

# Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors

Jaeyop Lee<sup>1,4</sup>, Yu Jerry Zhou<sup>1,4</sup>, Wenji Ma<sup>2,4</sup> , Wanwei Zhang<sup>1</sup>, Arafat Aljoufi<sup>1</sup>, Thomas Luh<sup>1</sup>, Kimberly Lucero<sup>1</sup>, Deguang Liang<sup>1</sup>, Matthew Thomsen<sup>3</sup>, Govind Bhagat<sup>3</sup>, Yufeng Shen<sup>2</sup> & Kang Liu<sup>1</sup>

The origin and specification of human dendritic cells (DCs) have not been investigated at the clonal level. Through the use of clonal assays, combined with statistical computation, to quantify the yield of granulocytes, monocytes, lymphocytes and three subsets of DCs from single human CD34<sup>+</sup> progenitor cells, we found that specification to the DC lineage occurred in parallel with specification of hematopoietic stem cells (HSCs) to the myeloid and lymphoid lineages. This started as a lineage bias defined by specific transcriptional programs that correlated with the combinatorial ‘dose’ of the transcription factors IRF8 and PU.1, which was transmitted to most progeny cells and was reinforced by upregulation of IRF8 expression driven by the hematopoietic cytokine FLT3L during cell division. We propose a model in which specification to the DC lineage is driven by parallel and inheritable transcriptional programs in HSCs and is reinforced over cell division by recursive interactions between transcriptional programs and extrinsic signals.

Efforts to construct generally accepted and coherent hierarchical relationships for the development of dendritic cells (DCs) have proven contentious<sup>1–4</sup>. This debate has been fueled by the observation that progenitor cells from either the myeloid branch or the lymphoid branch give rise to the same DC subsets<sup>5,6</sup> and by the fact that progenitor cells defined by the current markers are heterogeneous<sup>7–9</sup>. Moreover, most studies have focused on qualitative potency, and thus multipotency has traditionally been interpreted as equipotency<sup>10</sup>. In addition, suitable ways of quantifying, mathematically analyzing and identifying the significance of potency differentials have not been available. Single-cell RNA-based next-generation sequencing and functional clonal analysis have been used to reassess the homogeneity of progenitor subsets defined by the current markers<sup>8,11–13</sup>. Single-cell transplantation<sup>14</sup> and endogenous bar-coding<sup>15</sup> have suggested that most mouse myeloid cells are derived from hematopoietic stem cells (HSCs) that are restricted to the myeloid lineage, which has led to the idea of ‘early imprinting or commitment’ at the HSC stage<sup>10</sup>. However, specification to the human DC lineage has not been studied at single-cell resolution. In the mouse, expression of the gene encoding the transcription factor IRF8 and the function of IRF8 in regulating the development of DCs and monocytes are thought to occur after the lymphoid-primed multipotent progenitor (LMPP) stage<sup>9,16,17</sup>. However, the role and timing of the expression and regulation of IRF8 in specification to the human DC lineage remains unclear.

Here we investigated the developmental potency of human hematopoietic progenitor cells at the single-cell level and used quantitative analysis of clonal output to investigate the development of granulocytes, monocytes, CD141<sup>+</sup> conventional DCs (DC1 cells), CD1c<sup>+</sup> conventional DCs (DC2 cells), plasmacytoid DCs (pDCs) and lymphocytes from single cord-blood CD34<sup>+</sup> cells. We found that non-unipotent progenitor cells exhibited inherent lineage bias that was established *in vivo* in HSCs and was transmitted to most progeny. The combinatorial ‘dose’ of the transcription factors IRF8 and PU.1 was highly correlated with specific lineage biases, while the hematopoietic cytokine FLT3L drove and maintained the DC lineage program throughout cell division. These results indicate that the combinatorial dose of a common set of transcription factors in HSCs and multipotent progenitors (MPPs) can shape parallel and inheritable programs for distinct hematopoietic lineages, which are then reinforced through recursive interaction with environmental cytokines.

## RESULTS

### Functional heterogeneity of hematopoietic progenitor subsets

To map the developmental relationships among the DC, myeloid and lymphoid lineages, we isolated human CD34<sup>+</sup> hematopoietic progenitor cells from cord blood and divided them into ten non-overlapping progenitor populations: CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup>CD10<sup>-</sup>CD90<sup>+</sup> HSCs, CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>-</sup>CD10<sup>-</sup>CD90<sup>-</sup> MPPs, CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>+</sup>CD10<sup>-</sup>

<sup>1</sup>Department of Microbiology and Immunology, Columbia University Medical Center, New York, New York, USA. <sup>2</sup>Department of Systems Biology and Department of Biomedical Informatics, Columbia University Medical Center, New York, New York, USA. <sup>3</sup>Department of Pathology and Cell Biology, Columbia University Medical Center, New York, New York, USA. <sup>4</sup>These authors contributed equally to this work. Correspondence should be addressed to Y.S. (ys2411@cumc.columbia.edu) or K.L. (kl2529@cumc.columbia.edu).

Received 19 January; accepted 2 June; published online 26 June; corrected online 5 July 2017 (details online); doi:10.1038/ni.3789

**Table 1** Characterization of progenitor populations in human cord blood and bone marrow

Population	Phenotype	Cells (%)		Lineage output							Ref	
		CB	BM	CFU			MP plus FSG					
				ME	G	M	G	M	DC	L		
HSC	CD34 <sup>+</sup> CD38 <sup>-</sup> CD45RA <sup>-</sup> CD90 <sup>+</sup>	3.79	1.70	+	+	+	+	+	+	+	+	19
MPP	CD34 <sup>+</sup> CD38 <sup>-</sup> CD45RA <sup>-</sup> CD90 <sup>-</sup>	19.0	11.3	+	+	+	+	+	+	+	+	19
LMPP	CD34 <sup>+</sup> CD38 <sup>-</sup> CD45RA <sup>+</sup> CD10 <sup>-</sup>	1.36	16.5	-	+	+	+	+	+	+	+	20
MLP	CD34 <sup>+</sup> CD38 <sup>-</sup> CD45RA <sup>+</sup> CD10 <sup>+</sup> CD7 <sup>+</sup> or CD <sup>-</sup>	4.88	0.80	-	-	+	-	+	+	+	+	19
BNKP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>+</sup> CD123 <sup>int-neg</sup> CD115 <sup>-</sup> CD10 <sup>+</sup>	3.12	2.20	-	-	-	-	-	-	-	+	19
CMP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>-</sup> CD10 <sup>-</sup> CD123 <sup>int</sup>	40.0	33.8	+	+	+	+	+	+	+	-	18
GMDP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>-</sup> CD10 <sup>-</sup> CD123 <sup>int</sup>	11.4	11.2	-	+	+	+	+	+	+	-	7
MDP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>+</sup> CD123 <sup>int</sup> CD115 <sup>+</sup>	0.81	5.60	-	-	+	-	+	+	+	-	7
CDP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>-</sup> CD123 <sup>hi</sup> CD115 <sup>-</sup>	0.54	3.00	-	-	-	-	-	-	+	-	7
MEP	CD34 <sup>+</sup> CD38 <sup>+</sup> CD45RA <sup>-</sup> CD10 <sup>-</sup> CD123 <sup>-</sup>	15.3	11.4	+	-	-	-	-	-	-	-	18

Characterization of various subsets of progenitor cells (far left), including phenotype (second column), frequency of the population (from cord blood (CB) or bone marrow (BM)) among CD34<sup>+</sup> cells (third header), output (right), as assessed by colony-forming unit assay (CFU) or by culture on MP plus FSG. MEP, megakaryocyte-erythroid progenitor.

LMPPs, CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>+</sup>CD10<sup>+</sup> multi-lymphoid progenitors (MLPs), CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD10<sup>+</sup> B cell–natural killer (NK) cell progenitors (BNKPs), CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>-</sup>CD10<sup>-</sup>CD123<sup>+</sup> common myeloid progenitors (CMPs), CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD10<sup>-</sup>CD123<sup>+</sup>CD115<sup>-</sup> granulocyte-monocyte-DC progenitors (GMDPs), CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD10<sup>-</sup>CD123<sup>+</sup>CD115<sup>+</sup> monocyte-DC progenitors (MDPs), CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD10<sup>-</sup>CD123<sup>hi</sup>CD115<sup>-</sup> common DC progenitors (CDPs) and CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>-</sup>CD10<sup>-</sup>CD123<sup>-</sup> megakaryocyte-erythroid progenitors<sup>7,18–20</sup> (Table 1 and Fig. 1a). Because megakaryocyte-erythroid progenitors do not produce DCs or lymphoid or myeloid cells<sup>18,19</sup>, we evaluated the potential of the other nine progenitor populations to develop into eight mature lineages—the granulocyte (G), monocyte (M) megakaryocyte (Mk), erythrocyte (Er) and lymphocyte (L) lineages; and three DC subsets (DC1, DC2 and pDC)—using two *in vitro* systems: a colony-formation assay for the G, M, Mk and Er lineages (Supplementary Fig. 1a), and a culture containing MS5 and OP9 stromal cells and the cytokines FLT3L, SCF and GM-CSF (called ‘MP plus FSG’ here) to assess the G, M, L, DC1, DC2 and pDC lineages (Fig. 1b). Due to the lack of signaling via NOTCH receptors in the MP plus FSG culture, the L lineage is represented only by the output of B cells and NK cells. As expected, HSCs and MPPs produced all lineages and CMPs and GMDPs did not produce L cells, while LMPPs, MLPs and BNKPs did not produce cells of the Mk or Er lineage (Fig. 1b and Supplementary Fig. 1a). However, LMPPs and MLPs produced cells of the G and M lineages and the three DC subsets (Fig. 1b), indicative of some myeloid potential.

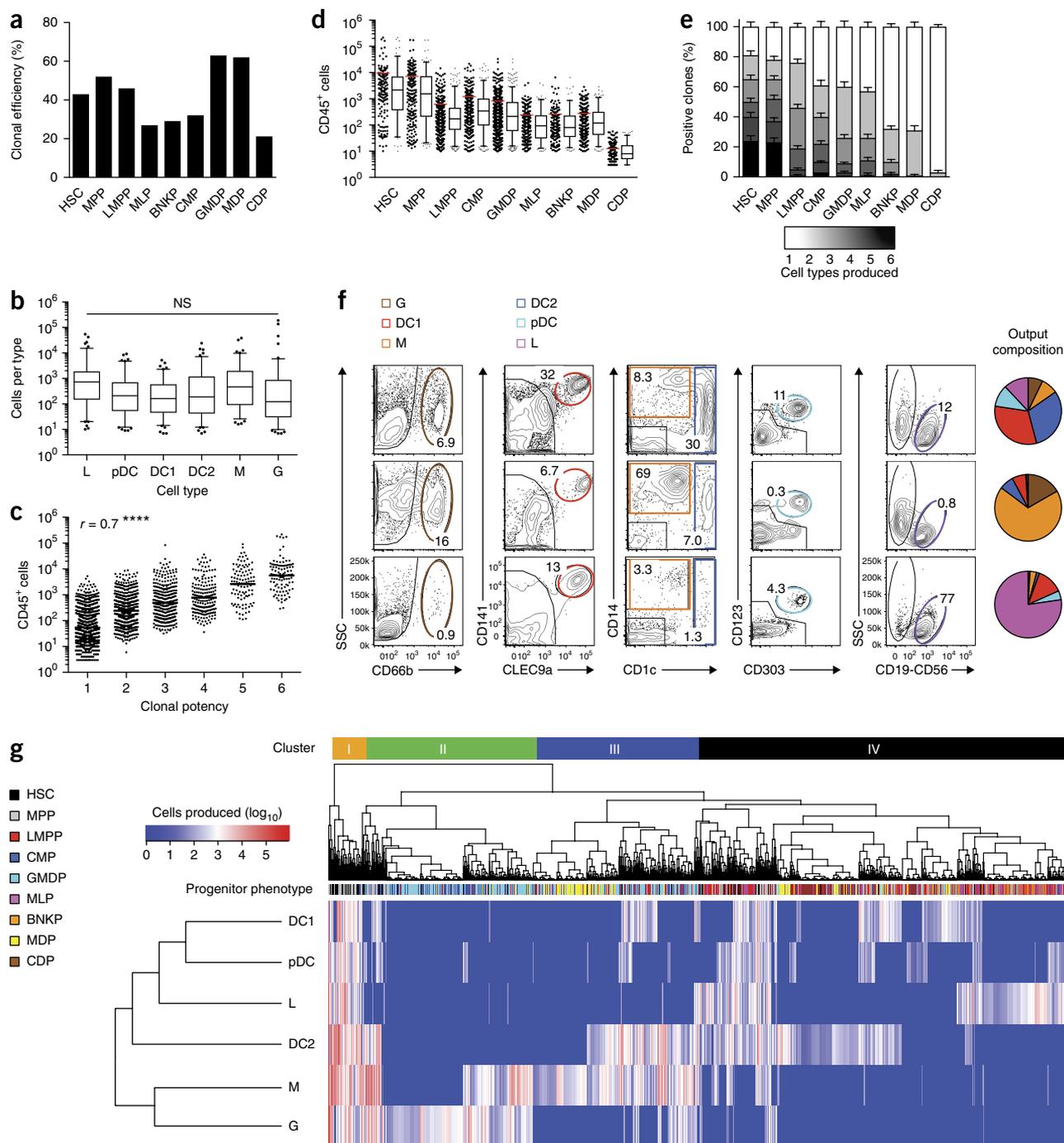
To determine the developmental sequence of the nine progenitor subsets, we labeled HSCs with the division-tracking dye CFSE (carboxyfluorescein succinimidyl ester), then cultured the labeled cells for 7 d on MP plus FSG, a period that allows the differentiation of intermediate progenitor cells<sup>7,21</sup> (Fig. 1c and Supplementary Fig. 1b). Various progenitor subsets were observed after a certain number of divisions: CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>+</sup>CD7<sup>-</sup> LMPPs appeared at divisions 1–2; CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>-</sup>CD7<sup>-</sup> CMPs and CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD7<sup>-</sup>CD123<sup>+</sup>CD115<sup>-</sup> GMDPs appeared at division 3; CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>+</sup>CD7<sup>+</sup> BNKPs and CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD7<sup>-</sup>CD123<sup>+</sup>CD115<sup>+</sup> MDPs appeared at division 5; and CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>+</sup>CD7<sup>-</sup>CD123<sup>hi</sup>CD115<sup>-</sup> CDPs appeared at division 7 (Fig. 1c). This indicated a hierarchy among progenitor phenotypes. When individual progenitor populations were cultured for 7 d *in vitro* at a density of 100 cells per well, HSC-MPPs produced both CD34<sup>+</sup>CD38<sup>+</sup>CD45RA<sup>-</sup>CD7<sup>-</sup> CMPs and CD34<sup>+</sup>CD38<sup>-</sup>CD45RA<sup>+</sup>CD7<sup>-</sup> LMPPs, CMPs and LMPPs did not differentiate into each other, LMPPs produced MLPs and BNKPs, CMPs

produced GMDPs, and GMDPs produced MDPs and CDPs (Fig. 1d and Supplementary Fig. 1c). In addition, MLPs produced GMDPs (Fig. 1d). Similar results were obtained at 7 d after *in vivo* transfer of HSC-MPPs, CMPs and LMPPs (1 × 10<sup>4</sup> cells each) intratibially into host mice of the non-obese diabetic–severe combined immunodeficiency (NOD-SCID) strain that lacked the cytokine receptor IL-2Rγ (NOD-SCID-IL-2Rγ<sup>null</sup> (NSG) mice) (Supplementary Fig. 1d). These results indicated that MPPs gave rise to CMPs and LMPPs, and that CMPs, LMPPs and MLPs all gave rise to GMDPs (Supplementary Fig. 1e).

Next we analyzed 5,559 single progenitor cells (each called a ‘clone’ here), including HSCs, MPPs, LMPPs, CMPs, MLPs, BNKPs, GMDPs, MDPs and CDPs, from human cord blood, of which 2,247 gave rise to progeny in the MP plus FSG culture (Fig. 2a). Of the 2,247 clones assessed, 105 clones were multipotent and generated all six lineages, including the L, G, M, DC1, DC2 and pDC lineages, and the average clonal yield of each lineage was statistically indistinguishable, ranging between 620 cells and 3,465 cells (Fig. 2b); this indicated that these culture conditions did not create bias toward any lineage. We divided the 2,247 clones into six groups on the basis of the number of lineages produced by each clone. The 105 clones that generated six lineages produced the largest number of CD45<sup>+</sup> progeny, while the 923 unipotent clones produced the smallest number of CD45<sup>+</sup> cells (Fig. 2c), which indicated that hematopoietic differentiation correlated with loss of lineage and proliferation potential. We then compared the clonal yield (Fig. 2d) and lineage yield (Fig. 2e) of all 2,247 clones grouped via progenitor subset. Although ranking the progenitor subsets by mean clonal yield correlated with ranking by developmental hierarchy, the yield of individual clones within each progenitor subset varied by orders of magnitude (Fig. 2d and Supplementary Fig. 2a). 24% of HSCs, 23% of MPPs and 0% of all other progenitor cells produced six lineages; all progenitor subsets displayed marked variation in lineage yield (Fig. 2e and Supplementary Fig. 2b). Notably, although all these progenitor cells are defined as common progenitors of several lineages<sup>7,19</sup> and are thus expected to produce more than one lineage, each population had many unipotent progenitor cells (Fig. 2e and Supplementary Fig. 2b), in confirmation of published observations<sup>22</sup>; this suggested that lineage specification might occur very early. Using flow cytometry to quantify terminally differentiated cells of each lineage (G, M, L, DC1, DC2 and pDC), we observed that the yield of various lineages, or the ‘quantitative potency’ of a given clone, was highly variable (Fig. 2f and Supplementary Table 1), which indicated that the multipotent progenitor cells were not equipotent. Unsupervised hierarchical clustering of 2,247 clones via their



**Figure 1** Marker-defined hematopoietic progenitors exhibit hierarchical and convergent potency. **(a)** Flow cytometry of human cord blood, showing the gating of progenitor populations with a starting gate of lineage-negative ( $\text{Lin}^-$ ) cells ( $\text{CD}3^- \text{CD}19^- \text{CD}56^- \text{CD}14^- \text{CD}16^- \text{CD}66\text{b}^- \text{CD}1\text{c}^- \text{CD}303^- \text{CD}141^-$ ). Numbers adjacent to outlined areas indicate percent cells in each subset (blue and red labels). **(b)** Flow cytometry of live, singlet  $\text{CD}45^+$  cells of various populations (left margin;  $n = 100$  cells in each) cultured for 14 d on MP plus FSG, showing the output of each (key). Numbers adjacent to outlined areas indicate percent cells in each subset (key) among total  $\text{CD}45^+$  cells. SSC, side scatter. **(c)** Flow cytometry of HSCs ( $n = 1,000$ ) sorted as in **a**, labeled with CFSE and cultured for 7 d on MP plus FSG, presented as concatenated plots showing the number of cell divisions (CFSE signal dilution) of various descendant populations (right margin; gated as in **d**). **(d)** Flow cytometry of HSC-MPPs, CMPs, GMDPs, LMPPs, MLPs and BNKPs (left margin;  $n = 1,000$  live, singlet  $\text{CD}45^+$   $\text{Lin}^- \text{CD}34^+$  cells ( $\text{Lin}^-$  defined as in **a**) of each population) after culture for 7 d on MP plus FSG, showing intermediate output (key). Numbers adjacent to outlined areas indicate percent cells in each subset (key) among total  $\text{CD}34^+$  cells. Data are from one experiment representative of 17 experiments, each with one of seventeen samples (**a**), or are from one experiment representative of five independent experiments (**b**) or four independent experiments (**c, d**).



**Figure 2** Clonal potency indicates heterogeneity of marker-pure progenitor populations and developmental distance from HSCs. **(a)** Clonal efficiency of HSCs ( $n = 360$ ), MPPs ( $n = 408$ ), LMPPs ( $n = 791$ ), MLPs ( $n = 720$ ), BNKPs ( $n = 542$ ), CMPs ( $n = 800$ ), GMDPs ( $n = 890$ ), MDPs ( $n = 357$ ) and CDPs ( $n = 691$ ) after clonal culture for 2–3 weeks on MP plus FSG; results are presented as the frequency of productive clones among total wells seeded. **(b)** Yield of each cell type (horizontal axis) from all multipotent clones ( $n = 105$ ). NS, not significant ( $P > 0.05$ ), multiple unpaired two-tailed Student's  $t$ -test. **(c)** Degree of correlation between clonal yield and potential.  $r$ , Spearman correlation factor; \*\*\*\* $P < 0.0001$ . **(d)** Yield of CD45<sup>+</sup> cells from all clones in each population (horizontal axis). **(e)** Qualitative potency of clones from each progenitor population (horizontal axis), presented as the frequency of clones producing one cell type (unipotent) to six different cell types (multipotent) (key). **(f)** Flow cytometry (left) of live CD45<sup>+</sup> cells, showing the phenotype produced from three individual multipotent clones (one per row); right, relative abundance of output cells of each (colors match key). Numbers adjacent to or in outlined areas (left) indicate percent cells in each subset (key). **(g)** Output (normalized values; top left key) of all six mature blood cell types (left margin; one per row) from each single progenitor cell (one per column (bar above plot); subset colors in far left key); top, four main clusters (I–IV) identified by unsupervised hierarchical clustering (brackets above plot and along left margin). Each symbol **(c,d)** represents an individual cell; small horizontal lines (red in **d**) indicate the mean. Data are representative of 17 experiments (**a–g**); in **b,d**: center line, median; vertical lines, 5% limit (upward) and 95% limit (downward), and dots indicate outliers **(b)** or all clones **(d)**; in **e**: bars, standard error of proportion of total positive clones).

quantitative potency revealed four main clusters that reflected progression of cell development: cluster I comprised highly proliferative and multipotent cells with five- or six-lineage developmental capacity; cluster II and cluster III consisted of oligopotent and unipotent progenitor cells with bias toward the G lineage and M lineage, respectively; and cluster IV comprised oligopotent and unipotent progenitor cells that tended to give rise to the L, DC1, DC2 or pDC lineage (Fig. 2g). All nine progenitor subsets analyzed were highly heterogeneous and were located in multiple clusters in aggregate (Fig. 2g) or as filtered by donor (Supplementary Fig. 2c). Thus, the progenitor subsets were heterogeneous but they could be ordered on a differentiation hierarchy on the basis of their proliferation potential.

### Quantitative clonal potency reveals lineage bias in progenitors

We sought to determine whether quantitative potency could determine each progenitor clone's developmental capacity. CFSE-labeled HSC-MPPs were cultured on MP plus FSG or were injected intratibially into NSG mice, were purified after three or six divisions, corresponding to intermediate or late developmental stages, and were evaluated in terms of clonal output (Supplementary Fig. 3a). When total progeny yield and lineage yield of each clone were compared, HSC-MPPs isolated after three divisions had lower progeny and lineage yields than those of undivided HSC-MPPs (Fig. 3a), and the magnitude of this decrease was even greater after six total divisions (Fig. 3a), which indicated that each clone's quantitative potency inversely correlated with developmental distance from HSCs.

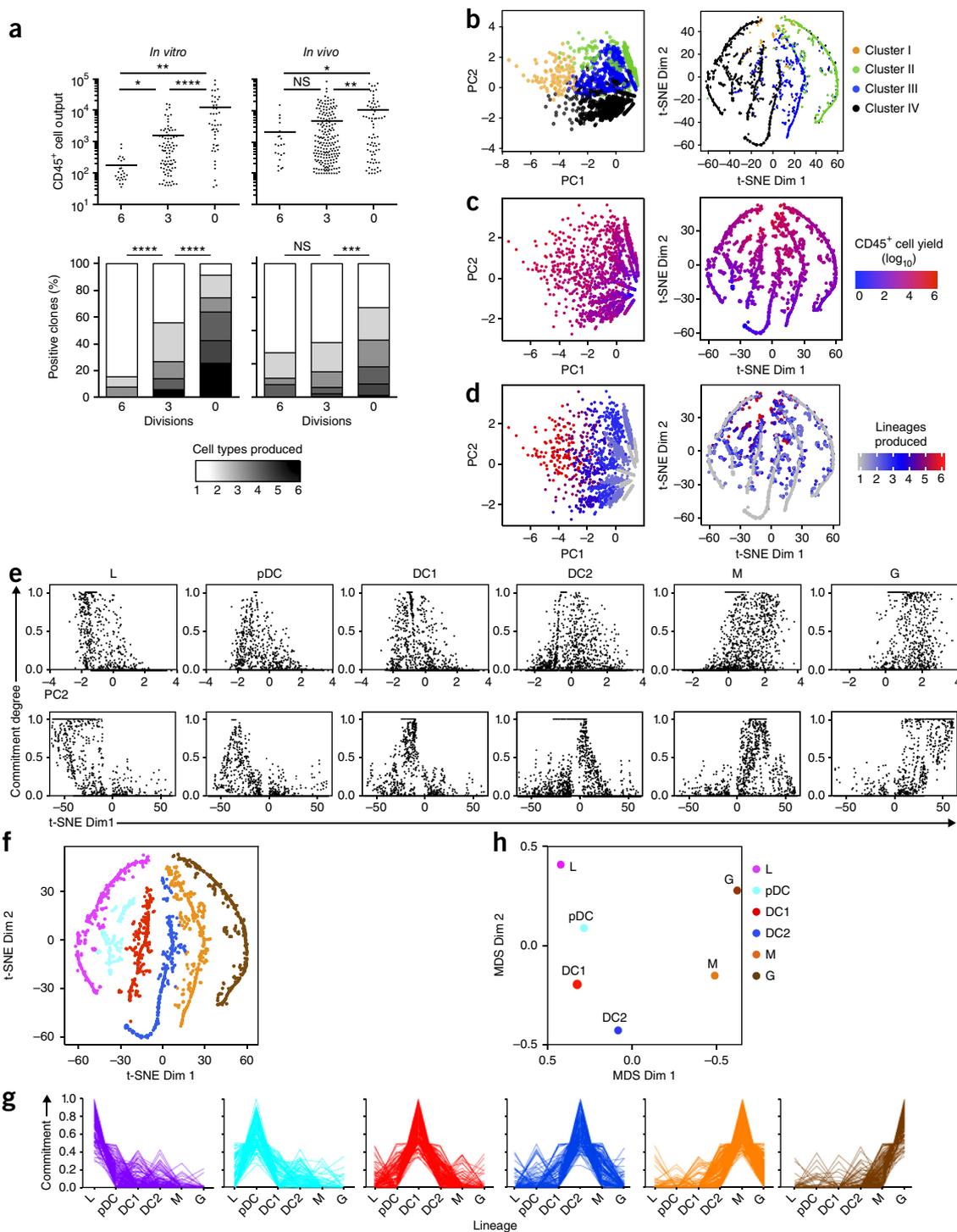
To investigate the developmental relationships among all progenitor clones, we analyzed the similarity of the 2,247 clones as determined by their quantitative potency. Each clone's quantitative potency was described as a six-dimensional vector on its output of each of the six lineages (G, M, L, DC1, DC2 and pDC), and their potency similarity was analyzed by principal-component analysis, which converts data into linearly uncorrelated variables, and by *t*-distributed stochastic neighbor embedding (*t*-SNE) combined with a Gaussian kernel diffusion model<sup>23,24</sup>, which preserves local structure in multi-dimensional space, to generate two-dimensional maps. Both analyses generated the same four clusters (I–IV) (Fig. 3b) that were identified by hierarchical clustering (Fig. 2g), with one dimension correlating with proliferative capacity (Fig. 3c) or number of lineages generated (Fig. 3d), and the other dimension's coordinate correlating with the predominant lineage yield (Fig. 3e). *t*-SNE allows the generation of a visualization map in which clones on a given track generate predominantly one lineage but are ordered in the spectrum from multipotency to unipotency and from high yield to low yield (Fig. 3c,d). Thus, progenitor clones that produced predominantly cells of the G, M, L, DC1, DC2 or pDC lineage fell on separate tracks (Fig. 3f and Supplementary Fig. 3a,b), and all clones on a given track had the same lineage bias, producing cells of one lineage in greater numbers than those of other lineages (Fig. 3f,g). When all clones' quantitative potency was used to compute the degree of ancestry sharing, the L and G lineages were considerably less likely to share ancestry than either the M and G lineages or the L and pDC lineages (Fig. 3h). Therefore, the distances between the lineage tracks reflected the likelihood of 'shared ancestry'. Critically, although repeating *t*-SNE mapping generated different maps, the clustering pattern was highly consistent (Supplementary Fig. 3d). These results indicated that such quantitative potency offered a meaningful indicator of a progenitor cell's developmental capacity, which allowed the grouping of progenitor cells on the basis of their predominant lineage yield and continuum of yield and lineage restriction.

### Hematopoietic lineage bias starts in HSCs

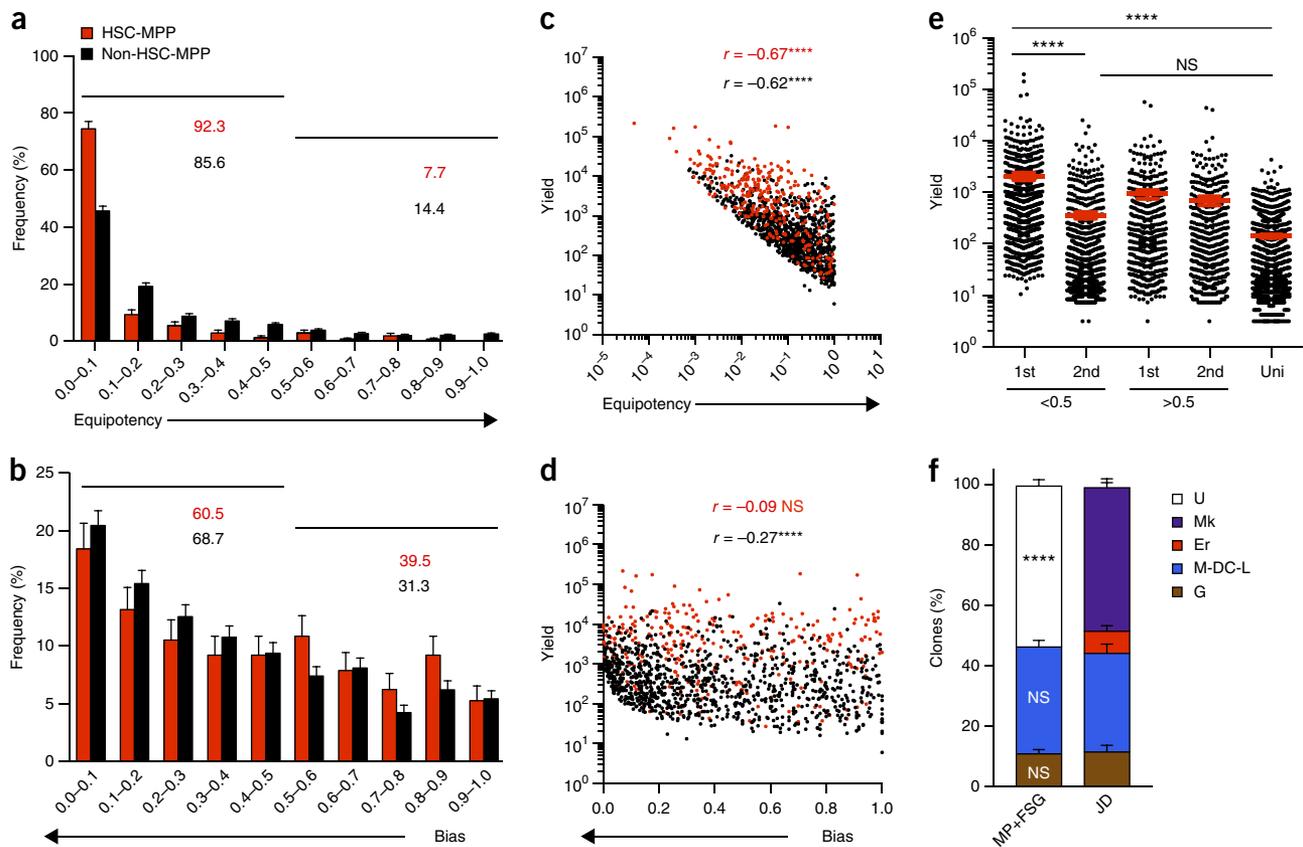
To determine whether the non-unipotent progenitor cells were equipotent, as assumed by classical differentiation models, or showed lineage bias, we calculated the 'equipotency ratio' of all non-unipotent clones by dividing the smallest lineage yield by the largest lineage yield; here, a ratio of 1 indicates a truly unbiased (i.e., equipotent) clone. Of 1,324 non-unipotent clones, 152 clones had a ratio of >0.5, and 1,172 had a ratio of <0.5 (Fig. 4a), which indicated that the vast majority of progenitor cells were not equipotent. Of HSCs and MPPs, 92.3% had a ratio of <0.5; among all other non-HSC-MPP clones, 85.6% had a ratio of <0.5 (Fig. 4a). This indicated that even the HSC and MPP clones were not equipotent. We also calculated the 'bias ratio' by dividing the second-largest lineage yield by the maximum lineage yield; here, a ratio of 0 indicates a wholly biased clone. We observed that 66.7% of non-unipotent progenitor cells, which included HSCs-MPPs, showed a bias ratio of < 0.5 (Fig. 4b), indicative of lineage bias.

To exclude the possibility that the lineage bias was due to artifacts *in vitro*, we first sought to determine whether the cultured multipotent progenitor clones were initially equipotent and the bias was caused by stochastic death of the progeny. We plotted bias degree against the yield of all non-unipotent clones and observed that highly biased HSC-MPPs or oligopotent progenitor cells tended to have higher offspring yields, whereas equipotent progenitors tended to have lower yields (Fig. 4c,d). Because stochastic death of progeny would diminish yields, this indicated that lineage bias was not caused by progeny death. Next, to address whether the bias was caused by random lineage expansion during culture, we compared the largest lineage yield from 878 biased progenitor clones that produced a single major lineage (bias ratio, <0.5) with the largest lineage yield from 438 unbiased clones with two major lineages (bias ratio, >0.5). The largest lineage yields of biased clones were significantly higher than those of unbiased clones (Fig. 4e), which indicated that lineage bias was the product of neither stochastic death nor random lineage expansion *in vitro* but was instead the product of *in vivo* establishment before isolation and culture and was intrinsically correlated with proliferative capacity.

To address whether the lineage bias was caused by the microenvironment of the medium, we compared the clonal composition of HSC-MPPs in MP plus FSG culture and in a different culture system of MS5 stromal cells and the cytokines SCF, FLT3L, TPO, EPO, IL-6, IL-3, IL-11 and GM-CSF (called 'JD culture' here), which supports differentiation of the Er and Mk lineages in addition to that of the G, M, DC and L lineages<sup>22</sup> (Supplementary Fig. 4a,b). In terms of clonal efficiency, 52% of total HSC-MPP clones were unproductive on MP plus FSG, while 2% of HSC-MPP clones were unproductive in the JD culture (Fig. 4f), which indicated that HSCs and MPPs were neither totipotent nor equipotent, as totipotent and equipotent HSC-MPP clones would expand to display similar clonal efficiency in either culture system. The clonal composition of the G and M-DC-L lineages was 46.75% in MP plus FSG and 44.56% in JD culture (Fig. 4f), which indicated that the culture conditions did not induce lineage biases on the basis of cytokine composition and concentration, which were different in the two cultures. About 5% of HSC-MPP clones produced all lineages in the JD system, versus ~11% in the MP plus FSG system (Supplementary Fig. 4c), which suggested that they were totipotent progenitor cells. On average, totipotent HSC-MPPs were more proliferative than non-totipotent HSC-MPPs were (Supplementary Fig. 4d). However, the HSC-MPP clones with the highest clonal yield were not totipotent (Supplementary Fig. 4d). In addition, the totipotent HSC-MPP clones were not equipotent and exhibited lineage bias, like the non-totipotent clones, both on MP plus FSG and in JD



**Figure 3** Statistical modeling of clonal potency reveals developmental patterns and lineage biases. **(a)** Clonal output (top) and potency composition (bottom) of CD45<sup>+</sup> cells derived from divisions 0, 3, and 6 of CFSE-labeled HSCs cultured for 6 d on MP plus FSG (*In vitro*) or transferred into NSG mice for 6 d (*In vivo*). **(b–d)** Principal-component analysis (left) and t-SNE (right) showing clustered clonal data (from **Fig. 2a**) in terms of assigned cluster **(b)**, yield **(c)** and lineages produced **(d)**. Each symbol represents an individual cell. Dim, dimension. **(e)** Pattern similarities between principal-component analysis and t-SNE for clones plotted according to their degree of commitment toward a specific cell type (above plots) versus either principle component 2 (PC2) (top row) or t-SNE dimension 1 (Dim 1) (bottom row). **(f)** t-SNE with each clone (one per symbol) assigned to a track on the basis of its shortest distance to the backbone of corresponding lineage track. **(g)** Degree of commitment of each clone (one per line), calculated in terms of offspring composition of each on the tracks in **f**. **(h)** Multidimensional scaling (MDS) showing developmental relationships among six cell types (key) in terms of the likelihood that two lineages would arise from a common progenitor. Each symbol **(a–f)** represents an individual cell; small horizontal lines **(a)** indicate the mean. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  and \*\*\*\* $P < 0.0001$  (unpaired two-tailed Student's *t*-test **(a, top)** or Fisher's exact test on the frequency of unipotent cells **(a, bottom)**). Data are from three independent experiments with results pooled from clones **(a)** or are representative of seventeen experiments with seventeen donors (one donor in each) **(b–h)**.



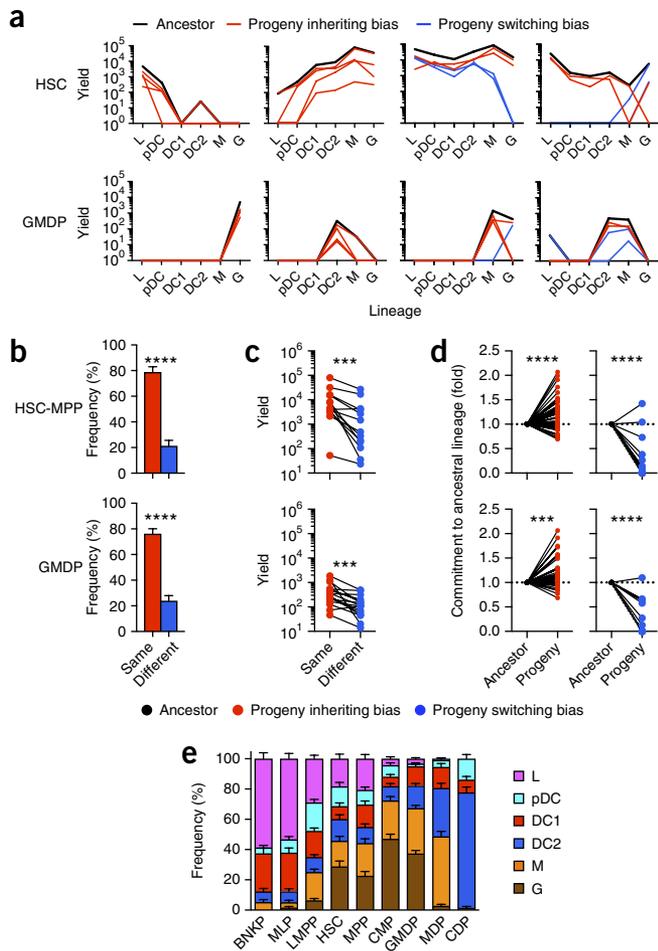
**Figure 4** Lineage bias is prevalent and starts early in HSCs. **(a,b)** Frequency distribution of all non-unipotent clones based on their degree of equipotency (horizontal axis), determined by the ratio of minimal lineage yield to maximal lineage yield **(a)** or by their degree of cell-type-specific potency bias (horizontal axis), determined by the ratio of second-highest lineage yield to maximal lineage yield **(b)**. Numbers below horizontal lines at top indicate the cumulative frequency of clones for which the ratio was  $<0.5$  (left line) or  $>0.5$  (right line). **(c,d)** Correlation between equipotency **(c)** or bias degree **(d)** and clonal yield for clones in **a,b**.  $r$ , Spearman correlation coefficient (colors match symbols in plot). **(e)** Yield of the largest lineage (1st) and second-largest lineage (2nd) produced by non-unipotent clones whose bias was  $<0.5$  ( $n = 115$ ) or  $>0.5$  ( $n = 162$ ), and of unipotent progenitor cells (Uni;  $n = 931$ ). **(f)** Proportion of HSC-MPP clones that were either unproductive (U) or biased toward the Mk, Er, M-DC-L or G lineage (key) on MP plus FSG (left;  $n = 768$  clones) or in JD culture (right;  $n = 286$  clones). \* $P < 0.05$  and \*\*\*\* $P < 0.0001$  (one-way analysis of variance **(e)** or Fisher's exact test of the proportion of productive and non-productive clones, or of the M-DC-L and G lineages between the two culture systems **(f)**). Each symbol **(c–e)** represents an individual cell; small red horizontal lines **(e)** indicate the mean ( $\pm$  s.e.m.). Data are representative of 17 experiments with 17 donors (MP+FSG) or three experiments with three donors (JD) (one donor in each), with results pooled from all clones (standard error of proportion in **a,b,f**).

culture (**Supplementary Fig. 4e,f**). These observations indicated that multipotent progenitor cells were not equipotent, that most progenitor cells—including rare totipotent clones—had an inherent lineage bias that was established *in vivo* early in HSCs, and that there was a correlation between lineage bias and proliferative capacity.

### Lineage bias is heritable and transmitted to progeny

To evaluate if lineage bias was maintained through the differentiation of progenitor cells into their progeny, we labeled single HSCs and GMDPs with the fluorescent dye DiD and cultured the labeled cells for 2–4 d on MP plus FSG, then individually cultured each of the four granddaughter cells for 2 more weeks (**Supplementary Fig. 5a**). We measured each granddaughter's quantitative potency and inferred each ancestor's quantitative potency as the sum of its granddaughters'. Tracing 198 granddaughter cells showed that the majority of HSC and GMDP progeny produced the same predominant lineage as their ancestor, suggestive of lineage inheritance, although some progeny produced a predominant lineage different from their ancestor, suggestive of bias 'switching' (**Fig. 5a** and **Supplementary Fig. 5b,c**). To quantify the relative rate of bias inheritance and bias switching, we

compared the lineage bias of each granddaughter cell with that of its ancestor. Notably, 79.6% of HSC progeny and 76.1% of GMDP progeny inherited ancestral bias, and 20.4% of HSC progeny and 23.9% of GMDP progeny switched bias to a different lineage (**Fig. 5b**), which indicated that the majority of progeny inherited ancestral bias. We then compared the clonal yield of the bias-inheriting progeny with that of their bias-switching siblings. The clonal yield of bias-inheriting progeny was significantly higher than that of bias-switching progeny (**Fig. 5c**). There was a significantly higher degree of commitment among all bias-inheriting progeny than among bias-switching ones (**Fig. 5d**), which indicated that bias-inheriting progeny amplified their inherited bias. For bias-switching progeny, there was considerable flexibility in the bias-switching direction, such that ancestors biased toward the G, M, DC1, DC2, pDC or L lineage could give rise to progeny with any other lineage bias (**Supplementary Fig. 5b,c**). However, bias switches in GMDP progeny were more likely to occur between the G and M lineages, M and DC2 lineages, or DC1 and DC2 lineages, while HSC progeny could switch between the G and L lineages (**Supplementary Fig. 5b,c**); this indicated a greater degree of bias-switch flexibility in HSCs than in GMDPs.



**Figure 5** Lineage bias is transmitted to most progeny and can be further amplified toward full commitment along division. **(a)** Yield for six lineages (horizontal axis) from ancestors and progeny with potency profiles that are similar to (inheriting bias) or different from (switching bias) that of the ancestral clone (key). **(b)** Frequency of granddaughter cells from HSCs ( $n = 89$ ) or GMDPs ( $n = 109$ ) that inherited or switched their ancestor's lineage bias (key). **(c)** Yield (average values) of siblings of HSCs ( $n = 13$ ) or GMDPs ( $n = 16$ ) with inherited or switched bias (key). **(d)** Change in the degree of commitment in the progeny of HSCs or GMDPs that inherited their ancestor's lineage bias (left;  $n = 66$  (HSC) or  $n = 47$  (GMDP)) or switched from it (right;  $n = 19$  (HSC) or  $n = 26$  (GMDP)). **(e)** Composition of clonal bias for progenitor subtypes isolated from cord blood. \*\*\* $P < 0.001$  and \*\*\*\* $P < 0.0001$  (unpaired two-tailed Student's  $t$ -test **(b)** or paired two-tailed Student's  $t$ -test **(c,d)**). Lines in **c,d** connect results for the same ancestor. Data are representative of three experiments with three donors **(a–d)**; result pooled from clones) or seventeen experiments with seventeen donors (one donor in each) **(e)**; result pooled from clones; mean and standard error of proportion (calculated from total positive clones for each progenitor)).

Next we calculated the frequency of clones biased toward different lineages in each of the marker-defined progenitor populations. Each population comprised groups of clones biased toward distinct lineages, and the proportion of these lineage groups was distinct and characteristic for each population analyzed (**Fig. 5e**). On the t-SNE visualization map, clones within marker-defined progenitor populations were distributed across multiple tracks of distinct lineage bias (**Supplementary Fig. 5d**), which indicated that clones attributed to each progenitor population by markers could fall on any track, consistent with the unique transcriptional patterns described by analysis

of mouse CMPs and GMPs by single-cell RNA-based next-generation sequencing<sup>13</sup>. We concluded that the hematopoietic lineage bias was established *in vivo* in HSCs and was heritable and amplified during proliferation.

### IRF8 expression marks DC1-lineage specification in HSC-MPPs

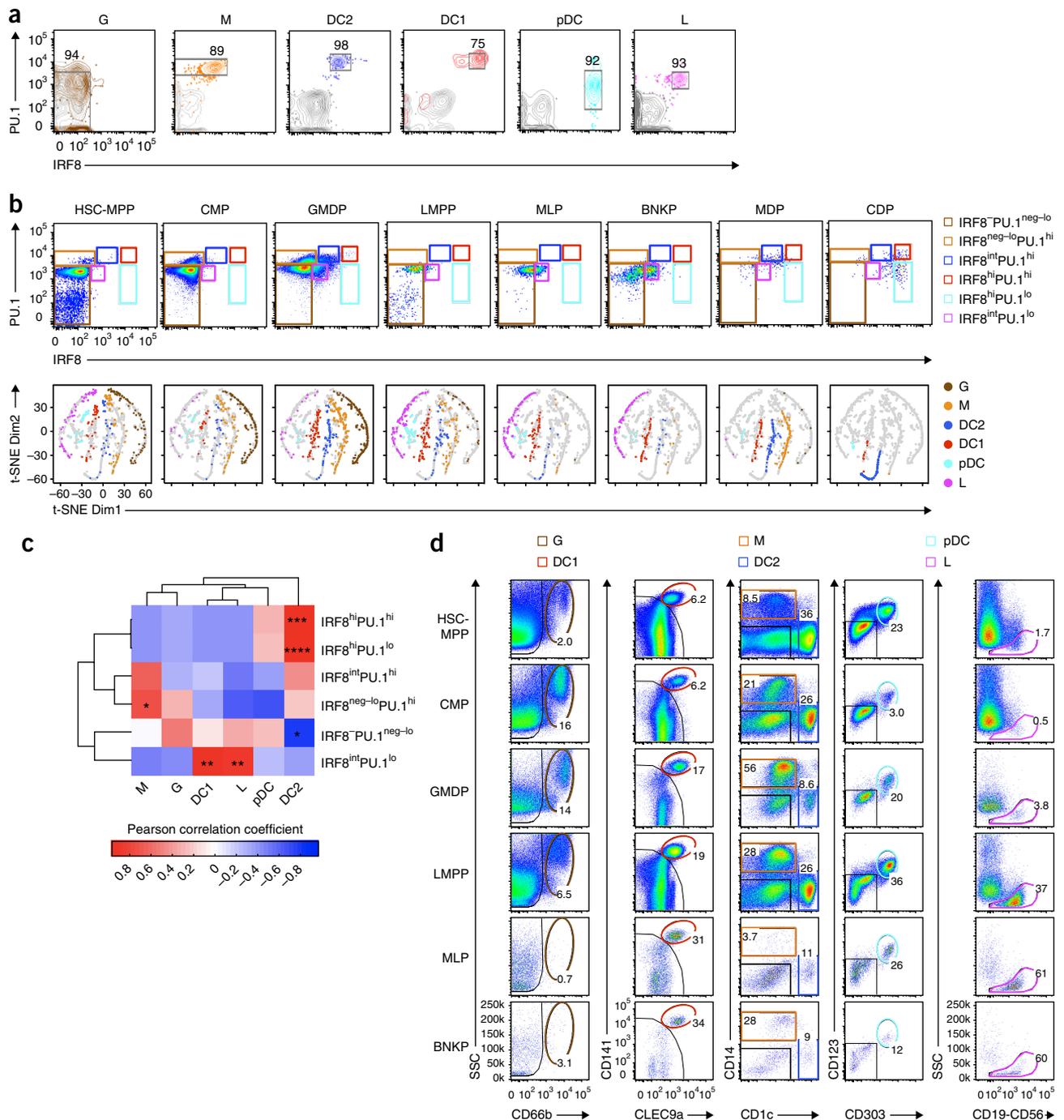
We next investigated the transcriptional program associated with lineage bias in HSCs. The transcription factors IRF8 and PU.1 are important for the development of multiple blood lineages, including DC subsets<sup>25–27</sup>. Because in mice PU.1 controls expression of the gene encoding IRF8 (ref. 17), which prevents the development of neutrophils from MDPs and common monocytic progenitors<sup>16,17,28</sup> and regulates the survival and function of terminally differentiated cells of the DC1 and pDC lineages<sup>9</sup>, we assessed the protein-expression kinetics of IRF8 and PU.1 during human DC hematopoiesis, by intracellular staining. We found distinct concentrations of IRF8 and PU.1 and ratios of IRF8 to PU.1 (called the 'IRF8-PU.1 dose' here) in differentiated cells of the G lineage (IRF8<sup>int</sup>PU.1<sup>lo</sup>), M lineage (IRF8<sup>neg-lo</sup>PU.1<sup>hi</sup>), L lineage (IRF8<sup>int</sup>PU.1<sup>lo</sup>), DC1 lineage (IRF8<sup>hi</sup>PU.1<sup>hi</sup>), DC2 lineage (IRF8<sup>int</sup>PU.1<sup>hi</sup>) and pDC lineage (IRF8<sup>hi</sup>PU.1<sup>lo</sup>) (**Fig. 6a**). The abundance of IRF8 protein was greater in the pDC and DC1 lineages than in the DC2 or other lineages (**Fig. 6a**). IRF8 and PU.1 were detectable as early as the HSC and MPP stages, albeit in a small number of cells (**Fig. 6b**), while LMPPs, GMDPs, MLPs, BNKPs, CMPs, MDPs and CDPs could be categorized into sub-populations with distinct dose combinations of IRF8 and PU.1 (**Fig. 6b**), reminiscent of those seen in mature cells of the L, pDC, DC1, DC2, M and G lineages (**Fig. 6a**). The IRF8<sup>int</sup>PU.1<sup>lo</sup> subpopulation was prominent among LMPPs, MLPs and BNKPs, whereas the IRF8<sup>int</sup>PU.1<sup>hi</sup> subpopulation was abundant among GMDP and MDPs (**Fig. 6b**). To determine the correlation between the frequency of subpopulations identified by IRF8-PU.1 dose and frequency of clones biased to the L, G, M, DC1, DC2 and pDC lineages among various progenitor populations measured by clonal assay on MP plus FSG (**Fig. 5e**), we calculated Pearson correlation coefficients for the frequency of all populations assessed. There was a positive correlation between the IRF8<sup>int</sup>PU.1<sup>lo</sup> subpopulation and DC1 lineage ( $r = 0.91$ ) and between the IRF8<sup>int</sup>PU.1<sup>hi</sup> subpopulation and the DC2 lineage ( $r = 0.46$ ) and M lineage ( $r = 0.64$ ) (**Fig. 6c**), which suggested a propensity of IRF8<sup>int</sup>PU.1<sup>lo</sup> cells to produce the DC1 lineage and of IRF8<sup>int</sup>PU.1<sup>hi</sup> cells to produce the DC2 and M lineages. To assess the relevance of the expression of IRF8 and PU.1 in terms of DC-subset potency *in vivo*, we purified HSC-MPPs, MLPs, BNKPs, LMPPs, GMDPs and CMPs from cord blood and injected them intratibially into NSG-SGM3 mice (NSG mice that express human IL-3, GM-CSF and SCF). At 2 weeks after the cell transfer, CMPs and GMDPs, which were predominantly IRF8<sup>int</sup>PU.1<sup>hi</sup>, produced abundant cells of the G, M and DC2 lineages but fewer cells of the DC1 and pDC lineages (**Fig. 6d**). In contrast, LMPPs, MLPs and BNKPs, which were predominantly IRF8<sup>int</sup>PU.1<sup>lo</sup>, produced abundant cells of the L, DC1 and pDC lineages but few cells of the DC2 and M lineages (**Fig. 6d**). This indicated that the progenitor cells' IRF8-PU.1 dose correlated with certain biases toward distinct DC subsets.

To determine whether expression of *Irf8* might mark specification to the DC lineage at the HSC-MPP stage in mice, we used *Irf8*<sup>gfp/gfp</sup> mice, which express enhanced green fluorescent protein (GFP) fused to the carboxyl terminus of endogenous IRF8 (ref. 29). About 34% of Lin<sup>-</sup>Sca-1<sup>+</sup>Kit<sup>+</sup> (LSK) cells, which include HSCs, MPPs and LMPPs, from *Irf8*<sup>gfp/gfp</sup> mice had intermediate expression of GFP (**Fig. 7a**), which correlated with intracellular staining of IRF8 with antibody (data not shown). The same number of GFP<sup>+</sup> LSK cells and GFP<sup>-</sup> LSK cells from *Irf8*<sup>gfp/gfp</sup> mice were seeded in a culture containing the

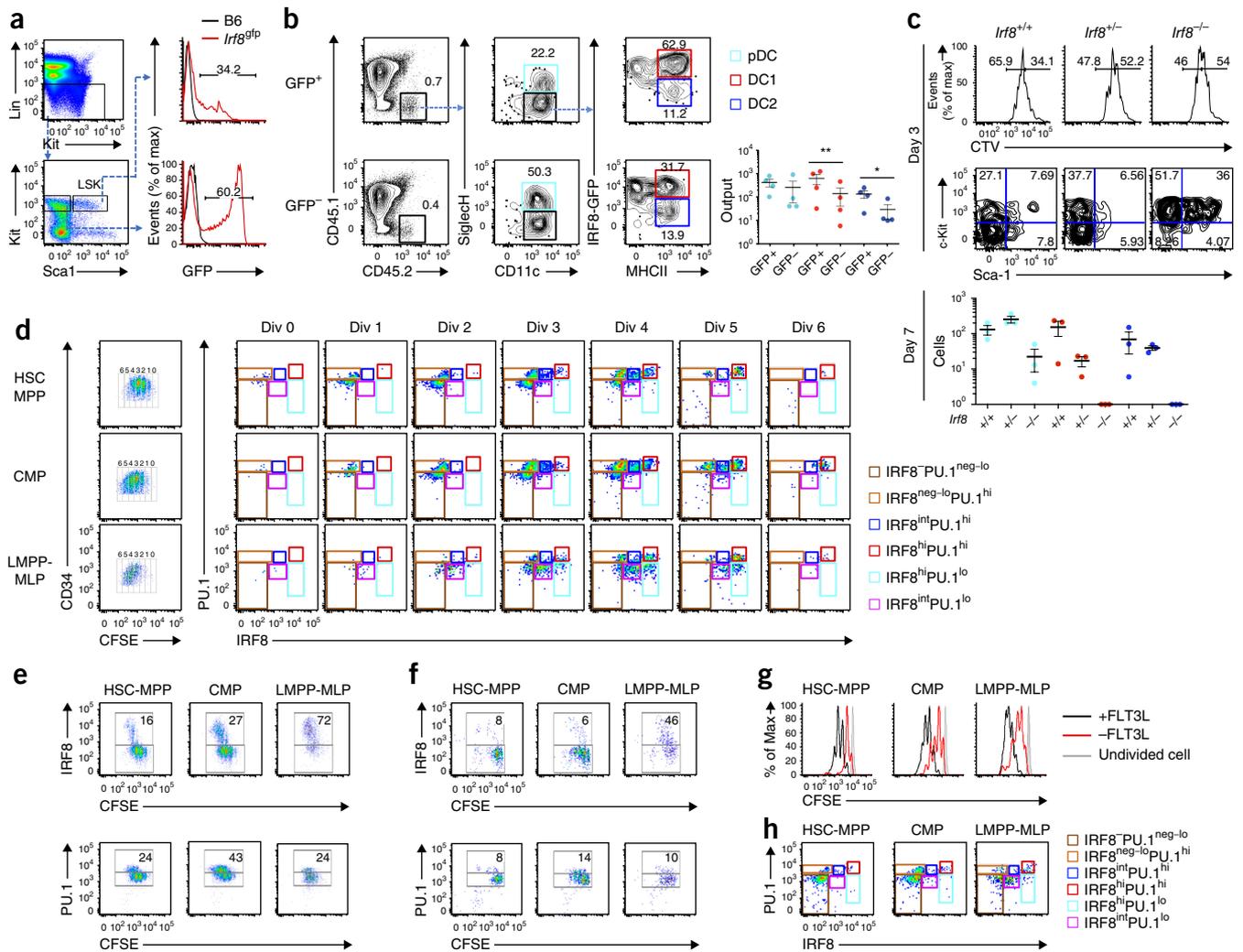
cytokine FLT3L, which supports differentiation of the DC1, DC2 and pDC lineages<sup>30</sup>. GFP<sup>+</sup> LSK cells produced fourfold more cells of the DC1 and DC2 lineages than did GFP<sup>-</sup> LSK cells from the same mice, although their pDC output was similar (Fig. 7b), which indicated that

IRF8 expression in LSK cells can be used to distinguish subpopulations with distinct DC subset potency.

To analyze the 'dose effect' of *Irf8*, we used mice carrying various number of *Irf8*<sup>-</sup> alleles, originally generated by crossing of



**Figure 6** Distinct and inheritable pattern of expression of IRF8 and PU.1 in progenitor cells correlates with lineage bias. **(a)** Flow cytometry of six types of mature immune cells (above plots) showing gating of each (outlines) on the basis of the expression of IRF8 and PU.1. Numbers adjacent to outlined areas indicate percent cells of each type (above plot). **(b)** Flow cytometry (as in **a**) of nine types of progenitor cells (above plots) (top row), showing cells with various pattern of expression of IRF8 and PU.1 (colors of outlined areas match key), and t-SNE (bottom row), showing the distribution of clones derived from progenitor cells (colors indicate lineage bias (key)). **(c)** Pearson correlation coefficient for the frequency of subpopulations identified by IRF8-PU.1 dose (right margin) and frequency of lineage composition of progenitor cells defined as in **b**; bracketing (top and left margin), hierarchical clustering of correlation matrices. **(d)** Flow cytometry of HSC-MPP, CMP, GMDP, LMPP, MLP and BNKP populations 14 d after intratibial injection into NSG-SGM3 mice, showing *in vivo* potency. Numbers adjacent to outlined areas indicate percent cells in each lineage (key) among parental cells. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  and \*\*\*\* $P < 0.0001$  (paired two-tail Student's *t*-test for transformed correlation). Data are representative of three independent experiments.



**Figure 7** Early expression of IRF8 in HSCs facilitates the specification of DC1 lineage. **(a)** Flow cytometry of bone marrow cells from B6 and *Irf8<sup>GFP</sup>* mice (key), showing gating scheme for LSK cells (left: *Irf8<sup>GFP</sup>*). Numbers above bracketed lines (right) indicate percent GFP<sup>+</sup> (IRF8-expressing) cells. **(b)** Flow cytometry (left and middle) of GFP<sup>+</sup> and GFP<sup>-</sup> LSK cell-derived cells from *Irf8<sup>GFP/GFP</sup>* mice after 7 d of co-culture with CD45.1<sup>+</sup> bone marrow cells plus FLT3L, showing production of differentiated DCs (key), and yield of differentiated DCs (far right). Numbers adjacent to outlined areas (left and middle) indicate percent DCs in each lineage (colors match key). **(c)** Flow cytometry of LSK cell-derived cells from *Irf8<sup>+/+</sup>*, *Irf8<sup>+/-</sup>* and *Irf8<sup>-/-</sup>* mice (above plots) after 3 d of culture as in **b**, showing proliferation (dilution of CellTrace Violet (CTV) and Sca-1 and c-Kit phenotype (middle row), and yield of differentiated DCs from such cells (genotype, horizontal axis) after 7 d of culture (bottom row). Numbers above lines (top row) indicate percent cells that had proliferated more slowly (right) or faster (left); numbers in quadrants (middle row) indicate percent cells in each. **(d)** Flow cytometry of cord-blood-derived HSC-MPPs, CMPs and LMPP-MLPs (left margin) after 6 d of culture on MP plus FSG, showing CFSE dilution (far left), and expression of IRF8 and PU.1 (middle and right) in cells at each division (above plots). Numbers above outlined areas (far left) indicate rounds of divisions; colors of outlined areas (middle and right) match key. **(e,f)** Flow cytometry of cells (above plots) after 6 d of culture on MP plus FSG **(e)** or MP plus SG **(f)**, showing expression of IRF8 (top row) and PU.1 (bottom row) at various stages of division (CFSE dilution). Numbers in outlined areas indicate percent IRF8<sup>+</sup> cells (top row) or PU.1<sup>+</sup> cells (bottom row). Numbers in outlined areas indicate percent IRF8<sup>+</sup> cells (top row) and PU.1<sup>+</sup> cells (bottom row). **(g)** Division of progenitor cells (above plots) after culture in MP plus FSG (+FLT3L) or MP plus SG (-FLT3L). **(h)** Flow cytometry of progenitor cells (above plots as in **g**) after 6 d of culture in MP plus SG, showing expression of IRF8 and PU.1 (outlined areas match key). Each symbol (**b**, bottom right; **c**, bottom row) represents an individual experiment; small horizontal lines indicate the mean ( $\pm$  s.e.m.). Data are representative of four independent experiments (**a**), four experiments with one to two mice per group (**b**), three experiments with one to two mice per group (**c**) or three independent experiments (**d-h**).

C57BL/6 mice with *loxP*-flanked *Irf8* alleles (B6(Cg)-*Irf8<sup>tm1.1Hm</sup>/J*) mice with mice expressing Cre recombinase from the gene encoding the transcription factor SOX2. We isolated LSK cells from *Irf8<sup>+/+</sup>*, *Irf8<sup>+/-</sup>* and *Irf8<sup>-/-</sup>* mice, labeled them with a division-tracking dye and assessed their proliferation and DC development in the FLT3L culture described above (**Fig. 7b**). On day 3 of culture, *Irf8<sup>+/-</sup>* and *Irf8<sup>-/-</sup>* LSK cells showed less proliferation than that of *Irf8<sup>+/+</sup>* LSK cells (**Fig. 7c**). Moreover, *Irf8<sup>-/-</sup>* LSK cells maintained higher expression of Sca-1 and Kit than that of *Irf8<sup>+/-</sup>* or *Irf8<sup>-/-</sup>* LSK cells (**Fig. 7c**), which

indicated that IRF8 deficiency impaired the differentiation of LSK cells. On day 7, *Irf8<sup>+/-</sup>* LSK cells produced ninefold fewer cells of the DC1 lineage and twofold fewer cells of the DC2 lineage than did *Irf8<sup>+/+</sup>* LSK cells, while *Irf8<sup>-/-</sup>* LSK cells failed to produce any cells of the DC1 or DC2 lineage (**Fig. 7c**). pDCs did not develop from *Irf8<sup>-/-</sup>* LSK cells but developed normally from *Irf8<sup>+/-</sup>* LSK cells (**Fig. 7c**). These data indicated that IRF8 functionally regulated the proliferation and specification of DC-subset lineages in a dose-dependent manner at around the HSC stage.

To trace the development of human IRF8<sup>int</sup>PU.1<sup>lo</sup> DC progenitor cells into IRF8<sup>hi</sup>PU.1<sup>hi</sup> DC1 cells, we purified HSCs, CMPs and LMPPs from cord blood, labeled them with CFSE and assessed the change in the expression of PU.1 and IRF8 over several cell divisions in MP plus FSG culture. Few LMPP progeny were IRF8<sup>int</sup>PU.1<sup>hi</sup> throughout all divisions, while the IRF8<sup>int</sup>PU.1<sup>lo</sup> LMPP progeny expanded and peaked at divisions 3–4, followed by an increase in the number of IRF8<sup>hi</sup>PU.1<sup>hi</sup> cells at divisions 4–5 (Fig. 7d); this suggested that the initial IRF8<sup>int</sup>PU.1<sup>lo</sup> expression profile of LMPPs was transmitted to most progeny and was further reinforced during cell division to establish a bias toward commitment toward the DC1 lineage. The expression of both IRF8 and PU.1 increased over the course of LMPP division but while IRF8 expression increased rapidly over the course of cell division, PU.1 expression remained relatively low and increased at a considerably slower rate (Fig. 7d). This suggested that IRF8<sup>int</sup>PU.1<sup>lo</sup> cells rapidly increased IRF8 expression over cell division and gave rise to IRF8<sup>hi</sup>PU.1<sup>hi</sup> cells.

Because cell division is driven by extrinsic cytokines, we investigated the role of extrinsic cytokines in strengthening lineage identity by assessing the effect of withdrawing FLT3L, the key cytokine that regulates DC development *in vivo*. We cultured CFSE-labeled HSCs, CMPs and LMPP-MLPs in a culture containing MS5 and OP9 stromal cells and the cytokines SCF and GM-CSF without FLT3L (called 'MP plus SG' here). HSC-MPPs, CMPs and, to a lesser degree, LMPPs underwent less division on MP plus SG than on MP plus FSG (Fig. 7e–g), and few of the cells that underwent division upregulated IRF8 expression (Fig. 7f); this resulted in significantly less generation of IRF8<sup>hi</sup>PU.1<sup>hi</sup> cells (Fig. 7h), which were associated with development of the DC1 lineage. This indicated that FLT3L not only facilitated the division of early progenitor cells but also drove the expression, maintenance and upregulation of IRF8. Together these data indicated that the IRF8-PU.1 dose correlated with the lineage bias established in HSCs (Supplementary Fig. 6a), that IRF8 expression started as early as in HSCs, where it regulated the propagation of LSK cells and their development into DCs, and that the maintenance and reinforcement of IRF8 expression over the course of cell division was dependent on FLT3L (Supplementary Fig. 6b).

## DISCUSSION

Here we have shown that specification to the human DC lineage occurred in parallel with that of the myeloid and lymphoid lineages in HSCs and was defined by specific transcriptional programs correlated with the ratio of IRF8 to PU.1. IRF8 expression in HSC-MPPs facilitated the propagation of DC progenitors and was driven by FLT3L during cell division.

Published single-cell studies have suggested early specification of the myeloid and DC lineages in mouse<sup>11,14,15</sup> and divergence of the erythro-megakaryocytic lineage from HSC-MPPs in humans<sup>22</sup>. We found that for HSC-MPPs, lineage specification began as a bias that was heritable and was transmitted to most progeny, where it was further amplified and reinforced toward commitment during cell division. Consistent with that, the proportion of Er- and Mk-biased HSC-MPPs in JD culture (i.e., supportive of the Er and Mk lineages) corresponded with the proportion of unproductive HSC-MPPs in MP plus FSG culture (i.e., not supportive of the Er and Mk lineages). Our granddaughter-tracing experiment suggested that ancestor cells were able to generate progeny that switched lineage biases, which would explain previous interpretations of this as a series of 'binary choice events' in multipotent progenitor cells. However, most progeny inherited the ancestral lineage bias, whereas bias switching happened infrequently, and these progeny tended to be less proliferative. Thus,

we estimate that the majority of mature blood cells are produced from lineage-specified, long-term progenitor cells that proliferate and transmit their lineage bias to their progeny, while bias switching contributes minimally to the overall production of mature blood cells. This is consistent with the finding that most mature blood myeloid cells descend from myeloid-restricted HSCs<sup>15</sup>.

Progenitor subsets contained clones with various dose combinations of IRF8 and PU.1, which correlated with the clonal lineage biases. That was consistent with the reported dose-dependent roles of IRF8 and PU.1 in regulating the development of DCs, monocytes and B cells<sup>26,31–33</sup> and could explain the heterogeneity of progenitor subsets reported in many studies<sup>8,11–13</sup>.

We observed IRF8 expression in HSCs with low PU.1 expression, and that IRF8 expression rapidly increased in HSC, CMP and LMPP progeny, consistent with the idea that *Irf8* transcription depends on PU.1 (ref. 17) and auto-activation<sup>9</sup>. Due to IRF8's low affinity for interferon-response elements, it must be recruited to DNA through interactions with PU.1 or the transcription factor BATF (AP-1)<sup>34,35</sup>. In mouse MDPs, PU.1 binds a distal enhancer of *Irf8* to drive its transcription<sup>17</sup>. Later, in precursors of conventional DCs, IRF8 binds the *Irf8* enhancer to reinforce its own transcription and thereby reinforces commitment to the CD8<sup>+</sup> DC1 lineage<sup>9</sup>. The transcription factor E2-2 employs similar autoactivation to reinforce the pDC lineage program<sup>36</sup>. IRF8 expression increased sharply within human LMPP progeny despite relatively low expression of PU.1 protein. Given that different *Irf8* enhancers are activated in the mouse MDP, DC1 and pDC lineages<sup>9,17</sup>, an alternative enhancer might facilitate *IRF8* transcription in human LMPPs.

FLT3L drove both the division of early progenitor cells and the upregulation of IRF8 expression throughout cell division, consistent with a requirement for FLT3L in mouse DC development<sup>37</sup>. Lineage bias in HSCs was transmitted and further amplified during cell division, and that cell division was coupled with the sequential acquisition of progenitor phenotypes, as defined by the expression of cell-surface receptors, including CD38, CD45RA, FLT3 (CD135), CD115, CD10 and CD123 (refs. 7,18). Although a receptor-expression phenotype is not equivalent to and does not synchronize with the transcriptional program, both can be linked with extrinsic signals and cell division. We speculate that combinatorial dose of a common set of transcription factors in HSCs and MPPs can shape parallel and inheritable programs for distinct hematopoietic lineages, which are then reinforced throughout cell division by recursive interactions between transcriptional programs and extrinsic signals.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank S. Goff, S. Reiner and R. Belsky for discussions. Supported by the Empire State Stem Cell Fund through the New York State Department of Health (C029562 to K.L.) and The US National Institute of Health (AI101251 to K.L.). Research reported in this manuscript was performed partly by the Columbia Center for Translational Immunology Flow Cytometry Core, supported in part by the Office of the Director of the US National Institutes of Health (S10RR027050 and S10OD020056).

## AUTHOR CONTRIBUTIONS

J.L., Y.J.Z. and K. Liu designed the study; J.L., Y.J.Z., A.A., K. Lucero and D.L. performed the experiments; J.L., Y.J.Z., D.L. and K. Liu performed data analysis;

W.M., W.Z., T.L. and Y.S. did the statistical computation and modeling; M.T. and G.B. provided human samples; and J.L., Y.J.Z., W.M., Y.S. and K. Liu wrote the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Manz, M.G., Traver, D., Miyamoto, T., Weissman, I.L. & Akashi, K. Dendritic cell potentials of early lymphoid and myeloid progenitors. *Blood* **97**, 3333–3341 (2001).
- D'Amico, A. & Wu, L. The early progenitors of mouse dendritic cells and plasmacytoid predendritic cells are within the bone marrow hemopoietic precursors expressing Flt3. *J. Exp. Med.* **198**, 293–303 (2003).
- Karsunky, H., Merad, M., Cozzio, A., Weissman, I.L. & Manz, M.G. Flt3 ligand regulates dendritic cell development from Flt3<sup>+</sup> lymphoid and myeloid-committed progenitors to Flt3<sup>+</sup> dendritic cells in vivo. *J. Exp. Med.* **198**, 305–313 (2003).
- Geissmann, F. *et al.* Development of monocytes, macrophages, and dendritic cells. *Science* **327**, 656–661 (2010).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I.L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
- Adolfsson, J. *et al.* Identification of Flt3<sup>+</sup> lympho-myeloid stem cells lacking erythromegakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
- Lee, J. *et al.* Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. *J. Exp. Med.* **212**, 385–399 (2015).
- Schlitzer, A. *et al.* Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat. Immunol.* **16**, 718–728 (2015).
- Grajales-Reyes, G.E. *et al.* Batf3 maintains autoactivation of *Irf8* for commitment of a CD8 $\alpha$ <sup>+</sup> conventional DC clonogenic progenitor. *Nat. Immunol.* **16**, 708–717 (2015).
- Mercier, F.E. & Scadden, D.T. Not all created equal: lineage hard-wiring in the production of blood. *Cell* **163**, 1568–1570 (2015).
- Naik, S.H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
- Perié, L., Duffy, K.R., Kok, L., de Boer, R.J. & Schumacher, T.N. The branching point in erythro-myeloid differentiation. *Cell* **163**, 1655–1662 (2015).
- Paul, F. *et al.* Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).
- Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112–1126 (2013).
- Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
- Becker, A.M. *et al.* IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors. *Blood* **119**, 2003–2012 (2012).
- Schönheit, J. *et al.* PU.1 level-directed chromatin structure remodeling at the *Irf8* gene drives dendritic cell commitment. *Cell Rep.* **3**, 1617–1628 (2013).
- Manz, M.G., Miyamoto, T., Akashi, K. & Weissman, I.L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc. Natl. Acad. Sci. USA* **99**, 11872–11877 (2002).
- Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).
- Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152 (2011).
- Breton, G. *et al.* Circulating precursors of human CD1c<sup>+</sup> and CD141<sup>+</sup> dendritic cells. *J. Exp. Med.* **212**, 401–413 (2015).
- Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2015).
- Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).
- Van Der Maaten, L. Accelerating t-sne using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
- Aliberti, J. *et al.* Essential role for ICSBP in the in vivo development of murine CD8 $\alpha$ <sup>+</sup> dendritic cells. *Blood* **101**, 305–310 (2003).
- Carotta, S. *et al.* The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. *Immunity* **32**, 628–641 (2010).
- Hambleton, S. *et al.* IRF8 mutations and human dendritic-cell immunodeficiency. *N. Engl. J. Med.* **365**, 127–138 (2011).
- Kurotaki, D. *et al.* IRF8 inhibits C/EBP $\alpha$  activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils. *Nat. Commun.* **5**, 4978 (2014).
- Wang, H. *et al.* A reporter mouse reveals lineage-specific and heterogeneous expression of IRF8 during lymphoid and myeloid cell differentiation. *J. Immunol.* **193**, 1766–1777 (2014).
- Naik, S.H. *et al.* Cutting edge: generation of splenic CD8<sup>+</sup> and CD8<sup>-</sup> dendritic cell equivalents in Fms-like tyrosine kinase 3 ligand bone marrow cultures. *J. Immunol.* **174**, 6592–6597 (2005).
- Carotta, S. *et al.* The transcription factors IRF8 and PU.1 negatively regulate plasma cell differentiation. *J. Exp. Med.* **211**, 2169–2181 (2014).
- Belz, G.T. & Nutt, S.L. Transcriptional programming of the dendritic cell network. *Nat. Rev. Immunol.* **12**, 101–113 (2012).
- Collin, M., Bigley, V., Haniffa, M. & Hambleton, S. Human dendritic cell deficiency: the missing ID? *Nat. Rev. Immunol.* **11**, 575–583 (2011).
- Tussiwand, R. *et al.* Compensatory dendritic cell development mediated by BATF-IRF interactions. *Nature* **490**, 502–507 (2012).
- Glasmacher, E. *et al.* A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* **338**, 975–980 (2012).
- Grajkowska, L.T. *et al.* Isoform-specific expression and feedback regulation of E protein TCF4 control dendritic cell lineage specification. *Immunity* **46**, 65–77 (2017).
- Schmid, M.A., Kingston, D., Boddupalli, S. & Manz, M.G. Instructive cytokine signals in dendritic cell lineage commitment. *Immunol. Rev.* **234**, 32–44 (2010).

## ONLINE METHODS

**Human samples.** Human umbilical cord blood was purchased from New York Blood Center (New York) and was processed 24–48 h after delivery. Human bone marrow was obtained from the Hematopathology Division or the Columbia Center for Translational Immunology at Columbia University Medical Center (New York). Informed consent was obtained from the patients, and/or samples were exempt from informed consent, being residual material after diagnosis and fully de-identified. All samples were collected according to protocols approved by the Institutional Review Board at Columbia University Medical Center.

**Mice.** NOD.Cg-Prkdcscid-IL2rg<sup>tm1Wjl</sup>/Sz (NOD/SCID/IL2rg<sup>null</sup> or NSG) mice and NOD.Cg-Prkdcscid-IL2rg<sup>tm1Wjl</sup> Tg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ (NSG-SGM3), C57BL/6J, CD45.1, *Irf8*<sup>-/-</sup> (stock number 018298) mice and IRF8<sup>flp</sup> reporter mice (stock number 027084) were purchased from Jackson Laboratory and were bred in a pathogen-free animal facility at CUMC. *Irf8*<sup>+/-</sup> mice were obtained by crossing of *Irf8*<sup>-/-</sup> mice to wild-type C57BL/6J mice. All experiments were performed according to the guidelines of IACUC at CUMC. For experiments, both sex of mice between 4 weeks and 8 weeks were used.

**Cell isolation and flow cytometry.** Fresh mononuclear cells were isolated from cord blood or bone marrow by density centrifugation using Ficoll-Hypaque (Amersham Pharmacia Biotech). Samples were incubated with fluorescence-labeled antibodies for direct analysis on BD LSR II flow cytometers (Becton Dickinson Immunocytometry Systems (BDIS)) or for further purification by fluorescence-activated cell sorting on a BD Influx or BD FACS Aria, both using HeNe and argon lasers. Sorted population showed >95% purity.

For human-hematopoietic-progenitor-cell analysis, single-cell lineage potential, developmental-hierarchy-relationship experiments, daughter-cell-lineage potential, and characterization of progenitor cells, CD34<sup>+</sup> cells were first enriched from cord blood using CD34 MicroBead Kit and LS MACS magnetic columns (Miltenyi Biotec). Enriched CD34<sup>+</sup> cells (70–95% purity) were incubated with anti-CD3 (OKT3, Brilliant Violet (BV) 650, BioLegend), anti-CD19 (HIB19, BV650, BioLegend), anti-CD56 (HCD56, BV650, BioLegend), anti-CD14 (TuK4, Qdot-655, Invitrogen), anti-CD66b (G10F5, PerCP-Cy5.5, BioLegend), anti-CD303 (201A, PerCP-Cy5.5, BioLegend), anti-CD141 (M80, PerCP-Cy5.5, BioLegend), anti-CD1c (L161, APC-Cy7, BioLegend), anti-CD34 (581, Alexa Fluor (AF) 700, BioLegend), anti-CD38 (HIT2, BV421, BioLegend), anti-CD90 (5E10, PE, BioLegend), anti-CD45RA (HI100, AF488, BioLegend), anti-CD123 (6H6, BV510, BD), anti-CD10 (HI10a, PE-Cy7, BioLegend) and anti-CD115 (9-4D2-1E4, APC, BioLegend). For culture experiments, progenitor cells were sorted from Lin<sup>-</sup> (CD3<sup>-</sup>CD19<sup>-</sup>CD56<sup>-</sup>CD14<sup>-</sup>CD66b<sup>-</sup>CD303<sup>-</sup>CD141<sup>-</sup>CD1c<sup>-</sup>) cells and according to the surface phenotypes in **Table 1**.

For inter-developmental relationship experiments, cells from either culture or NSG bone marrow were stained for LIVE/DEAD (Life Technologies), CD45 (HI30, AF700, BioLegend), CD14 (Qdot-655), CD3 (OKT3, BV650, BioLegend), CD19 (HIB19, BV650, BioLegend), CD56 (HCD56, BV650, BioLegend), CD16 (3G8, BV650, BioLegend), CD11c (3.9, PerCP-Cy5.5, BioLegend), CD66b (PerCP-Cy5.5), CD303 (PerCP-Cy5.5), CD141 (PerCP-Cy5.5), CD34 (581, APC-Cy7, BioLegend), CD38 (BV421), CD90 (PE), CD7 (CD7-6B7, PE-Cy7, BioLegend), CD45RA (AF488), CD123 (BV510) and CD115 (APC). For *in vivo* transfer experiments, mouse CD45 (30-F11, PB, BD) was also stained.

For characterization of terminally differentiated cells in single cell cultures or NSG bone marrow, cells were stained for LIVE/DEAD (Life Technologies), CD45 (AF700), CD66b (PerCP-Cy5.5), CD56 (B159, Pacific Blue (PB), BD), CD19 (HIB19, PB, BioLegend), CD14 (Qdot-655), CLEC9a (8F9, PE, BioLegend), CD1c (L161, PE-Cy7, BioLegend), CD303 (201A, FITC, BioLegend), CD123 (6H6, Brilliant Violet (BV) 510, BioLegend), CD141 (AD5-14H12, APC, Miltenyi), CD235a (GA-R2, APC, BD Pharmingen) and CD41a (HIP8, APC-H7, BD Pharmingen) for 40 min on ice. 4  $\mu$ l or 10  $\mu$ l of antibody mix was used to stain cells harvested from 96-well plates or 24-well plates, respectively. For *in vivo* transfer experiments, mouse CD45 (30-F11, PB, BD) was also stained.

For intracellular staining of PU.1 and IRF8, cells were first stained with antibodies to surface markers, then were fixed and permeabilized using the Foxp3 Fixation/Permeabilization Concentrate and Diluent Kit (eBioscience) for 20 min

on ice, and then were stained with anti-IRF8 (V3GYWCH, PE, eBioscience) and anti-PU.1 (7C6B05, AF647, BioLegend) in 1 $\times$  Permeabilization buffer (eBioscience) for more than 1 h on ice.

Differentiated DCs from mouse bone marrow progenitor cells were identified by staining of CD45.2 (104, Pacific Blue), CD45.1 (A20, PerCP-Cy5.5), CD11c (N418, APC-Cy7), I-Ab (M5/114.15.2, A700), SiglecH (551, PE) and CD172a (P84, APC) (all antibodies from BioLegend).

**Cell culture.** Two culture system were used for clonal assay of cord-blood-derived progenitor cells. For MP plus FSG culture, MS5 and OP9 stromal cells were maintained and passed in complete alpha MEM medium (Invitrogen) with 10% FCS and penicillin-streptomycin (Invitrogen) as previously described<sup>7</sup>. In brief, after 2 h of treatment with 10  $\mu$ g/ml of mytomicin C (Sigma) and washing with PBS, MS5 and OP9 cells were seeded at a 1:6 ratio in 96- or 24-well plates 24 h before culture of hematopoietic cells. For 96-well plates, 3.75  $\times$  10<sup>4</sup> MS5 cells and 6.25  $\times$  10<sup>3</sup> OP9 cells were seeded per well, and for 24-well plates, 1.5  $\times$  10<sup>5</sup> MS5 and 2.5  $\times$  10<sup>4</sup> OP9 cells were seeded per well. Purified progenitor populations were cultured in medium containing 100 ng/ml FLT3L (Celldex), 20 ng/ml SCF (Peprotech) and/or 10 ng/ml GM-CSF (Peprotech), with half of the medium changed every 7 d. Cells were harvested between day 3 and day 21 for flow-cytometry analysis. For JD culture, we used published conditions<sup>22</sup>. In brief, MS5 cells were plated in a flat-bottomed 96-well plate at the density of 5  $\times$  10<sup>3</sup> cells per well in Myelocult medium (H5100, Stem cell technologies) per well and were given 24–48 h to attach. Before cell sorting, Myelocult media was carefully removed and 200  $\mu$ l medium was added. We used serum-free media (StemPro34 SFM with nutrient, Life Technologies) supplemented with SCF (100 ng/ml), FLT3 (20 ng/ml), TPO (100 ng/ml), EPO (3 units/ml), IL-6 (50 ng/ml), IL-3 (10 ng/ml), IL-11 (50 ng/ml), GM-CSF (20 ng/ml), LDL (4  $\mu$ g/ml), 2-mercaptoethanol, L-glutamine and penicillin-streptomycin. At week 2, half of the medium was changed. Colony-forming unit assays were performed using MethoCult (Stemcell, H4434), containing SCF, GM-CSF, IL-3 and EPO. Colony-forming unit cells (CFU-C) were counted after 14 d of culture.

For FLT3L culture of mouse progenitor cells, 200 purified LMPPs from CD45.2<sup>+</sup>*Irf8*<sup>+/+</sup>, *Irf8*<sup>+/-</sup> or *Irf8*<sup>-/-</sup> mice were seeded with 3  $\times$  10<sup>5</sup> CD45.1<sup>+</sup> total bone marrow cells in 200  $\mu$ l of RPMI culture with 10% FCS, 1 mM L-glutamine, 1 mM sodium pyruvate, 10 mM HEPES, NEAA  $\mu$ g/ml FLT3L in 96-well round-bottomed plates, and were cultured for 2–7 d before analysis.

To determine cellular divisions in culture, input populations were labeled for 15 min with 5  $\mu$ M carboxyfluorescein diacetate succinimidyl ester (CFSE, Molecular Probes) or CellTrace violet (Molecular Probes) at 37 °C and were washed with complete alpha MEM before culture or *in vivo* transfer.

**Tracing of single-cell progeny.** For daughter-cell tracing, HSCs-MPPs and GMDPs were first sorted as a population based on their surface marker phenotype described in **Table 1**. Washed cells in cold PBS were incubated in 500  $\mu$ l of alpha MEM medium (Invitrogen) without serum containing Vybrant DiD cell-labeling solution (1:200 dilution, Life Technologies) for 20 min at 37 °C in a water bath. Cells were spun down at 1,500 r.p.m. for 5 min and were washed twice with complete alpha MEM medium (Invitrogen) with 10% FCS and penicillin-streptomycin (Invitrogen). Cells were then resuspended in PBS and were re-sorted as DiD<sup>+</sup> directly into 96-well plates in MP plus FSG at a density of 1 cell per well. Each cell was monitored daily for division using either an EVOS FL Cell imaging system (Life Technologies) or an Inverted Leica fluorescent microscope DM16000 (Leica) equipped with a Cy5 light source. This method allowed us to trace up to more than five divisions (>50 daughter cells) from a single initial cell (data not shown).

When the initial cell generated four granddaughter cells, as detected by microscopy, we collected and manually aliquoted them into eight separate wells of a 96-well plate in MP plus FSG (0.5 cells per well) to increase the probability of seeding one granddaughter cell into secondary wells. GMDP-derived granddaughter cells were cultured for 2 weeks and HSC-MPP-derived granddaughter cells were cultured for 3 weeks before harvest. Ancestors that only had one viable granddaughter cell by the end of the culture were not included for analysis.

The ancestor's potency was inferred by the sum of the all granddaughters. The lineage bias was determined by lineage that exhibited highest yield.

The progeny exhibited the same lineage bias with its ancestor was considered bias inherited, and the progeny exhibited different lineage bias from its ancestor was considered bias switched.

**In vivo transplantation into NSG mice.** NSG mice were given intraperitoneal injection of busulfan (Sigma, 30  $\mu\text{g/g}$  of body weight) to ablate endogenous hematopoietic system 2 d before transfer of human CD34<sup>+</sup> cells. Human progenitor cells purified from cord blood were resuspended in 10  $\mu\text{l}$  PBS and injected intratibially into mice with a Hamilton syringe and a 27-gauge needle. 7 or 14 d after transplantation, bone marrow was harvested from recipient mice and was analyzed for human CD45<sup>+</sup> cells. NSG mice were used to characterize progenitor hierarchy and for *in vivo* CFSE-labeled HSC-MPP-transfer experiments. NSG-SGM3 mice were used to determine *in vivo* progenitor lineage potential.

**Clonal analysis of progenitor cells.** Progenitor cells were individually sorted as single cells directly into 96-well plates containing mitomycin C-treated stromal cells. Immediately afterward, medium containing cytokine mix was added. Each well was harvested after 7–21 d of culture and were stained for LIVE/DEAD, CD45, CD66b, CLEC9a, CD14, CD1c, CD303, CD123, CD141, CD19 and CD56 (antibodies identified above). Positive clones were determined by the detection of at least two events (for CDPs) or seven events (for all other progenitors) in any of the lineage-specific gates.

**Heat map, principal-component analysis and multidimensional scaling.** Clonal output data were normalized with the procedure described in DESeq<sup>39</sup>, assuming that the geometric mean of total clonal output for a single progenitor phenotype across different donors should be similar. Normalized cell counts were scaled by log base 10, were clustered by unsupervised hierarchical clustering function with `hclust {stats}` (R Statistical Software) and were visualized with `heatmap.2 {gplots}`. Complete linkage method was used for clustering, with the distance metric between progenitors defined by Euclidean distance. The ordering of leaves was optimized with the `cba` package, so that the sum of similarities between adjacent leaves could be maximized while keeping the hierarchical tree structure unchanged. Principal component analysis was performed with the function `prcomp()` in R, with the centering, scaling and `cor` options on. Ancestral similarity between each pair of cell lineages was calculated as Spearman's rho with `cor {stats}`, using their yield from 2,247 progenitors as six dimensions. The distance ( $d = 1 - \rho$ ) between each cell type was calculated. The distance matrix was reduced to two dimensions with multidimensional scaling via `cmdscale {stats}`, with `eig = True` and `k = 2`. Potency similarity between each pair of progenitors was calculated in a similar way differing by transposing the counts matrix first.

**Visualization of development trajectories using t-SNE mapping.** To identify putative developmental trajectories from HSCs to six individual blood lineages through clonal output, we used t-distributed stochastic neighbor embedding (t-SNE) technique for dimension reduction for visualization. First, we further normalized the yield of each lineage with DESeq to make sure the geometric mean of each progeny type yield was similar across all progenitors (the culturing system produced fewer pDCs than other types of progeny). Then, we took the normalized clonal output as input to the Barnes-Hut t-SNE package<sup>24</sup> with the parameters `perplexity = 20` and `theta = 0.3` for visualization (cord blood samples). t-SNE minimized the Kullback-Leibler divergence between two similarity distributions, with one measuring pairwise similarities of the input objects and the other measuring pairwise similarities of the projected

low-dimensional points in the embedding space. In our case, the similarities in the high-dimension space between pairwise progenitor cells was calculated using the joint probabilities with an isotropic Gaussian kernel over the number of their terminal outputs by symmetrizing two conditional probabilities as follows:

$$P_{j|i} = \frac{\exp\left(\frac{-(x_i - x_j)^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-(x_i - x_k)^2}{2\sigma_i^2}\right)}, P_{i|i} = 0$$

$$P(i, j) = \frac{P_{j|i} + P_{i|j}}{2N}$$

where  $x_i$  and  $x_j$  are the logarithm of terminal cells number vectors for progenitors  $i$  and  $j$ ,  $\sigma_i$ , the bandwidth of the Gaussian, as cells are moving toward more differentiated state in heterogeneous and stochastic way similar to diffusion as input similarities to t-SNE for visualization in two-dimensional space and to generate the diffusion map.

**Distance computation of progenitors to track and assignment of cell-type-specific lineage bias.** To determine the distance of each cell to every lineage in the diffusion map, we first established a backbone for each lineage using cells with 70% commitment degree to that lineage. Commitment degree was defined as the ratio of one lineage yield over the sum of all six lineages yield, ranging from 0 to 1 (where 0 means no potential and 1 means fully committed). We then computed the Euclidean distance between every pair of cells. The distance from a cell to a track is defined as the closest distance to any of the cells on the backbone for all tracks. We finally assigned, as the closest track, the track to which the cell was closest.

**Calculation of correlation between transcription-factor dose and lineage potency.** For correlation between IRF8-PU.1 dose and lineage potency of all progenitor cells, we first calculated the percentage of subpopulations identified by relative IRF8-PU.1 dose and lineage-bias composition of each progenitor cell, then calculated the Pearson correlation coefficient between them for all progenitor cells. Student's *t*-test for transformed correlation<sup>38</sup> was used to access the statistical significance of correlation.

**Statistical analysis.** Statistical tests are described in their corresponding figure legends. All values indicated are mean + s.e.m., or mean and standard error of proportion, unless specified otherwise. For comparison of results, we used one-way analysis of variance, unpaired or paired two-tailed Student's *t*-test, Spearman's correlation test, Fisher's exact test and Pearson correlation test. Statistical analysis was done with GraphPad prism v7.0, Microsoft Excel, R or R Studio. Significance was set at  $P < 0.05$ . Data exclusion criteria was only applied to determine unproductive clones.

**Data availability statement.** Clonal data (for Figs. 2–5) are in **Supplementary Table 1**. All other data that support the findings of this study are available from the corresponding authors upon request.

38. Rahman, N. *A Course in Theoretical Statistics* (Charles Griffin and Company, 1968).

39. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

---

**Erratum:** Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors

Jaeyop Lee, Yu Jerry Zhou, Wenji Ma, Wanwei Zhang, Arafat Aljoufi, Thomas Luh, Kimberly Lucero, Deguang Liang, Matthew Thomsen, Govind Bhagat, Yufeng Shen & Kang Liu

*Nat. Immunol.*; doi:10.1038/ni.3789; corrected online 5 July 2017

In the version of this article initially published online, the flow cytometric dots were missing in the middle plot of the leftmost column in Figure 2f. The error has been corrected in the print, PDF and HTML versions of this article.