0.140705	0.921014	0.898179
MPP-	M 6	0.35927	0.949004	0.292558	0.803683	0.756328	0.261334	-0.05913	0.881404	0.847847
MPP-	M 7	0.003464	0.167506	0.002584	0.704911	0.179613	-0.00687	0.056732	0.313992	0.204161
	average	0.146012	0.556553	0.062624	0.541899	0.371774	0.019612	0.463386	0.471048	0.473029
		MPP-	Gr	CD4T	MPP+	CLP	CD8T	HSC	MEP	GMP
B	M 1	-0.04358	0.190943	0.765933	0.157182	0.157075	0.627561	-0.17047	0.199689	0.207692
B	M 2	0.130588	0.143621	0.98082	0.559569	0.229755	0.989323	0.133688	0.394293	0.502624
B	M 3	0.037741	0.687201	0.991107	0.003974	-0.06846	0.9682	-0.06387	-0.04023	0.180505
B	M 4	-0.30521	-0.23881	0.999992	-0.13582	0.729414	0.999848	-0.29754	0.185612	0.017937
B	M 5	0.839805	0.955512	0.695489	0.952512	0.957565	0.473744	-0.11335	0.953043	0.958354
B	M 6	0.35927	0.444257	0.960541	0.635516	0.811722	0.947035	-0.09059	0.583446	0.70229
B	M 7	0.003464	0.571111	0.542834	0.14749	0.623363	0.514123	-0.13401	0.556791	0.625035
	average	0.146012	0.393404	0.848102	0.331489	0.49149	0.788548	-0.10516	0.404664	0.456348
		MPP-	B	CD4T	MPP+	CLP	CD8T	HSC	MEP	GMP
Gr	M 1	0.204141	0.190943	0.198656	0.77336	0.820365	-0.06095	-0.13827	0.864395	0.893742
Gr	M 2	0.574309	0.143621	0.113983	0.329383	0.396084	0.108154	-0.14175	0.62031	0.623211
Gr	M 3	0.13374	0.687201	0.659808	0.075694	-0.05473	0.572907	-0.04348	-0.03432	0.268827
Gr	M 4	0.951969	-0.23881	-0.23945	0.452566	-0.24247	-0.24568	0.874354	-0.14901	0.083942
Gr	M 5	0.915203	0.955512	0.501527	0.993639	0.994351	0.275297	-0.05672	0.998663	0.999016
Gr	M 6	0.949004	0.444257	0.354461	0.907646	0.839236	0.311746	0.040668	0.947192	0.910083
Gr	M 7	0.167506	0.571111	0.536925	0.333314	0.809845	0.476723	-0.10899	0.822394	0.809821
	average	0.556553	0.393404	0.303701	0.552229	0.508955	0.205457	0.060831	0.581375	0.65552
		MPP-	B	Gr	MPP+	CLP	CD8T	HSC	MEP	GMP
CD4T	M 1	-0.15721	0.765933	0.198656	0.119916	0.094496	0.765171	-0.22658	0.13991	0.165414

		MPP-	B	Gr	CD4T	MPP+	CLP	CD8T	MEP	GMP
HSC	M 1	0.816008	-0.17047	-0.13827	-0.22658	0.194054	0.234389	-0.10332	0.131918	-0.09343
HSC	M 2	0.431082	0.133688	-0.14175	0.201597	0.127773	-0.04145	0.167954	0.1114	0.09872
HSC	M 3	0.94736	-0.06387	-0.04348	-0.12252	-3.56E-05	0.056322	-0.11988	0.122102	0.099791
HSC	M 4	0.91094	-0.29754	0.874354	-0.29837	0.50101	-0.1724	-0.30395	-0.02884	0.452154
HSC	M 5	0.140705	-0.11335	-0.05672	-0.19341	-0.0399	-0.06836	-0.19176	-0.07143	-0.08538
HSC	M 6	-0.05913	-0.09059	0.040668	-0.19542	-0.1556	-0.15884	-0.21947	-0.14707	-0.1624
HSC	M 7	0.056732	-0.13401	-0.10899	-0.13225	-0.17244	-0.05466	-0.14793	-0.07474	-0.03339
	average	0.463386	-0.10516	0.060831	-0.13814	0.06498	-0.02929	-0.13119	0.006191	0.039438
		MPP-	B	Gr	CD4T	MPP+	CLP	CD8T	HSC	GMP
MEP	M 1	0.415896	0.199689	0.864395	0.13991	0.952458	0.986488	-0.07193	0.131918	0.960612
MEP	M 2	0.725771	0.394293	0.62031	0.495671	0.877871	0.778421	0.457236	0.1114	0.977891
MEP	M 3	0.160448	-0.04023	-0.03432	-0.04139	0.28671	0.952507	0.037543	0.122102	0.94702
MEP	M 4	-0.12119	0.185612	-0.14901	0.188332	0.62804	0.525799	0.191159	-0.02884	0.48843
MEP	M 5	0.921014	0.953043	0.998663	0.503459	0.994722	0.992609	0.277934	-0.07143	0.997687
MEP	M 6	0.881404	0.583446	0.947192	0.510823	0.980524	0.93626	0.468229	-0.14707	0.974415
MEP	M 7	0.313992	0.556791	0.822394	0.613732	0.591198	0.951638	0.547418	-0.07474	0.973402
	average	0.471048	0.404664	0.581375	0.344362	0.758789	0.874817	0.272513	0.006191	0.90278
		MPP-	B	Gr	CD4T	MPP+	CLP	CD8T	HSC	MEP
GMP	M 1	0.261053	0.207692	0.893742	0.165414	0.921361	0.929621	-0.06878	-0.09343	0.960612
GMP	M 2	0.71611	0.502624	0.623211	0.582412	0.855337	0.720646	0.553761	0.09872	0.977891
GMP	M 3	0.184216	0.180505	0.268827	0.171077	0.37494	0.914643	0.221867	0.099791	0.94702
GMP	M 4	0.199636	0.017937	0.083942	0.018652	0.454211	0.292532	0.018212	0.452154	0.48843
GMP	M 5	0.898179	0.958354	0.999016	0.505772	0.994503	0.997071	0.278209	-0.08538	0.997687
GMP	M 6	0.847847	0.70229	0.910083	0.660172	0.96983	0.981849	0.617017	-0.1624	0.974415
GMP	M 7	0.204161	0.625035	0.809821	0.708851	0.523626	0.982786	0.641777	-0.03339	0.973402
	average	0.473029	0.456348	0.65552	0.401764	0.727687	0.831307	0.323152	0.039438	0.90278