ARTICLE

doi:10.1038/nature13038

Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia

Liran I. Shlush[1]*, Sasan Zandi[1]*, Amanda Mitchell[1], Weihsu Claire Chen[1], Joseph M. Brandwein[1,2,3], Vikas Gupta[1,2,3], James A. Kennedy[1], Aaron D. Schimmer[1,2,3,4], Andre C. Schuh[1,2,3], Karen W. Yee[1,2,3], Jessica L. McLeod[1], Monica Doedens[1], Jessie J. F. Medeiros[1], Rene Marke[1,5], Hyeoung Joon Kim[6], Kwon Lee[6], John D. McPherson[4,7], Thomas J. Hudson[4,7,8], The HALT Pan-Leukemia Gene Panel Consortium†, Andrew M. K. Brown[7], Fouad Yousif[7], Quang M. Trinh[7], Lincoln D. Stein[7,8], Mark D. Minden[1,2,3,4], Jean C. Y. Wang[1,2,3] & John E. Dick[1,8]

In acute myeloid leukaemia (AML), the cell of origin, nature and biological consequences of initiating lesions, and order of subsequent mutations remain poorly understood, as AML is typically diagnosed without observation of a pre-leukaemic phase. Here, highly purified haematopoietic stem cells (HSCs), progenitor and mature cell fractions from the blood of AML patients were found to contain recurrent *DNMT3A* mutations (*DNMT3A*^{mut}) at high allele frequency, but without coincident *NPM1* mutations (*NPM1c*) present in AML blasts. *DNMT3A*^{mut}-bearing HSCs showed a multilineage repopulation advantage over non-mutated HSCs in xenografts, establishing their identity as pre-leukaemic HSCs. Pre-leukaemic HSCs were found in remission samples, indicating that they survive chemotherapy. Therefore *DNMT3A*^{mut} arises early in AML evolution, probably in HSCs, leading to a clonally expanded pool of pre-leukaemic HSCs from which AML evolves. Our findings provide a paradigm for the detection and treatment of pre-leukaemic clones before the acquisition of additional genetic lesions engenders greater therapeutic resistance.

There is overwhelming evidence that virtually all cancers are clonal and represent the progeny of a single cell[1–3], but the evolutionary trajectory that leads from the first somatic mutation to the eventual development of cancer is not well mapped. The simplest models predict that each newly acquired somatic mutation confers selective advantage to drive successive waves of clonal expansion, with the fittest clone becoming dominant. However, the modern era of cancer genomics has exposed a more complex clonal architecture in many tumour types[4], where multiple genetically distinct subclones co-exist with the dominant clone[5,6]. Comparison of diagnostic and recurrent/metastatic samples obtained from the same patient has established that the latter frequently do not evolve from the dominant clone, but instead can be traced either to a minor subclone present at diagnosis, or to a putative, undetected ancestral clone[7–15]. Thus, a clear understanding of the genomic landscape of tumours is required to devise targeting strategies that eliminate not only the dominant clone, but also the subclonal reservoirs from which recurrence can arise.

Although the clonal composition of cancer lineages within individual tumours is coming into focus, the very first steps in cancer development remain poorly defined. Early and possibly initiating mutations have been identified from analysis of pre-neoplastic lesions in breast[16], lung[17], skin[18] and colon cancer[19], as well as from studies of AML cases that evolved from a prior myelodysplastic syndrome (MDS)[20]. However, key questions remain unanswered. In particular, can clinically relevant clones be traced back to a non-tumorigenic cell? Do pre-cancerous ancestral clones persist after tumour development? If so, are they present in the diagnostic sample, and do they survive treatment and persist in remission samples?

Human leukaemia is a disease model particularly suited to addressing these fundamental questions, due to the depth of our understanding of normal haematopoiesis and the availability of functional assays and analytic tools that allow examination of phenotypically defined populations at the single-cell level[21]. In AML, a subset of cases evolve from a preceding clinically overt phase such as MDS or chronic myeloid leukaemia (CML), characterized by clonal expansion of one or more blood lineages[22,23]. The founder mutations present in pre-leukaemic cells are retained in the AML blasts, implicating them as putative initiating events and establishing clonal expansion as the first step in leukaemogenesis. Interestingly, somatic mutations in some leukaemia-associated genes such as *TET2* have also been linked to multilineage clonal haematopoiesis in ageing healthy individuals[24]. Insight into the phenotype of the normal cell from which clonal expansion can initiate was first provided by pioneering studies in CML, which demonstrated that *BCR-ABL1* arises in a multipotential HSC[25]. However, for the majority of AML cases that arise *de novo* without any prior clinical perturbations, insight into the cellular context and functional consequences of the earliest genetic lesions requires identification and examination of ancestral cells within the diagnostic sample. Recent studies have found that only a subset of mutations contained in AML blasts were present in HSC-enriched cell fractions isolated from AML patient samples, and that these cells were capable of non-leukaemic differentiation[26,27]. Here we establish that these ancestral pre-leukaemic HSCs present at diagnosis are able to regenerate the entire haematopoietic hierarchy while possessing competitive repopulation advantage over non-leukaemic HSCs leading to clonal expansion. These pre-leukaemic HSCs are found in a high proportion of AML patients that carry mutations in *DNMT3A* and *IDH2*, and unlike AML blasts, they survive induction chemotherapy and persist in the bone marrow at remission, providing a potential reservoir for leukaemic progression.

[1]Princess Margaret Cancer Centre, University Health Network (UHN), Toronto, Ontario M5G 2M9, Canada. [2]Department of Medicine, University of Toronto, Toronto, Ontario M5S 2J7, Canada. [3]Division of Medical Oncology and Hematology, UHN, Toronto, Ontario M5G 2M9, Canada. [4]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2M9, Canada. [5]Radboud University, Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands. [6]Chonnam National University Hwasun Hospital, Genome Research Center for Hematopoietic Diseases, Gwangju 519-809, South Korea. [7]Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. [8]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
*These authors contributed equally to this work.
†Lists of participants and their affiliations appear in Supplementary Information.

Detection of *DNMT3A*^{mut} in non-leukaemic cells

During studies to examine intra-tumoral genetic heterogeneity in AML, deep targeted sequencing (Tar-seq, read depth ~250×) of 103 commonly mutated leukaemia genes (Fig. 1a and Supplementary Table 1) was carried out on peripheral blood samples from 12 patients at diagnosis (blasts > 80%) (Supplementary Tables 2, 3). Normal T cells were expanded *in vitro* to provide non-leukaemic tissue for genetic comparison. Consistent with mutant allele frequencies reported in recent studies[28], *DNMT3A*^{mut} was found in 4 out of 12 samples (mutant allele frequency ~50%) (Fig. 1a). Unexpectedly, in 3 of these 4 patients, *DNMT3A*^{mut} was detected in T cells at a low allele frequency (1–20%). Other genetic alterations, including *NPM1c*, were found only in peripheral blood but not in T-cell samples, ruling out AML cell contamination of cultured T cells. To estimate the proportion of AML cases with *DNMT3A*^{mut}-bearing T cells, 71 additional samples, taken at diagnosis from patients with normal cytogenetics, were screened by Sanger sequencing for *DNMT3A*^{mut} along with other common AML mutations (Supplementary Table 4). Consistent with published data[29–31], 17 out of 71 AML samples (24%) carried mutations in *DNMT3A*, and 15 of these 17 (88%) also carried *NPM1c*. For these 17 patients, the allele frequency of *DNMT3A*^{mut} and *NPM1c* in CD33^{+} blasts, as well as corresponding freshly isolated T-cell controls, was measured by droplet digital PCR (ddPCR) at a sensitivity of 1 mutated allele in 1,000 reference alleles. Whereas both *DNMT3A*^{mut} and *NPM1c* were always present in blasts at similar allele frequency, *DNMT3A*^{mut} with no evidence of *NPM1c* was detected in T cells from 12 of these 17 patients (70.5%) (Fig. 1b). In addition, *FLT3* internal tandem duplications (*FLT3*-ITD) were detected in blasts, but not in the T cells of two patients bearing this mutation (Extended Data Fig. 1). These data reveal the sequential order of mutation acquisition in these patients, with *DNMT3A*^{mut} arising earlier in leukaemogenesis than *NPM1c* and *FLT3*-ITD, a conclusion predicted from recent studies on bulk AML blasts showing that *NPM1c* and *FLT3*-ITD occur late and are the only genes recurrently mutated in *DNMT3A*^{mut} AML[15,28,32]. Moreover, our findings establish that *DNMT3A*^{mut} occurs in an ancestral cell that gives rise to both T cells and the dominant AML clone present at diagnosis.

DNMT3A^{mut} HSCs undergo multilineage differentiation

To gain insight into the properties of the ancestral cell within which *DNMT3A*^{mut} first arises, we examined additional non-leukaemic haematopoietic cell populations from 11 *DNMT3A*^{mut}/*NPM1c* AML patients. A high-resolution 12-parameter sorting strategy[33–35] was used to isolate non-leukaemic haematopoietic stem and progenitor cell populations, including haematopoietic stem cells/multipotent progenitors (HSCs/MPPs), multilymphoid progenitors (MLPs), common myeloid progenitors (CMPs), granulocyte monocyte progenitors (GMPs), and megakaryocyte erythroid progenitors (MEPs), as well as mature B, T and natural killer (NK) cells within the CD33^{-} cell fraction. Together with CD45^{dim}CD33^{+} AML blasts, these highly purified, phenotypically defined normal cell populations were assessed by ddPCR for *DNMT3A*^{mut} and *NPM1c* (Fig. 2 and Extended Data Figs 2 and 3). *DNMT3A*^{mut} was found together with *NPM1c* in CD33^{+} blasts from all patients. By contrast, we found *DNMT3A*^{mut} at variable allele frequency without *NPM1c* across the spectrum of mature and progenitor cell populations. Results for a representative patient (no. 11) are shown in Fig. 2a, b. In this patient, *DNMT3A*^{mut} was present in HSCs/MPPs at an allele frequency of 12–30% without detectable *NPM1c*. Although the clonal contribution that an individual normal HSC makes in humans is unknown, studies in higher primates estimated that single HSCs provide approximately 0.5% clonal contribution during steady-state haematopoiesis[36]. Thus, the high *DNMT3A*^{mut} allele frequency in HSCs/MPPs points to their clonal expansion as compared to non-mutated HSCs. In patient no. 11, *DNMT3A*^{mut} was also present in all downstream progenitors at variable frequencies. The mean allele frequency among HSCs/MPPs, MLPs and CMPs was 24.6% across all patients analysed (Extended Data Fig. 2). Importantly, even for patients in whom *DNMT3A*^{mut} was not detected in mature cells, *DNMT3A*^{mut} without *NPM1c* was found in stem/progenitor populations (Fig. 2c),

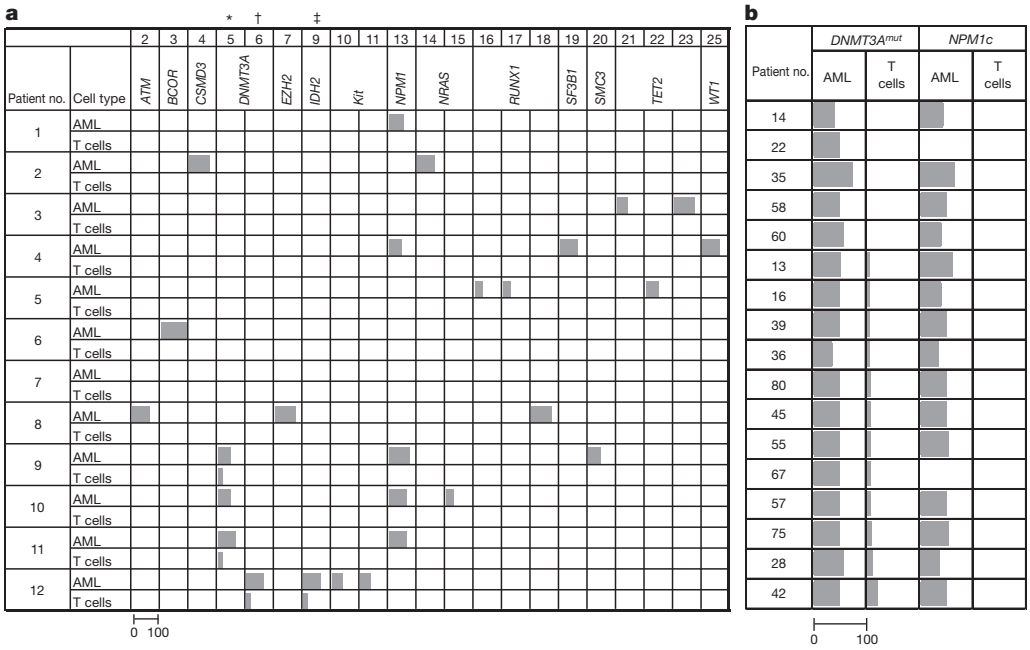


Figure 1 | Recurrent somatic *DNMT3A* mutations are common in T cells from AML patients. **a**, Summary of the allele frequency (%) of missense and frameshift somatic single nucleotide variants (sSNV) in AML-related genes assessed by deep targeted sequencing (read depth 250×) in AML blasts and T cells from the peripheral blood of 12 AML patients. The sSNV numbers indicated at the top of the table correspond to the numbers in Supplementary Table 3. Somatic mutations in *DNMT3A* (*R882H; †R137C) were found in both T cells and AML blasts in patients no. 9, 11 and 12. Patient no. 12 also had a low frequency *IDH2* mutation (‡R140L) in T cells. **b**, Frequency (%) of *DNMT3A*^{mut} and *NPM1c* in freshly isolated CD33^{+} blasts (AML) and matched T-cell controls from 17 patients with normal karyotype AML, as determined by droplet digital PCR. For **a** and **b**, the length of the bars is proportional to the mutant allele frequency (the scale bar under the first column applies to all columns).

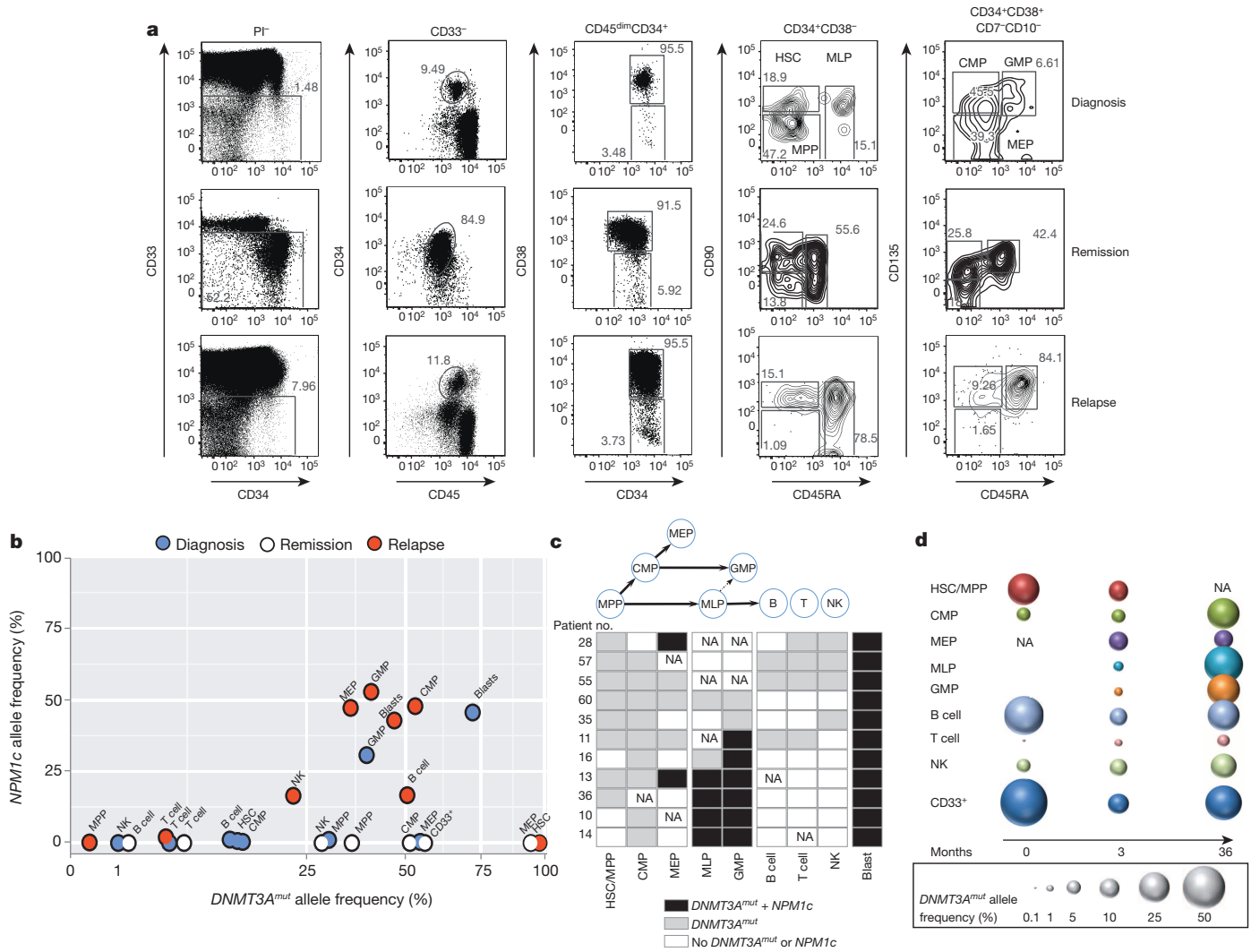


Figure 2 | *DNMT3A* mutation precedes *NPM1* mutation in human AML and is present in stem/progenitor cells at diagnosis and remission. **a**, Flow cytometric analysis showing the gating strategy used to isolate phenotypically normal stem cell and progenitor cell populations from AML patient samples. Plots show analysis of samples from patient no. 11: diagnosis (day 0, peripheral blood mononuclear cells), remission (day 62, $CD34^+$ enriched bone marrow) and relapse (day 379, peripheral blood mononuclear cells). **b**, Allele frequency of *DNMT3A* and *NPM1* mutations in stem/progenitor, mature lymphoid and blast ($CD45^{dim}$ $CD33^+$) cell populations, as indicated, isolated from diagnosis (blue), remission (white) and relapse (red) samples of patient no. 11 as determined by droplet digital PCR (ddPCR). At remission, $CD33^+$ myeloid cells were also analysed. **c**, Summary of the occurrence of $DNMT3A^{mut}$ and *NPM1c* in isolated stem/progenitor, mature and blast cell populations from 11 AML patient peripheral blood samples as determined by ddPCR. White, neither $DNMT3A^{mut}$ nor *NPM1c* detected; grey, $DNMT3A^{mut}$ alone detected; black, both $DNMT3A^{mut}$ and *NPM1c* detected. NA, no cell population detected; HSC, haematopoietic stem cell; MPP, multipotent progenitor; MLP, multilymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte monocyte progenitor; MEP, megakaryocyte erythroid progenitor; NK, natural killer cells. **d**, Graphic representation of $DNMT3A^{mut}$ allele frequency in sorted cell populations isolated from diagnosis (0 months), early (3) and late (36) remission samples of patient no. 57.

providing further strong evidence that $DNMT3A^{mut}$ precedes *NPM1c* during leukaemogenesis. Interestingly, in three patients (no. 10, 14, 16), $DNMT3A^{mut}$ was detected in CMPs but not HSCs/MPPs, an outcome consistent with the existence of $DNMT3A^{mut}$-bearing HSCs below our detection limit that generated a clonally expanded CMP population, or possibly the existence of a preceding lesion in HSCs/MPPs with later acquisition of $DNMT3A^{mut}$ in CMPs. Resolution of this question will require whole-genome sequencing of sorted blast and phenotypically normal populations from patient samples. Our analysis also showed that in 6 out of 11 patients, both $DNMT3A^{mut}$ and *NPM1c* were found together in MLP and/or GMP populations, pointing to the probable progenitor cell types where overt AML driven by *NPM1c* arises. Collectively, our findings provide key insights into the leukaemogenic process in human AML and confirm historical predictions from early clonality studies of the existence of a pre-leukaemic state[3,37].

Pre-leukaemic HSCs survive chemotherapy

To examine how $DNMT3A^{mut}$ affects population dynamics during leukaemic progression, we undertook temporal analysis of mature and progenitor cells from 5 patients (no. 11, 28, 35, 55, 57) sampled at diagnosis, remission (3 months) or relapse (Fig. 2b, d and Extended Data Figs 3 and 4). Compared to diagnosis, the allele frequency of $DNMT3A^{mut}$ alone was similar or higher at remission (patients no. 11, 28, 35, 57) and relapse (patients no. 11, 55). Although $CD33^+$ leukaemic blasts at diagnosis always carried both mutations, $CD33^+$ myeloid cells at remission bore only $DNMT3A^{mut}$, indicating that they are not AML blasts but the progeny of $DNMT3A^{mut}$-bearing progenitors with preserved myeloid differentiation capacity. In the relapse sample of patient no. 11, both mutations were present in the majority of cells, with the exception of HSCs/MPPs in which a proportion carried only $DNMT3A^{mut}$. Patient no. 57 was a long-surviving patient

that allowed a comparison of early and late (36 months) remission samples; this revealed a marked increase in *DNMT3A*^{mut} allele frequency in most cell populations over time (Fig. 2d). In addition, a small proportion of CD33^{+} myeloid cells in the late remission peripheral blood sample contained both *DNMT3A*^{mut} and the *NPM1c* mutation found at diagnosis, indicating either regrowth of the diagnostic leukaemic clone or emergence of a new clone following an independent *NPM1c* mutation event within the pre-leukaemic pool. Collectively, our data indicate that the ancestral cell that bears *DNMT3A*^{mut} without *NPM1c* is an HSC/MPP capable of multilineage differentiation. Moreover, these ancestral HSCs/MPPs survive chemotherapy, expand during remission, and might serve as a reservoir for clonal evolution leading to recurrent disease.

Pre-leukaemic HSCs undergo clonal expansion

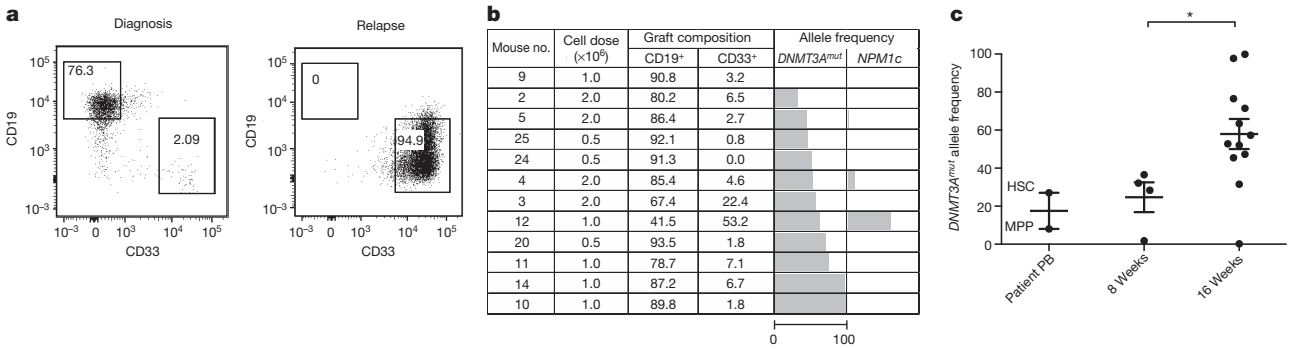
To establish conclusively whether phenotypically defined *DNMT3A*^{mut}-bearing HSCs/MPPs are functional HSCs and whether competitive repopulation advantage underlies their *in vivo* clonal expansion, we undertook xenograft repopulation assays. Mononuclear cells from the peripheral blood of two patients at diagnosis (no. 11, 55) with *DNMT3A*^{mut} allele frequency in HSCs/MPPs of 30% and 20%, respectively, were transplanted into cohorts of immunodeficient mice using a limiting dilution approach and analysed after 8 and 16 weeks (Fig. 3 and Extended Data Fig. 5). In this xenograft model, leukaemic engraftment is characteristically seen as a dominant myeloid (CD45^{dim}CD33^{+}) graft, whereas non-leukaemic grafts are multilineage and contain both lymphoid (predominantly CD19^{+} B cells) and myeloid (CD33^{+}) cells (Extended Data Fig. 6). For patient no. 11, multilineage engraftment was seen in 24 out of 35 mice, giving a calculated frequency of one repopulating HSC in 7.3×10^5 cells (Extended Data Fig. 5a). Only a single graft contained more than 50% CD33^{+} myeloid cells, consistent with co-engraftment by a leukaemia stem cell (LSC) that was present at low frequency[38]. We analysed by ddPCR 12 of the multilineage xenografts following 16 weeks of repopulation. Ten of these contained a high proportion of cells bearing *DNMT3A*^{mut} without *NPM1c* (mean allele frequency 57%), whereas both *DNMT3A*^{mut} and *NPM1c* were present in the single mouse with significant myeloid engraftment (Fig. 3b). Kinetic analysis demonstrated increasing *DNMT3A*^{mut} allele frequency in multilineage grafts over time (Fig. 3c). Similar results were found for patient no. 55 (Extended Data Fig. 5 and data not shown). In contrast, cells from the relapse sample of both patients generated leukaemic grafts and no multilineage grafts (Fig. 3a and data not shown), consistent with a higher LSC frequency at relapse compared to diagnosis[38]. Together, these data provide evidence that *DNMT3A*^{mut} occurs in HSCs/MPPs capable of generating a long-term multilineage lympho-myeloid graft, confirming their designation as pre-leukaemic HSCs[26]. *DNMT3A*^{mut} also endows pre-leukaemic HSCs with a competitive repopulation advantage over non-mutated HSCs, explaining the clonal expansion of pre-leukaemic HSCs in patients at the time of diagnosis and during remission.

Our xenograft results indicate that when pre-leukaemic HSCs exist at higher frequency than LSCs, non-leukaemic multilineage grafts, rather than leukaemic grafts, are frequently generated. Examination of our historical xenograft data from 264 diagnostic AML samples revealed that 37% did not generate any graft, 40% generated leukaemia, and 23% gave rise to non-leukaemic multilineage grafts (Extended Data Fig. 6). Sanger sequencing data was available for 25 samples that generated multilineage grafts (Supplementary Table 4), and revealed that 10 out of 25 (40%) came from patients bearing *DNMT3A*^{mut}; *IDH1*/*IDH2* mutations were present in 12 patients, including three who had both *DNMT3A* and *IDH1*/*IDH2* mutations (Fig. 4a, b). It was not possible to determine conclusively whether these xenografts were generated by pre-leukaemic HSCs, as they were not available for mutation testing. To examine whether pre-leukaemic cells also exist in patients with *IDH1*/*IDH2* mutations, we analysed samples from three patients with *IDH1* mutation and three patients with *IDH2* mutation by high-resolution cell sorting and ddPCR. In four patients, no pattern of preceding mutation was detected in non-leukaemic cell populations. However, in two patients we found *IDH2* mutation without *NPM1c* in a number of progenitor and mature cell populations (Fig. 4c), indicating that *IDH2* mutation might also occur as a pre-leukaemic event.

Our data predict that *DNMT3A* mutation may occur in healthy adults and pre-date AML diagnosis by months or even years. Through searches of exome sequence databases derived from peripheral blood (https://esp.gs.washington.edu/drupal/) we found that the frequency of the *DNMT3A* R882H variant (rs147001633) was 0.066% (3 in 4,545). Although this was considered to be a germline variant in this healthy adult cohort, our findings raise the possibility that the mutations detected in these studies may have originated from an HSC/MPP containing an acquired somatic *DNMT3A* mutation that underwent clonal expansion.

Discussion

Our study provides a number of key insights into the leukaemogenic process in human AML. Our findings establish the sequential order of mutation acquisition for the patients reported here: *DNMT3A*^{mut} occurs before *NPM1c* and *FLT3*-ITD. Additionally, we provide strong evidence for the presence, at diagnosis, of pre-leukaemic HSCs that are ancestral to the dominant AML clone. Based on our data, pre-leukaemic HSCs are prevalent among patients with *DNMT3A*^{mut}, which account for



Mouse no.	Cell dose (×10^6)	Graft composition		Allele frequency	
		CD19$^+$	CD33$^+$	*DNMT3A*^{mut}	*NPM1c*
9	1.0	90.8	3.2		
2	2.0	80.2	6.5		
5	2.0	86.4	2.7		
25	0.5	92.1	0.8		
24	0.5	91.3	0.0		
4	2.0	85.4	4.6		
3	2.0	67.4	22.4		
12	1.0	41.5	53.2		
20	0.5	93.5	1.8		
11	1.0	78.7	7.1		
14	1.0	87.2	6.7		
10	1.0	89.8	1.8		

Figure 3 | Pre-leukaemic HSCs bearing *DNMT3A*^{mut} generate multilineage engraftment and have a competitive advantage in xenograft repopulation assays. **a**, Representative flow cytometric analysis of engrafted human cells collected from NSG mouse bone marrow (BM) 16 weeks after intrafemoral transplantation of peripheral blood mononuclear cells (PBMNCs) from diagnosis and relapse samples of patient no. 11. **b**, Analysis of human graft composition in NSG mouse BM 16 weeks after intrafemoral transplantation of PBMNCs from the diagnosis sample of patient no. 11 across a range of cells doses. The percentage of human (CD45$^+$) B (CD19$^+$) and myeloid (CD33$^+$) cells was determined by flow cytometry. Mutant allele frequency (%) in the human graft was determined by droplet digital PCR (ddPCR) analysis of sorted human cells. The length of the bars is proportional to the mutant allele frequency (the scale bar under the first column applies to all columns). **c**, Summary of *DNMT3A*^{mut} allele frequency in the human graft from mice analysed by ddPCR 8 and 16 weeks after transplantation of PBMNCs from patient no. 11, compared to isolated haematopoietic stem cells/multipotent progenitors (HSCs/MPPs) from the patient's peripheral blood at diagnosis. *$P < 0.05$. Bars indicate mean and standard deviation.

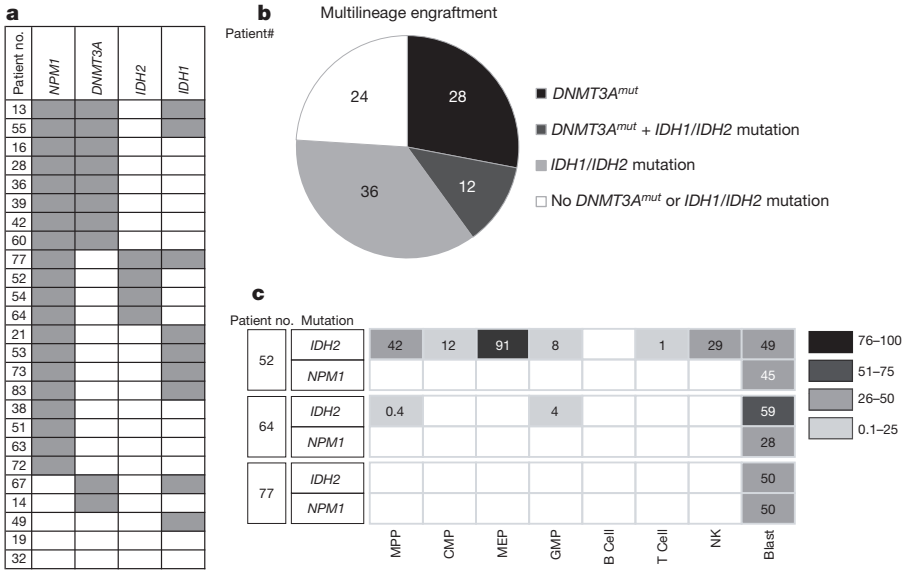


Figure 4 | Identification of pre-leukaemic HSCs with *IDH2* mutation. **a**, Summary of the occurrence of mutations in *NPM1*, *DNMT3A* and *IDH1/IDH2* determined by Sanger sequencing, in AML patient peripheral blood samples ($n = 25$) that generated a non-leukaemic multilineage graft after transplantation into immunodeficient mice. **b**, Representation of the proportion (%) of AML patient samples with *DNMT3A* and/or *IHD1/2* mutations among samples that generated a non-leukaemic multilineage graft in xenotransplanted mice. **c**, *IDH2* and *NPM1* mutant allele frequency (%) in stem/progenitor, mature lymphoid and blast ($CD45^{dim}CD33^{+}$) cell populations isolated from the peripheral blood of patients no. 52, 64 and 77 at diagnosis, as determined by droplet digital PCR (ddPCR). Blank boxes indicate no mutation detected.

25% of adult AML cases; additionally, our multilineage engraftment data indicate that pre-leukaemic HSCs may also exist in a proportion of AML patients with *IDH2* mutations. Pre-leukaemic progenitors of varying phenotypes have been reported in other types of hematologic malignancies[25,39–41], although functional studies were limited. Our work supports previous studies identifying phenotypic primitive cells that bear only a subset of mutations found in AML blasts[26,27,42]. The more precise analysis of highly resolved HSCs and progenitor populations that we have undertaken provides novel insight into the identity and proportional contribution of the stem/progenitor populations that acquire pre-leukaemic lesions. Furthermore, our work demonstrates that $DNMT3A^{mut}$ confers a functional repopulation advantage to pre-leukaemic HSCs over wild-type HSCs in xenograft assays, which probably underlies the clonal expansion of pre-leukaemic HSCs observed in patients at the time of diagnosis. Our study is consistent with mouse studies showing that HSCs lacking *DNMT3A* have a competitive growth advantage[43,44], and with a recent report predicting that the human $DNMT3A^{mut}$ results in loss of function[45].

Collectively, our results support a model wherein the cell of origin for $DNMT3A^{mut}$ AML is an HSC and the initiating *DNMT3A* mutation results in the generation of an expanded pool of HSCs and downstream progenitors, within which additional mutations such as *NPM1c* are acquired, driving progression to AML. In the samples studied here, our findings point to GMP and/or MLP as the likely populations in which *NPM1c* was acquired. Our experimental design provides a framework for the future identification of other early events in leukaemogenesis, using generation of multilineage xenografts as a surrogate assay to identify AML samples that may contain pre-leukaemic HSCs, and for examining how these changes disrupt normal HSC function, cause clonal expansion and initiate leukaemic development.

Our results have broad clinical implications. Previous studies in T-ALL[46] and B-ALL[7,11,14] revealed the existence of genetically diverse subclones at diagnosis. As found originally in these diseases[47] and now in AML, in approximately 50% of patients the relapse clone is not related to the predominant clone at diagnosis, but rather to a minor leukaemic subclone[10,15] or to a predicted ancestral clone[47]. Our direct demonstration that ancestral clones persist at remission suggests that pre-leukaemic HSCs are resistant to induction chemotherapy, and for some patients they might represent a reservoir from which relapse arises. If future phylogenetic single-cell lineage analysis establishes this possibility, then pre-leukaemic HSCs should be directly targeted to prevent relapse. As new drugs are developed that effectively target mutations in *DNMT3A* or other genes that give rise to pre-leukaemic HSCs (for example, AG-221, an IDH2 inhibitor), there may be an opportunity to eradicate these pre-leukaemic HSC clones before the acquisition of additional mutations renders them more resistant to therapy. Our findings also support broadening the definition of minimal residual disease to include not only the post-therapy survival of AML blasts and LSCs but also pre-leukaemic HSCs. Practically, this suggests that, for patients with both $DNMT3A^{mut}$ and *NPM1c*, the residual level of both mutations and not *NPM1c* alone should be monitored. Finally, our database analysis showing that the *DNMT3A* R882H variant is present in blood samples from normal adults should stimulate new studies investigating whether pre-leukaemic HSCs are present in healthy individuals and determining the risk of progression to AML; these may in turn enable earlier diagnosis for those patients who present without prior overt haematologic disturbances. Of note, a recent report describes the development of donor-derived leukaemia in a patient 27 months after receiving allogeneic HSCs from an HLA-matched sibling, whose donated peripheral blood cells were later found to carry mutations in *IDH2* and *DNMT3A* at low frequency[48]. Interestingly, the donor remains free of leukaemia 10 years after his stem-cell donation, indicating that the mechanisms underlying the progression from subclinical pre-leukaemic haematopoiesis to overt leukaemia are complex and context-dependent.

METHODS SUMMARY

All patient samples listed in Supplementary Tables 2 and 4 were obtained under Research Ethics Board approval with informed consent. Non-leukaemic stem, progenitor and mature cells were sorted from diagnostic samples and subjected to genomic analysis. Illumina sequencing libraries were constructed and target enrichment was performed using a custom Agilent SureSelect kit (following manufacturer's protocol). Sequencing was conducted on the Illumina HiSeq 2000 platform to an average on target coverage of 250×. Reads were aligned to the reference human genome build hg19 using Novoalign (Novocraft), and a BAM file was produced for each tumour and T-cell pair. Variant calls were made using the genome analysis tool kit (GATK). Significance levels (*P* values) were determined by chi-square test.

Targeted Sanger sequencing and ddPCR were performed for specific point mutations. Primary AML samples were also transplanted in xenograft assays using standard conditions.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 August 2013; accepted 20 January 2014.
Published online 12 February 2014.

1. Fialkow, P. J. *et al.* Clonal development, stem-cell differentiation, and clinical remissions in acute nonlymphocytic leukemia. *N. Engl. J. Med.* **317,** 468–473 (1987).
2. McCulloch, E. A., Howatson, A. F., Buick, R. N., Minden, M. D. & Izaguirre, C. A. Acute myeloblastic leukemia considered as a clonal hemopathy. *Blood Cells* **5,** 261–282 (1979).
3. Vogelstein, B., Fearon, E. R., Hamilton, S. R. & Feinberg, A. P. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science* **227,** 642–645 (1985).
4. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
5. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
6. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nature Rev. Genet.* **13,** 795–806 (2012).
7. Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* **469,** 356–361 (2011).
8. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–1113 (2010).
9. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464,** 999–1005 (2010).
10. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
11. Notta, F. *et al.* Evolution of human *BCR–ABL1* lymphoblastic leukaemia-initiating cells. *Nature* **469,** 362–367 (2011).
12. Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461,** 809–813 (2009).
13. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467,** 1114–1117 (2010).
14. Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322,** 1377–1380 (2008).
15. Shlush, L. I. *et al.* Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120,** 603–612 (2012).
16. Sgroi, D. C. Preinvasive breast cancer. *Annu. Rev. Pathol.* **5,** 193–221 (2010).
17. Wistuba, I. I., Mao, L. & Gazdar. Smoking molecular damage in bronchial epithelium. *Oncogene* **21,** 7298–7306 (2002).
18. Balaban, G. B., Herlyn, M., Clark, W. H. Jr & Nowell, P. C. Karyotypic evolution in human malignant melanoma. *Cancer Genet. Cytogenet.* **19,** 113–122 (1986).
19. Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319,** 525–532 (1988).
20. Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366,** 1090–1098 (2012).
21. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10,** 120–136 (2012).
22. Raza, A. & Galili, N. The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. *Nature Rev. Cancer* **12,** 849–859 (2012).
23. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Rev. Cancer* **12,** 599–612 (2012).
24. Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nature Genet.* **44,** 1179–1181 (2012).
25. Fialkow, P. J., Gartler, S. M. & Yoshida, A. Clonal origin of chronic myelocytic leukemia in man. *Proc. Natl Acad. Sci. USA* **58,** 1468–1471 (1967).
26. Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.* **4,** 149ra118 (2012).
27. Miyamoto, T., Weissman, I. L. & Akashi, K. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc. Natl Acad. Sci. USA* **97,** 7521–7526 (2000).
28. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368,** 2059–2074 (2013).
29. Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43,** 309–315 (2011).
30. Ley, T. J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).
31. Patel, J. P. *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N. Engl. J. Med.* **366,** 1079–1089 (2012).
32. Krönke, J. *et al.* Clonal evolution in relapsed *NPM1*-mutated acute myeloid leukemia. *Blood* **122,** 100–108 (2013).
33. Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature Immunol.* **11,** 585–593 (2010).
34. Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333,** 218–221 (2011).
35. Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature Immunol.* **14,** 756–763 (2013).
36. Kim, H. J. *et al.* Many multipotential gene-marked progenitor or stem cell clones contribute to hematopoiesis in nonhuman primates. *Blood* **96,** 1–8 (2000).
37. Fialkow, P. J., Janssen, J. W. & Bartram, C. R. Clonal remissions in acute nonlymphocytic leukemia: evidence for a multistep pathogenesis of the malignancy. *Blood* **77,** 1415–1417 (1991).
38. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature Med.* **17,** 1086–1093 (2011).
39. Jankowska, A. M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: *UTX*, *EZH2*, and *DNMT3A*. *Blood* **118,** 3932–3941 (2011).
40. Kikushige, Y. *et al.* Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* **20,** 246–259 (2011).
41. Walter, M. J. *et al.* Recurrent *DNMT3A* mutations in patients with myelodysplastic syndromes. *Leukemia* **25,** 1153–1158 (2011).
42. Chan, S. M. & Majeti, R. Role of DNMT3A, TET2, and IDH1/2 mutations in pre-leukemic stem cells in acute myeloid leukemia. *Int. J. Hematol.* **98,** 648–657 (2013).
43. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genet.* **44,** 23–31 (2012).
44. Tadokoro, Y., Ema, H., Okano, M., Li, E. & Nakauchi, H. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J. Exp. Med.* **204,** 715–722 (2007).
45. Kim, S. J. *et al.* A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* **122,** 4086–4089 (2013).
46. Clappier, E. *et al.* Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. *J. Exp. Med.* **208,** 653–661 (2011).
47. Inaba, H., Greaves, M. & Mullighan, C. G. Acute lymphoblastic leukaemia. *Lancet* **381,** 1943–1955 (2013).
48. Yasuda, T. *et al.* Leukemic evolution of donor-derived cells harboring *IDH2* and *DNMT3A* mutations after allogeneic stem cell transplantation. *Leukemia* http://dx.doi.org/10.1038/leu.2013.278 (15 October 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all members of the Dick laboratory for critical assessment of this work, A. Khandani, P. Penttilä, N. Simard, T. Velauthapillai and the SickKids-UHN flow facility for technical support, J. Claudio for management of the HALT studies that enabled the genetic analysis described herein, and J. Cui and X.-Z. Yang for curating the human AML samples used in these studies. This work was supported by a Postdoctoral Fellowship Award from the McEwen Centre for Regenerative Medicine with funding made available through the Gentle Ben Charity (L.I.S.), a Canadian Institutes for Health Research (CIHR) fellowship in partnership with the Aplastic Anemia and Myelodysplasia Association of Canada and an award from Vetenskapsradet (S.Z.), and by grants from CIHR, Canadian Cancer Society, Terry Fox Foundation, Genome Canada through the Ontario Genomics Institute, Ontario Institute for Cancer Research with funds from the province of Ontario, a Canada Research Chair, and the Ontario Ministry of Health and Long Term Care (OMOHLTC). The views expressed do not necessarily reflect those of the OMOHLTC. This work was also supported by the Cancer Stem Cell Consortium with funding from the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-047), and through the Canadian Institutes of Health Research (CSC-105367). Contributors to the HALT Pan-Leukemia Gene Panel are listed in Supplementary Note 1.

Author Contributions L.I.S. and S.Z. designed and performed experiments, analysed data and wrote the manuscript; A.M., W.C.C screened AML engraftment in xenotransplantation assays; J.M.B., V.G., J.A.K., A.D.S., A.C.S., K.W.Y., M.D.M. collected AML samples and assembled clinical information; J.A.K. correlated xenotransplantation engraftment data with clinical information; J.L.M., M.D. performed xenotransplantation experiments; J.J.F.M., R.M. performed ddPCR; H.J.K., K.L. performed Sanger sequencing; J.D.M., T.J.H., supervised the targeted sequencing; A.M.K.B. and F.Y. performed and analysed targeted sequencing; Q.M.T., L.D.S. performed *DNMT3A* data mining. M.D.M. designed the study; J.C.Y.W. supervised AML xenotransplantation screening experiments, designed the study and wrote the manuscript; J.E.D. supervised the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.D. (jdick@uhnresearch.ca).

METHODS

Targeted sequencing of leukaemia-associated genes. Genomic DNA was subjected to limited whole-genome amplification (RepliG, Qiagen) to obtain the required amount of input DNA for the SureSelect protocol. Amplified genomic DNA was mechanically sheared using the Covaris M220 Focused-ultrasonicator, and Illumina sequencing adaptors were ligated to fragments to make a sequencing library, which was then hybridized with 120mer biotinylated RNA library baits to capture the regions of interest. Baits were designed to capture the coding sequence of the 103 leukaemia-associated genes listed in Supplementary Table 1 (total target size ~370 kilobases). The targeted regions were pulled out using magnetic streptavidin beads and amplified. The resulting amplified library was quantified and sequenced on the Illumina HiSeq 2000 platform to an average on target coverage of 250×. Reads were aligned to the reference human genome build hg19 using Novoalign (Novocraft) and on-target single nucleotide variants (SNVs) and indels were called using the genome analysis tool kit (GATK). Somatic SNVs were called in AML blasts with a read depth of at least 30×. The list of contributors to the gene list is provided in Supplementary Note 1.

T cell isolation and expansion from primary AML samples. $CD3^+$ cells were isolated from peripheral blood AML patient samples using EasySep (Stem Cell Technologies) and re-suspended at a concentration of 1×10^7 cells in 2 ml in RPM1 + 10% FBS-HI + rhIL-2 (250 IU per mL, Proleukin, Chiron) + anti-CD28 antibody (5 μg ml^{-1}, clone CD28.2, eBioscience). Cells were then added to one well of a 24-well plate that had been pre-coated for 2 h with anti-CD3 antibody (Clone OKT3, eBioscience) and cultured for 4 days at 37 °C with 5% CO_2. Cells were collected on day 4, resuspended in fresh RPMI + 10% FBS-HI + rhIL2 (250 IU ml^{-1}) and replated into one well of a six-well plate. Cells were further cultured and expanded for 14–20 days, feeding with fresh full medium containing rhIL-2 (250 IU ml^{-1}) every 3–4 days. At the end of T cell expansion, the purity of $CD3^+$ T cells was checked by flow cytometry. DNA from the cultured T cells was extracted by PureGene Cell kit (Qiagen).

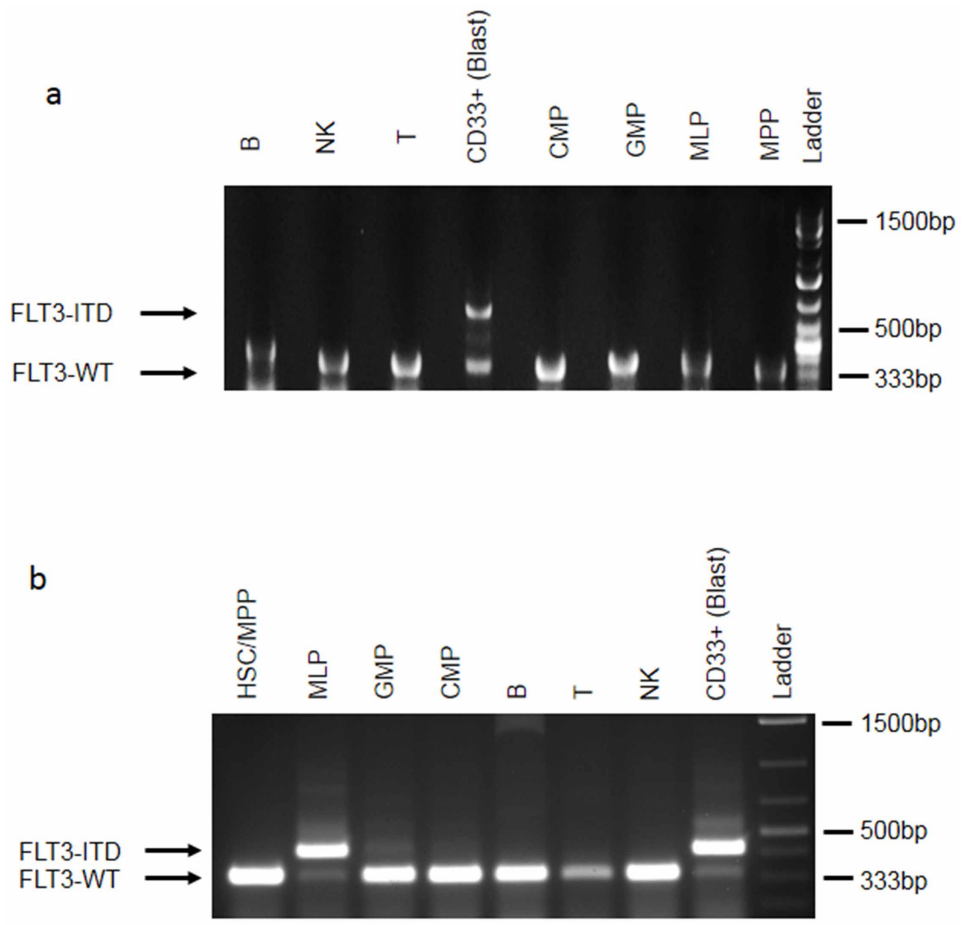
Droplet digital PCR (ddPCR). Genomic DNA (25 ng) or amplified DNA (2 μl from a 1:20 dilution of a 16 h RepliG whole-genome amplification) was subjected to ddPCR in a 96-well plate according to the manufacturer's protocol. Each sample was tested in duplicate. The plate was then loaded onto a droplet reader with a two colour FAM/VIC fluorescence detector. The mutant allele frequency was calculated as the fraction of positive droplets divided by total droplets containing a target. The TaqMan probes and primers used for ddPCR are listed in Supplementary Table 5. To evaluate the detection limits of the ddPCR assay, a standard curve was generated using serial dilutions of DNA with a known mutation frequency mixed with non-mutated DNA. For probes listed in Supplementary Table 5, the minimum detection level was 1:1,000 (0.1%).

Fluorescence activated cell sorting of human stem/progenitor and mature cell populations. Mononuclear cells (10^6 in 100 μl) from peripheral blood or bone marrow of AML patients were stained with the following antibodies (all from BD unless stated otherwise, dilution used and catalogue number in parentheses): anti-CD45RA-FITC (1:25, 555488), anti-CD90-APC (1:50, 561971), anti-CD135-Biotin (1:10, 624008), anti-CD38-PE-Cy7 (1:200, 335790), anti-CD10-Alexa-700 (1:10, 624040), anti-CD7-Pacific Blue (1:50, 642916), anti-CD45-V500 (1:200, 560777), anti-CD34-APC-Cy7 (1:100, custom made by BD, CD34 clone 581), anti-CD34-PerCP-Efluor 710 (1:100, e-Bioscience 46-0344-42), anti-CD33-PE-Cy5 (1:100, Beckman Coulter PNIM2647U), anti-CD19-PE (1:200, 349204), anti-CD3-FITC (1:100, 349201), anti-CD56-Alexafluor 647 (1:100, 557711), and Streptavidin-QD605 (1:200, Invitrogen Q10101MP). Samples from patients no. 1, 10, 11 (remission sample only), 32, 35, and 55 were enriched for $CD34^+$ cells using a Miltenyi CD34 MicroBead kit according to the manufacturer's protocol before antibody staining. Cells were sorted on a FACS AriaIII to a post-sort purity of >95%.

Xenotransplantation assays. Animal experiments were performed in accordance with institutional guidelines approved by the UHN Animal Care Committee. 8 to 12-week-old female NOD/SCID/IL-2Rg_c-null (NSG) mice were sublethally irradiated (225 cGy) 6–24 h before transplantation. Mononuclear cells from AML patients were depleted of $CD3^+$ cells by EasySep (Stem Cell Technologies) before intrafemoral transplantation. Mice were killed 8 or 16 weeks after transplantation and human engraftment in the injected femur and non-injected bone marrow was evaluated by flow cytometry using the following human-specific antibodies (all used at 1:200, all from BD unless stated otherwise, catalogue number in parentheses): anti-CD45-APC (340943), anti-CD19-PE, anti-CD33-PE-Cy5, anti-CD3-FITC, anti-CD14-PE Texas Red (Beckman Coulter PNIM2707U), anti-CD15-Pacific Blue (642917), anti-CD38-PE-Cy7, and anti-CD34-APC-Cy7. The threshold for detection of human engraftment was 0.1% $CD45^+$ cells. All flow cytometric analysis was performed on the LSRII (BD Biosciences). For limiting dilution assays, the frequency of repopulating cells was calculated using ELDA software[49].

Statistical analysis. For the initial targeted sequencing analysis, 12 independent patient samples were studied to capture the biologic diversity of AML. For validation of the *DNMT3A* findings, 71 samples were screened to identify at least 15 with *DNMT3A* mutations, as predicted by the known prevalence of *DNMT3A* mutation in AML. For limiting dilution analyses, at least 25 xenografts were analysed for each patient sample to ensure a large enough sample for statistical comparison. No animals or samples were excluded from any analysis. No formal randomization method was applied when assigning animals to different experimental groups. Group allocation and outcome assessment was not done in a blinded manner, including for animal studies. Frequency estimations were generated using the ELDA software, which takes into account whether the assumptions for LDA are met (http://bioinf.wehi.edu.au/software/elda/index.html, provided by the Walter and Eliza Hall Institute)[49]. *P* values were derived using two-tailed Student's *t*-tests. In each group of data, estimate variation was taken into account and is indicated as standard deviation. For all graphs, $*P = 0.01–0.05$, $**P = 0.001–0.01$, and $***P < 0.001$.

49. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347,** 70–78 (2009).
50. Kottaridis, P. D. *et al.* Studies of *FLT3* mutations in paired presentation and relapse samples from patients with acute myeloid leukemia: implications for the role of *FLT3* mutations in leukemogenesis, minimal residual disease detection, and possible therapy with FLT3 inhibitors. *Blood* **100,** 2393–2398 (2002).
51. Heinrich, V. *et al.* The allele distribution in next-generation sequencing data sets is accurately described as the result of a stochastic branching process. *Nucleic Acids Res.* **40,** 2426–2431 (2012).



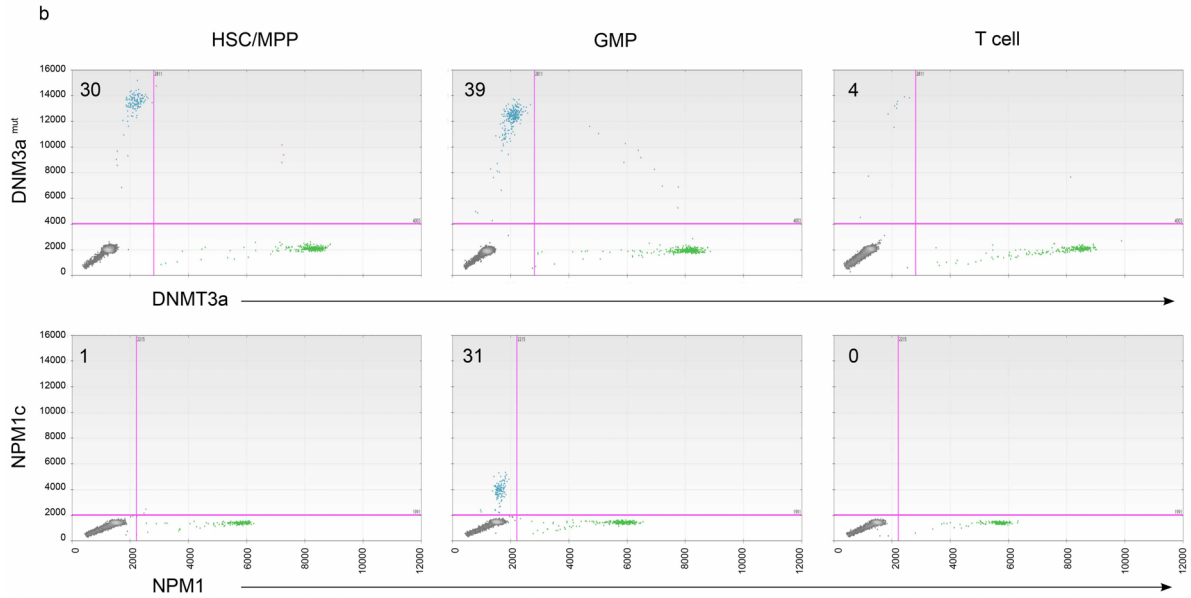
Extended Data Figure 1 | *FLT3*-ITD is a late event in patients carrying *DNMT3A* mutation. PCR analysis of *FLT3*-ITD[50] in stem/progenitor, mature lymphoid and blast ($CD45^{dim}$ $CD33^{+}$) cell populations from patient no. 13 (**a**) and no. 14 (**b**). *FLT3*-ITD was present in the blasts from both patients, and also in MLPs from patient no. 14. In contrast, $DNMT3A^{mut}$ without *FLT3*-ITD was detected in multiple non-blast cell populations (see Extended Data Fig. 2). HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MLP, multilymphoid progenitor; GMP, granulocyte monocyte progenitor; NK, natural killer cells.

a

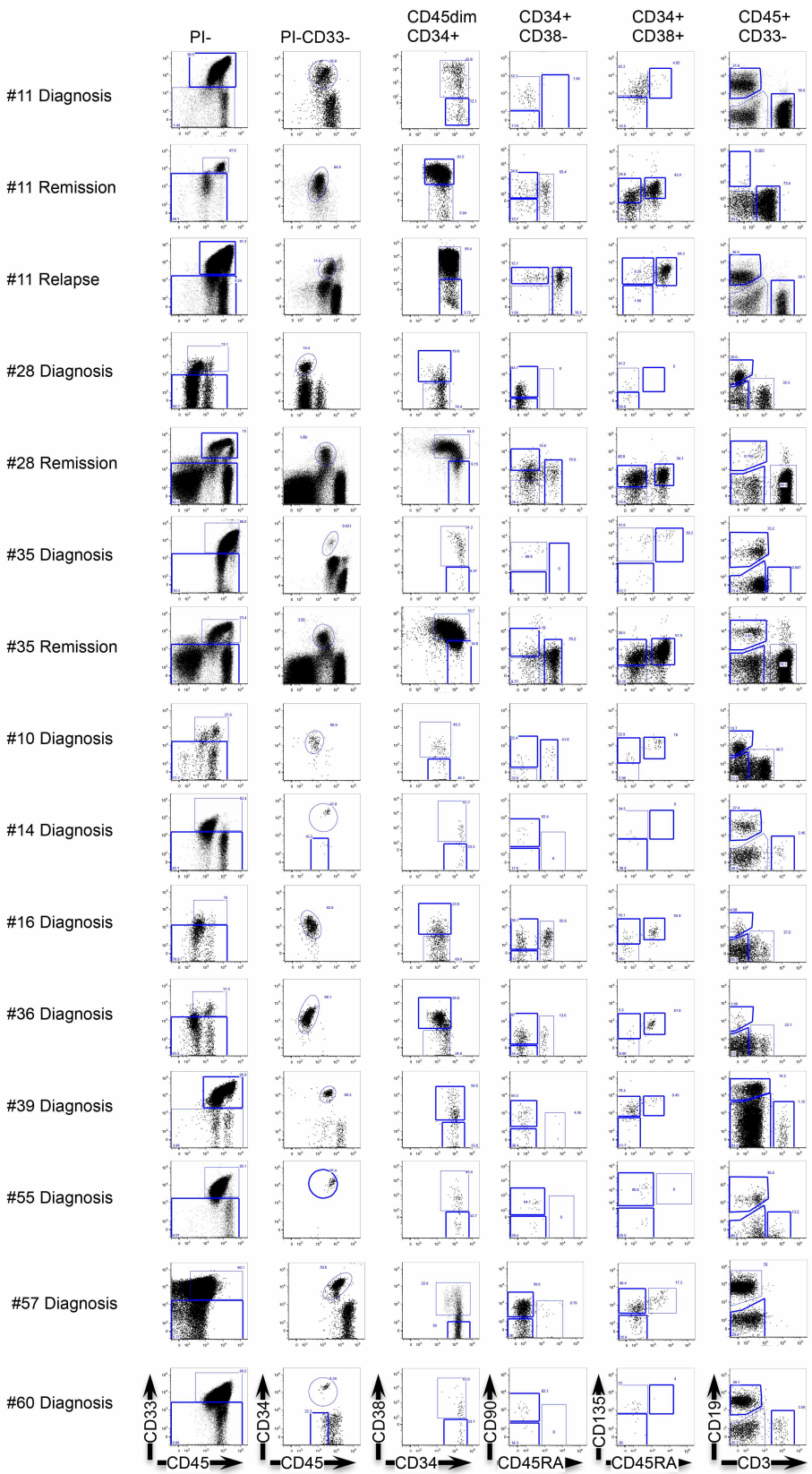
		HSC/MPP	CMP	MEP	MLP	GMP	B cell	T cell	NK	Blast
28	DNMT3a	2		60	NA	NA	0.2	3	27	40
	NPM1		NA	7	NA	67	0.2			39
57	DNMT3a	18	3	NA			28	0.1	3	41
	NPM1			NA						35
55	DNMT3a	21	66	28	NA	NA	11	6	37	52
	NPM1			1	NA	2	0.2	1	1	56
60	DNMT3a	61	68	94	69	68				41
	NPM1									43
35	DNMT3a	20	21	0.1		47	0.1		2	24
	NPM1									68
11	DNMT3a	30	13	55	NA	39	11	4	1	72
	NPM1				NA	31	1			46
16	DNMT3a		3		36	78				43
	NPM1					39			3	42
13	DNMT3a	69	12	23	87	76	NA	0.3		54
	NPM1			13	61	36	NA			63
36	DNMT3a	58	NA		57	96			0.4	64
	NPM1				3	94			1	36
10	DNMT3a		50	NA	64	93				36
	NPM1	1	0.1	NA	81	80			1	46
14	DNMT3a		2		67	33	1	NA		59
	NPM1				76	23				44

76-100
51-75
26-50
0.1-25

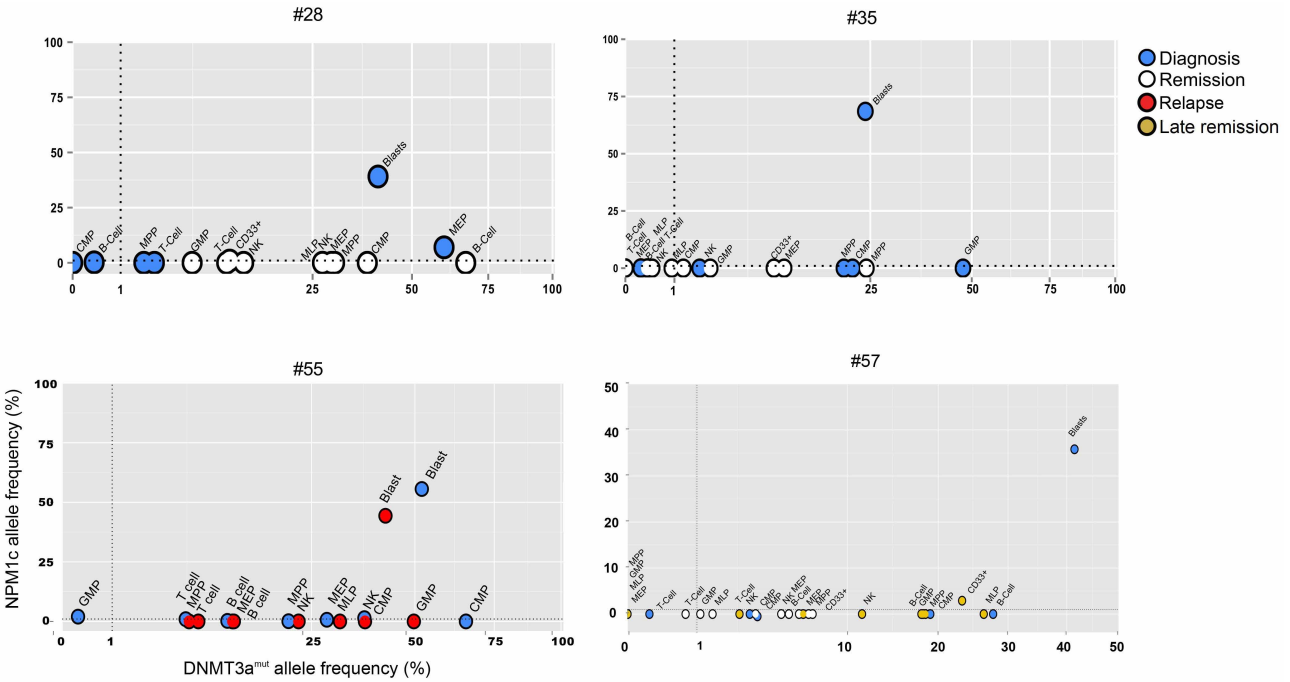
b



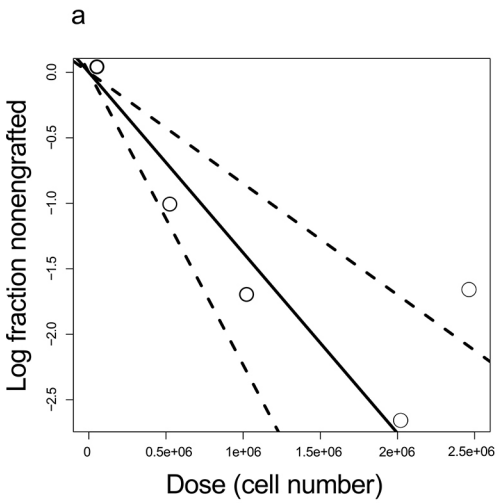
Extended Data Figure 2 | Frequent occurrence of *DNMT3A* mutation without *NPM1* mutation in stem/progenitor and mature lymphoid cells in AML patients at diagnosis. **a**, Summary of the allele frequency (%) of *DNMT3A* and *NPM1* mutations in stem/progenitor, mature lymphoid, and blast ($CD45^{dim}$ $CD33^{+}$) cell populations from 11 AML patient peripheral blood samples obtained at diagnosis, as determined by droplet digital PCR (ddPCR). Phenotypically normal cell populations were isolated by fluorescence activated cell sorting according to the strategy depicted in Fig. 2a. Mutant allele frequency ~50% is consistent with a heterozygous cell population. Departures from 50% mutant allele frequency may be stochastic[51], related to clonal heterogeneity, or due to the presence of copy number variations, for example loss of the wild type allele (loss of heterozygosity) or amplification of the mutant allele. NA, no cell population detected; HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; MEP, megakaryocyte erythroid progenitor; MLP, multilymphoid progenitor; GMP, granulocyte monocyte progenitor; NK, natural killer cells. Blank boxes indicate no *DNMT3A* or *NPM1* mutation detected. **b**, Representative plots showing ddPCR analysis of $DNMT3A^{mut}$ and *NPM1c* allele frequency in sorted cell populations from patient no. 11. The mutant allele frequency (%) is indicated on each plot.



Extended Data Figure 3 | Phenotypically normal stem/progenitor and mature cell populations are present in AML patient samples at diagnosis, remission and relapse. Flow cytometric analysis showing the gating strategy used to isolate phenotypically normal stem/progenitor and mature lymphoid cell populations from AML patient samples. Diagnosis and relapse samples are from peripheral blood; remission samples are from bone marrow.

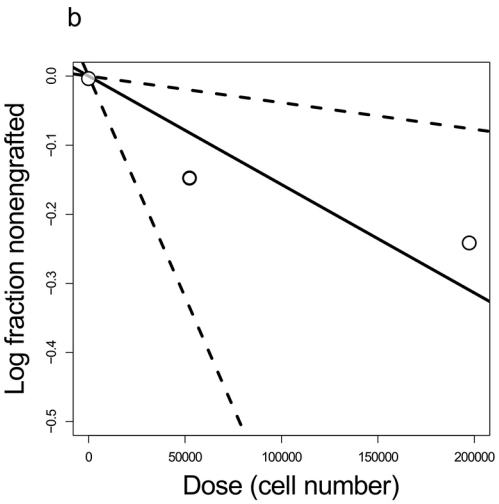


Extended Data Figure 4 | Cells bearing mutations in *DNMT3A* but not *NPM1* are present at diagnosis in AML patients and persist at remission and relapse. Allele frequency of *DNMT3A* and *NPM1* mutations of patients no. 28, 35, 55, and 57 in stem/progenitor, mature and blast ($CD45^{dim}$ $CD33^{+}$) cell populations, as determined by droplet digital PCR (ddPCR). Cells were isolated from diagnosis (blue), early remission (white), relapse (red) or late remission (yellow) samples. At remission, $CD33^{+}$ myeloid cells were also analysed. HSC, haematopoietic stem cell; MPP, multipotent progenitor; MLP, multilymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte monocyte progenitor; MEP, megakaryocyte erythroid progenitor; NK, natural killer cells.



Dose	Tested	Engrafted
2,500,000	5	4
2,000,000	7	7
1,000,000	10	8
500,000	8	5
100,000	5	0

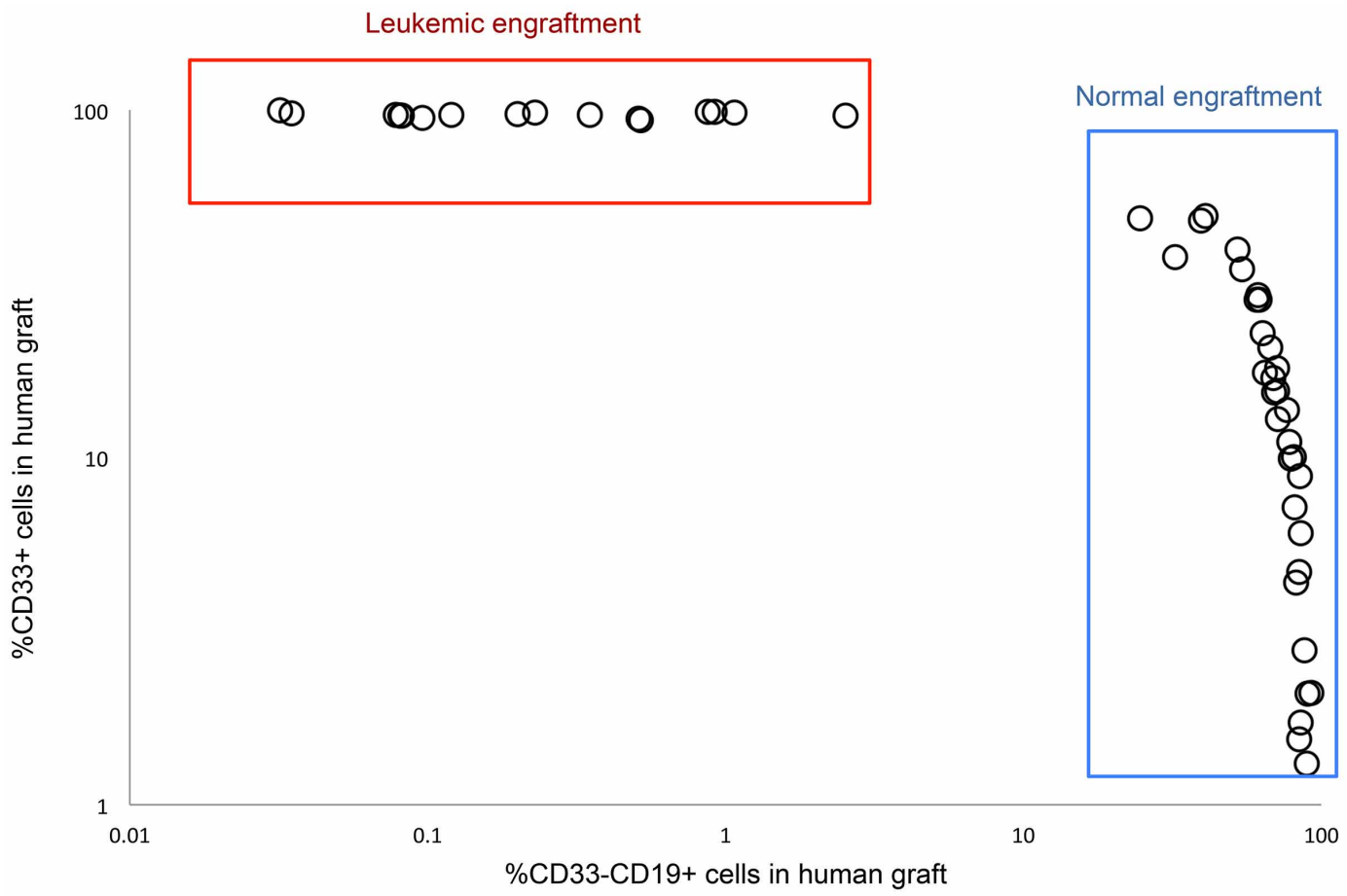
Estimated frequency : 1:725,897



Dose	Tested	Engrafted
200,000	5	1
50,000	8	1
8,000	7	0
1,000	5	0
100	5	0

Estimated frequency : 1:637,487

Extended Data Figure 5 | PreL-HSCs in the peripheral blood of AML patients generate multilineage human grafts in immunodeficient mice. Summary of results of limiting dilution experiments to assess frequency of pre-leukaemic HSCs generating multilineage grafts after xenotransplantation. Cohorts of NSG mice were transplanted intrafemorally with varying numbers of peripheral blood mononuclear cells from diagnostic samples of AML patient no. 11 (**a**) and no. 55 (**b**) and analysed after 8 or 16 weeks by flow cytometry. Engraftment was defined as >0.1% human $CD45^+$ cells in the injected right femur. Shown is the number of mice with multilineage human grafts containing both $CD33^+$ myeloid cells and $CD33^-CD19^+$ cells. The frequency of pre-leukaemic HSCs was calculated using the ELDA platform[49].



Extended Data Figure 6 | Frequent generation of non-leukaemic multilineage human grafts following xenotransplantation of peripheral blood cells from AML patients. Summary of xenograft characteristics in 123 sublethally irradiated NSG mice transplanted intrafemorally with mononuclear peripheral blood cells from 20 AML patients at diagnosis and analysed after 8 weeks by flow cytometry. The proportion of myeloid ($CD33^+$) and B-lymphoid ($CD33^-CD19^+$) cells in the human ($CD45^+$) graft is shown. Leukaemic (AML) engraftment is characterized by a dominant myeloid ($CD45^{dim}CD33^+$) graft, whereas non-leukaemic multilineage grafts contain both lymphoid (predominantly $CD33^-CD19^+$ B cells) and myeloid ($CD33^+$) cells. No leukaemic or multilineage graft could be detected in 65/123 mice (53%) in this cohort. Red box indicates AML grafts (27 mice, 22%); blue box indicates multilineage grafts (31 mice, 25%).

CORRIGENDUM

doi:10.1038/nature13190

Corrigendum: Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia

Liran I. Shlush, Sasan Zandi, Amanda Mitchell, Weihsu Claire Chen, Joseph M. Brandwein, Vikas Gupta, James A. Kennedy, Aaron D. Schimmer, Andre C. Schuh, Karen W. Yee, Jessica L. McLeod, Monica Doedens, Jessie J. F. Medeiros, Rene Marke, Hyeoung Joon Kim, Kwon Lee, John D. McPherson, Thomas J. Hudson, The HALT Pan-Leukemia Gene Panel Consortium, Andrew M. K. Brown, Fouad Yousif, Quang M. Trinh, Lincoln D. Stein, Mark D. Minden, Jean C. Y. Wang & John E. Dick

Nature **506,** 328–333 (2014); doi:10.1038/nature13038

Author Fouad Yousif (of the Ontario Institute for Cancer Research, Toronto, Canada) should have been included in the author list after Andrew M. K. Brown with affiliation number 7 and listed in the Author Contributions as performing and analysing targeted sequencing; these omissions have been corrected in the online versions of this Article. In addition, in the legend to Fig. 1a, "Somatic mutations in *DNMT3a* (*, R882H; †, R137C)" should read "Somatic mutations in *DNMT3a* (*, R882H; †, R326C)".