LETTER

doi:10.1038/nature22993

Tracing the origins of relapse in acute myeloid leukaemia to stem cells

Liran I. Shlush[1,2,3]*, Amanda Mitchell[1]*, Lawrence Heisler[4], Sagi Abelson[1], Stanley W. K. Ng[1], Aaron Trotman-Grant[1], Jessie J. F. Medeiros[1], Abilasha Rao-Bhatia[1], Ivana Jaciw-Zurakowsky[1], Rene Marke[5], Jessica L. McLeod[1], Monica Doedens[1], Gary Bader[6,7], Veronique Voisin[7], ChangJiang Xu[7], John D. McPherson[4], Thomas J. Hudson[4,6,8], Jean C. Y. Wang[1,9,10], Mark D. Minden[1,8,9,10] & John E. Dick[1,6]

In acute myeloid leukaemia, long-term survival is poor as most patients relapse despite achieving remission[1]. Historically, the failure of therapy has been thought to be due to mutations that produce drug resistance, possibly arising as a consequence of the mutagenic properties of chemotherapy drugs[2]. However, other lines of evidence have pointed to the pre-existence of drug-resistant cells[3]. For example, deep sequencing of paired diagnosis and relapse acute myeloid leukaemia samples has provided direct evidence that relapse in some cases is generated from minor genetic subclones present at diagnosis that survive chemotherapy[3–5], suggesting that resistant cells are generated by evolutionary processes before treatment[3] and are selected by therapy[6–8]. Nevertheless, the mechanisms of therapy failure and capacity for leukaemic regeneration remain obscure, as sequence analysis alone does not provide insight into the cell types that are fated to drive relapse. Although leukaemia stem cells[9,10] have been linked to relapse owing to their dormancy and self-renewal properties[11–13], and leukaemia stem cell gene expression signatures are highly predictive of therapy failure[14,15], experimental studies have been primarily correlative[7] and a role for leukaemia stem cells in acute myeloid leukaemia relapse has not been directly proved. Here, through combined genetic and functional analysis of purified subpopulations and xenografts from paired diagnosis/relapse samples, we identify therapy-resistant cells already present at diagnosis and two major patterns of relapse. In some cases, relapse originated from rare leukaemia stem cells with a haematopoietic stem/progenitor cell phenotype, while in other instances relapse developed from larger subclones of immunophenotypically committed leukaemia cells that retained strong stemness transcriptional signatures. The identification of distinct patterns of relapse should lead to improved methods for disease management and monitoring in acute myeloid leukaemia. Moreover, the shared functional and transcriptional stemness properties that underlie both cellular origins of relapse emphasize the importance of developing new therapeutic approaches that target stemness to prevent relapse.

To track clonal dynamics during leukaemia initiation and progression, we used a combined genetic and functional approach. Peripheral blood mononuclear cells collected from 11 patients with acute myeloid leukaemia (AML) at diagnosis and relapse were obtained from our biobank. Whole-genome sequencing (WGS) was performed (coverage ~50×) on leukaemic blasts isolated from each sample (clinical and immunophenotypic data in Supplementary Tables 1 and 2, respectively). T cells purified from diagnosis samples were used as the germline reference and to identify ancestral pre-leukaemic mutations[16]. To evaluate the genetic diversity of leukaemia stem cells (LSCs), diagnosis and relapse samples were transplanted into NOD/SCID/IL-2Rg$_c$-null (NSG) or NSG-SGM3 (humanized cytokine) mice[17]. Human myeloid ($CD45^+CD33^+$) and B-cell ($CD45^+CD19^+$) populations were sorted from xenografts and genotyped by droplet digital PCR (ddPCR, sensitivity ~1 in 1,000) for a subset of the variants identified by WGS. Additionally, patient samples from diagnosis, relapse, and remission time points (where available) were sorted into four progenitor ($CD33^-$ haematopoietic stem cells/multipotent progenitors (HSCs/MPPs), multi-lymphoid progenitors (MLPs), common myeloid progenitors/megakaryocyte erythroid progenitors (CMPs/MEPs), and granulocyte monocyte progenitors (GMPs)) and four mature ($CD45^{dim}CD33^+$ blasts, and T, B, and natural killer (NK) cells) populations using our established strategy[16], and genotyped (5–20 variants per sample) by ddPCR (experimental design is outlined in Extended Data Fig. 1a).

To investigate both the early stages of leukaemic development and the role of pre-leukaemic haematopoietic stem or progenitor cells (preL-HSPC) as a potential source of relapse[16,18], we identified somatic variants predicted to have a damaging effect on the encoded protein (protein-damaging variants, PDVs) (Supplementary Table 3). These were defined as pre-leukaemic (preL-PDV) if they were present in T cells sorted from patient samples, or in B cells sorted from xenografts, or leukaemic (L-PDV) if absent in these populations. Patients had an average of 22.3 ± 14.6 PDVs and, remarkably, 30–50% of these PDVs were pre-leukaemic (Fig. 1a). Cell populations carrying preL-PDVs but not L-PDVs were classified as pre-leukaemic, whereas those with both preL-PDVs and L-PDVs were considered leukaemic. We found evidence of a pre-leukaemic cell population in all patients except for number 9, who bore a *KMT2A* (also known as *MLL*) translocation; the lack of cells with pre-leukaemic mutations in this patient was consistent with previous reports of few cooperating mutations in *MLL* acute leukaemias[19,20]. Analysis of the occurrence and variant allele frequency (VAF) of PDVs enabled estimation of the order of acquisition of mutations in each patient with AML (Fig. 1b and Extended Data Fig. 1b). However, because 97.5% of PDVs identified at diagnosis were also present at relapse, they were not useful for identifying the cellular origin of the relapse clone. Nevertheless, this analysis revealed considerable genetic evolution occurring during the pre-leukaemic phase.

To investigate clonal dynamics between diagnosis and relapse time points, we identified somatic variants with VAF >20% in relapse blasts and <5% in blasts at diagnosis (relapse variants) (Supplementary Table 4). Relapse variants were largely intronic or intergenic, and compared with somatic mutations from the diagnosis sample were enriched for transversions (Supplementary Tables 5 and 6), consistent

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario M5G 2M9, Canada. [2]Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. [3]Division of Hematology Rambam Healthcare Campus, Haifa 31096, Israel. [4]Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. [5]Laboratory of Pediatric Oncology, Radboud University Medical Center, Nijmegen 6525 GA, The Netherlands. [6]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [7]Donnelly Centre for Cellular and Biomolecular Research, Toronto, Ontario M5S 3E1, Canada. [8]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. [9]Department of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [10]Division of Medical Oncology and Hematology, University Health Network, Toronto, Ontario M5G 2M9, Canada.
*These authors contributed equally to this work.

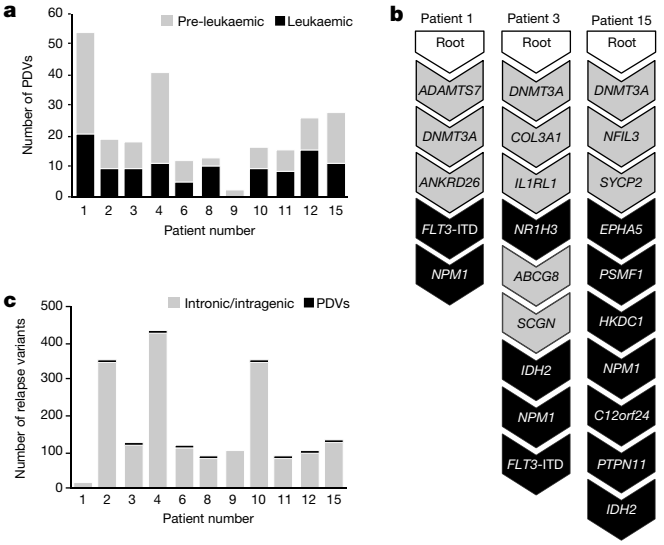


Figure 1 | Mutational analysis reveals extensive pre-leukaemic genetic evolution and major clonal shifts at relapse. a, Number of pre-leukaemic and leukaemic PDVs identified in 11 patients. **b**, Estimated order of acquisition of pre-leukaemic (grey) and leukaemic (black) PDVs in three patients, based on phylogenetic analysis. *FLT3*-ITD, *FLT3* with an internal tandem duplication; *C12orf24* is also known as *FAM216A*. **c**, Number of intronic or intragenic (grey) and protein-damaging (black) relapse variants identified in each patient.

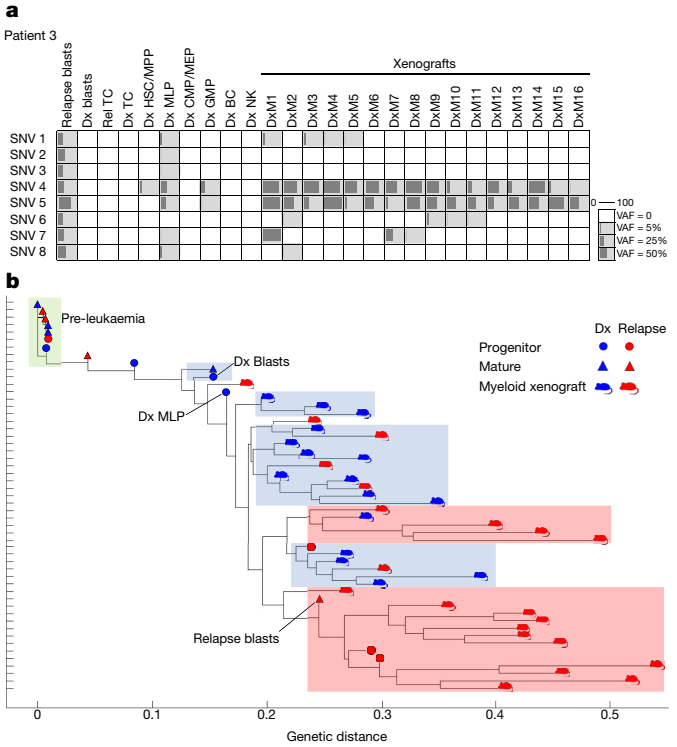


Figure 2 | AML relapse can originate from rare cells with a primitive functional and cellular phenotype. a, VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic (Dx) or relapse (Rel) sample of patient 3, or in myeloid ($CD33^+$) cells isolated from xenografts (DxM) generated by the diagnostic sample of the same patient. BC, B cell; CMP/MEP, common myeloid progenitor/megakaryocyte erythroid progenitor; GMP, granulocyte monocyte progenitor; HSC/MPP, haematopoietic stem cell/multipotent progenitor; MLP, multi-lymphoid progenitor; SNV, single nucleotide variant; TC, T cell. White squares indicate populations without relapse variant detection; relapse variants were detected in populations with grey squares, with the dark grey bar indicating VAF. **b**, Phylogenetic tree showing clonal relationships for patient 3, based on analysis of VAFs of all relapse variants, and pre-leukaemic and leukaemic mutations. The genetic distance between any two respective symbols on the phylogenetic tree was estimated using a neighbour joining method based on the degree of genetic relatedness (Nei's genetic distance). Each coloured box contains highly related cell populations and xenografts inferred to represent a clone (green, pre-leukaemic; blue, diagnostic; red, relapse).

with previous reports[3,5]. To track the cellular origins of the relapse clone(s), we analysed the occurrence and VAF of relapse variants in primitive and mature cell populations sorted from the diagnostic sample as well as in xenografts generated from the diagnostic and relapse samples. Patients had an average of 169 ± 135 relapse variants, consistent with a clonal switch between diagnosis and relapse. However, in one patient (number 1) the existence of only a very small number of relapse variants indicated that the diagnosis and relapse blasts were highly related (Fig. 1c and Extended Data Fig. 2a). In this patient, all preL-PDVs and L-PDVs were present at high VAF in the blast-like myeloid population at remission. These data suggest that, despite the achievement of a clinical and morphological remission in this patient, LSCs from the dominant clone at diagnosis (as functionally defined by the xenograft assays) survived chemotherapy and regenerated the same clone upon relapse. Indeed, reconstruction of a phylogenetic lineage tree using VAFs of PDVs in sorted cell populations and xenografts from diagnosis and relapse demonstrated the presence of several genetically diverse LSC subclones at diagnosis that persisted in relapse, with the dominant diagnostic clone re-emerging as the dominant relapse clone (Extended Data Fig. 2a).

In three patients (numbers 3, 6, and 9), relapse variants could not be detected by ddPCR in diagnostic blasts. Xenografts generated from diagnostic samples contained L-PDVs and were thus leukaemic, and frequently carried one or more relapse variants (Fig. 2a and Extended Data Figs 2b and 3), establishing that LSCs bearing relapse variants were already present at diagnosis. In patient 3, all diagnostic xenografts bore relapse variants, suggesting that the LSCs responsible for generating the dominant clone in the patient at the time of diagnosis were consistently out-competed in mice by LSC subclones bearing relapse variants. Phylogenetic analysis based on xenografts demonstrated genetic diversity in the LSC compartment at diagnosis, with at least four or five distinguishable subclones (Fig. 2b), most of which were also detected at relapse (Supplementary Table 8). These results contrast with those of a previous study that was not able to demonstrate propagation of relapse-fated subclones in xenografts[21], probably because of differences in xenograft methodology. The results of our functional studies formally establish that LSCs are genetically diverse both at diagnosis and relapse, indicating continued branching evolution after leukaemic transformation. Although T cells from both diagnosis and relapse contained preL-PDVs, the absence of detectable relapse variants in these T-cell populations suggests that relapse did not originate from preL-HSPCs, although we cannot exclude the possibility that relapse variants in T cells might be below our limit of detection and/or that some pre-leukaemic mutations might engender a lineage bias precluding differentiation into lymphoid cells. In this patient (number 3), phenotypic MLPs carried most of the tested relapse variants at low VAF (Fig. 2a), whereas phenotypic HSC/MPPs and granulocyte monocyte progenitors carried only one or two relapse variants. These progenitor populations all contained L-PDVs at high VAF, establishing that they are leukaemic (Supplementary Table 8). Thus, compared with other subpopulations, the MLP fraction exhibited the closest genetic relationship to all the evolving LSC subclones of the diagnosis sample, including the specific LSC subclone fated to drive the dominant relapsing clone. These findings are consistent with previous studies suggesting that LSCs can have an MLP phenotype[22] and can drive AML intiation[16]. On the basis of the low frequency of MLPs in the peripheral blood sample (<0.001% of total mononuclear cells) and low VAF of relapse variants in the MLP fraction (~0.1%), we estimated that the frequency of relapse variant-bearing MLPs was <0.0002% of the total

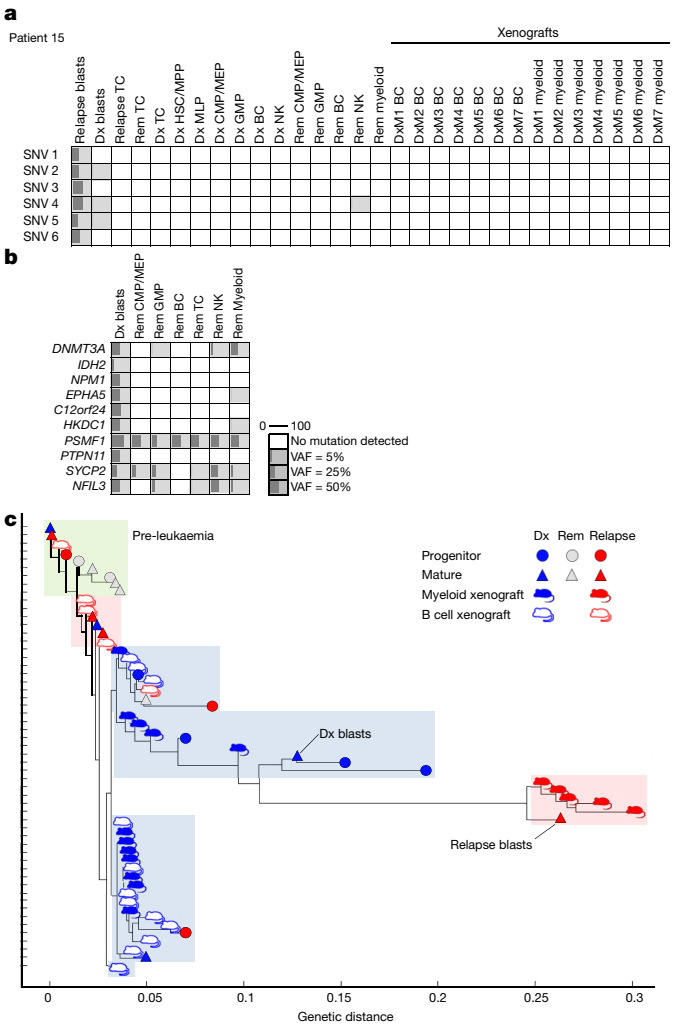


Figure 3 | AML relapse can originate from cells with a committed phenotype. a, VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission (rem), or relapse samples of patient 15, or in CD19$^+$ B or CD33$^+$ myeloid cells isolated from xenografts (DxM) generated from the diagnostic sample of the same patient. Abbreviations and display criteria as in Fig. 2a. **b**, VAFs of PDVs as detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 15. **c**, Phylogenetic tree showing clonal relationships for patient 15, based on analysis of VAFs of all relapse variants pre-leukaemic and leukaemic mutations. Display criteria as in Fig. 2b.

mononuclear population in this patient's peripheral blood. A similar pattern of relapse variants was seen in patients 6 and 9 (Extended Data Figs 2b and 3). Overall, our findings provide functional and phenotypic evidence that, in some patients, the cellular origin of relapse is a rare population of cells with a primitive HSPC-like phenotype, which are already present at diagnosis before initiation of therapy; we term these cases as having a primitive relapse origin (relapse origin-primitive, RO$_P$).

In four other patients (2, 8, 10, and 15), relapse variants were detectable at low VAF (0.2–12.5%) in the non-HSPC blast population, but not in any other sorted progenitor population from the diagnostic sample (detection limit ~0.1%) (Fig. 3a and Extended Data Figs 4–6). For these patients, a bone marrow sample from remission (after second consolidation) was available, in which L-PDVs were detected in CD33$^+$ cells as well as in some progenitor populations (Fig. 3b and Extended Data Figs 4–6), indicating persistence of leukaemic cell populations. Interestingly, relapse variants were below the limit of detection in the remission sample. To determine the clonal makeup of the LSC compartment, we evaluated xenografts generated by the patients' diagnostic samples. The sample from patient 10 generated leukaemic grafts in NSG mice. The sample from patient 2 generated only multilineage grafts in both NSG and NSG-SGM3 mice. The samples from patients 8 and 15 did not generate any graft in NSG mice; however, cells from patient 8 generated a leukaemic graft and those from patient 15 generated a multilineage graft in NSG-SGM3 mice (Supplementary Table 7). Among these diagnostic xenografts, relapse variants were detected only in NSG-SGM3 xenografts from patient 8. For patient 15, phylogenetic analysis including relapse xenografts demonstrated that relapse was oligoclonal, and that the cell fraction at diagnosis most genetically related to the major relapse clone was the CD33$^+$ blast population (Fig. 3c). Together, these data suggest that, in this group of patients, relapse originated from cells with a more committed immunophenotype (relapse origin-committed, RO$_C$), and that LSCs bearing relapse variants in RO$_C$ cases might possess more stringent growth requirements in the xenograft model than those of RO$_P$ patients.

In the remaining three patients (4, 11, and 12) (Extended Data Figs 7a, b and 8), no relapse variants could be validated by ddPCR in any of the sorted cell populations or xenografts from the diagnostic sample, preventing conclusive analysis of relapse origins; deeper sequencing will be required to determine whether ultra-rare cell types responsible for relapse are present. Overall our data demonstrate that AML undergoes complex clonal evolution within both the pre-leukaemic HSC and LSC compartments, and we identify at least two distinct patterns of relapse based on the cell types from which relapse can originate. The presence at diagnosis of LSCs bearing relapse variants strongly supports the concept that chemotherapy does not induce mutations leading to emergence of new clones, but rather selects for pre-existing subclones that are already therapy resistant, probably because of dormancy and/or epigenetic plasticity; these findings are consistent with studies of mouse models of AML[23]. Indeed, the existence of minor LSC subclones within the bulk leukaemic populations is suggestive of their dormancy. Nevertheless, relapse variants were enriched for transversions, as typically seen after mutagen exposure[2]; further investigation is needed to understand why this mutagenic signature is present in relapse-generating cells before therapy.

To investigate the molecular pathways underlying the different patterns of relapse (RO$_C$ and RO$_P$), we undertook transcriptional profiling of the bulk diagnosis and relapse samples (Supplementary Table 9). Unsupervised hierarchical clustering of the normalized read counts revealed two major clusters encompassing RO$_P$ (3, 6, and 9) and RO$_C$ (2, 8, 10, and 15) patients (Fig. 4a). Since our patient numbers were small, we applied the same unsupervised clustering to patients who relapsed in The Cancer Genome Atlas (TCGA) AML cohort ($n = 84$). Again, two distinct clusters were generated; gene set enrichment analysis[24] revealed that the same gene expression programs were driving the distinct clusters in both datasets (Fig. 4a and Extended Data Fig. 9a). The TCGA RO$_C$-like cluster ($n = 58$) was enriched for patients with more primitive AML subtypes (French–American–British (FAB) M0/1/2); in this group, there was a trend to a higher proportion of patients with complex karyotype (Extended Data Fig. 9b). Patients in the RO$_P$-like TCGA cluster were older and had higher white blood-cell counts at diagnosis; this group was enriched for patients with myelomonocytic subtypes (FAB M4/M5), intermediate risk cytogenetics, *NPM1c* mutation, and chromosomal alteration inv(16) (Extended Data Fig. 9b). The primitive immunophenotype of the rare MLP-like cells that generate relapse in RO$_P$ patients was distinct from the gene expression profiles of the bulk population; bulk cells were more similar to those of mature monocytic/granulocytic lineage cells, in keeping with their more differentiated FAB subtypes (Fig. 4b). By contrast, the bulk gene expression profiles of RO$_C$ patients were most similar to those of normal HSCs and erythroid/megakaryocytic progenitors, consistent with their primitive FAB subtypes. Collectively, these data suggest that, despite their phenotypic and genetic dissimilarities, relapse in both patient groups is linked to stem cell properties, manifested

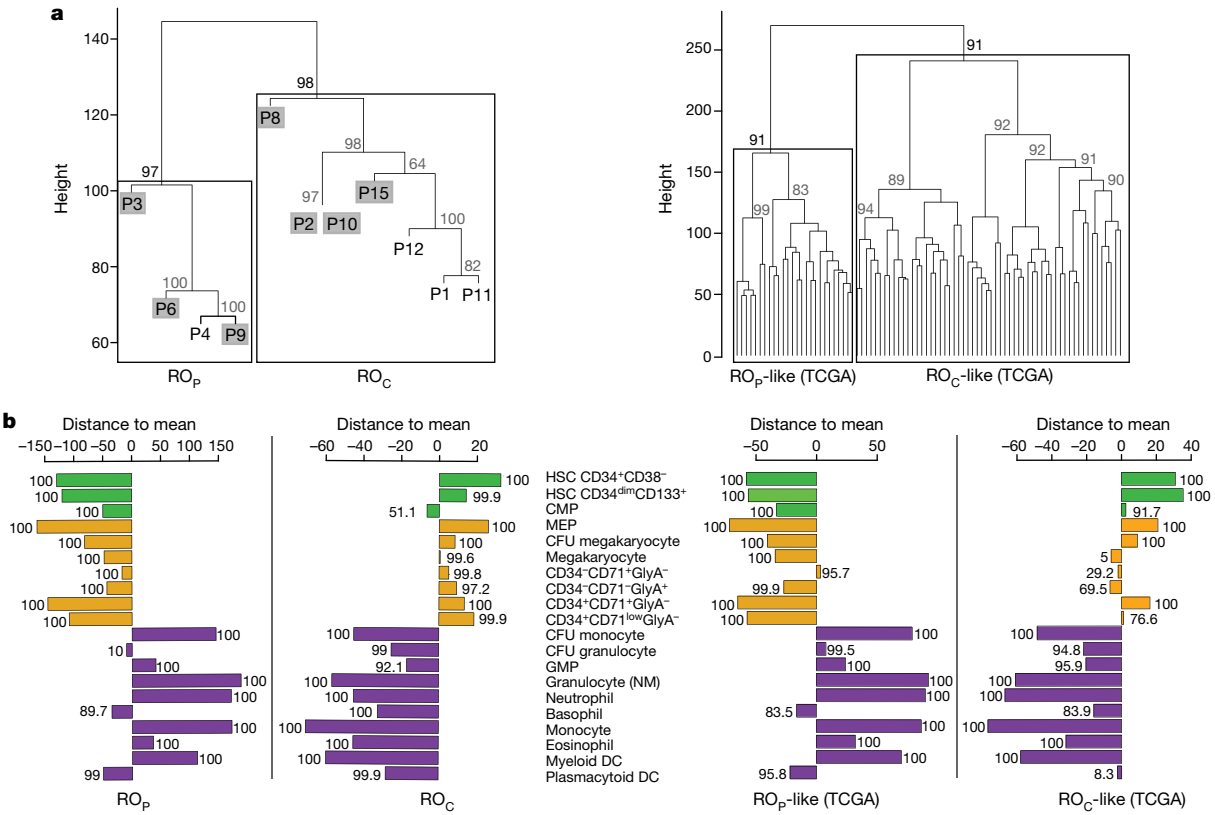


Figure 4 | Distinct gene expression patterns are associated with relapse origin. a, Unsupervised hierarchical clustering of gene expression data from bulk diagnostic samples from the Princess Margaret Cancer Centre (PM) cohort ($n = 11$, left) or the TCGA relapsed AML cohort ($n = 84$, right). RO_P-like and RO_C-like clusters are indicated by boxes. Patients (P#) for whom the relapse origin was functionally established are shaded. Values above each edge represent approximately unbiased *P* values (percentage) calculated using pvclust. **b**, Differentially expressed genes between RO_P-like and RO_C-like clusters in the PM cohort (left) or the TCGA relapsed AML cohort (right) compared with Gene Expression Omnibus dataset GSE24759 (DMAP) populations. Numbers beside each bar indicate the percentage of time for which the observed value (set of up- or downregulated genes) was better represented in that population than random values (equal number of randomly selected genes based on 1,000 trials). CFU, colony-forming unit; DC, dendritic cell; NM, neutrophilic metamyelocyte.

either as a primitive LSC population giving rise to relapse (RO_P group), or as stemness transcriptional programs that are retained in the bulk leukaemia cell population (RO_C group) (summarized in Supplementary Table 10).

To investigate whether the clonal switch from diagnosis to relapse was associated with gene expression changes, a perturbation model was used to examine the transcription profiles of our paired samples[25]. Patients in the RO_P group exhibited an increase in the proportion of cells with HSC gene expression (Extended Data Fig. 10a) and a concomitant decrease in the proportion of cells with mature myeloid lineage programs between diagnosis and relapse time points (Extended Data Fig. 10b). Consistent with this observation, the LSC frequency increased in the relapse samples (Extended Data Fig. 10c). The RO_C group exhibited more inter-sample variation; however, in general, the already high proportion of diagnosis cells expressing HSC programs in the absence of mature myeloid programs changed little at relapse (Extended Data Fig. 10a, b). Additionally, an LSC-based gene expression score that was highly predictive of response to standard AML induction therapy[15] was also higher when measured at relapse compared with diagnosis (Extended Data Fig. 10d). Our overall interpretation of these datasets is that the phenotypic and transcriptional properties reflective of monocytic lineage maturation seen in RO_P cases is indicative of a deep developmental hierarchy wherein a small pool of LSC generate bulk blasts that exhibit extensive myeloid differentiation; the seeds of relapse lie buried within these rare LSCs. At relapse, the higher LSC to blast ratio and the primitive blast phenotype with increased expression of stemness programs reflect a switch towards a more shallow developmental hierarchy upon leukaemic progression. By contrast, the more primitive blast phenotype in RO_C cases denoted by strong expression of stemness transcriptional programs suggests that a shallow leukaemic hierarchy already exists at diagnosis in these cases, perhaps because of their poor-risk molecular properties that result in greater impairment of myeloid commitment programs.

Here, we have tracked the complex evolutionary history of AML within individual patients from the early stages of pre-leukaemic development to diagnosis and through progression to relapse. Our results have considerable implications for cancer biology in general as well as for how AML is monitored and treated. AML is the best example of a tumour that follows the cancer stem cell model. Our findings, together with recent evidence that stemness transcriptional signatures are highly predictive of therapy response[14,15], provide strong evidence that cancer stem cells and stemness properties are clinically relevant in human AML, helping to resolve longstanding controversies surrounding this model and the experimental methods required to investigate cancer stem cells[26]. Accordingly, therapeutic strategies must not only eradicate bulk tumour cells but also effectively target cancer stem cells. The presence of genetically diverse LSCs at diagnosis highlights a major limitation of therapies that target only the specific properties of the dominant clone. Our findings predict that RO_P versus RO_C cases will require different therapies given their different biology. The minor populations of dormant LSCs will need to be targeted in RO_P cases, whereas relapse might be prevented in RO_C cases by targeting the dominant blast population, for example with anti-CD33-drug toxin conjugates or other bi-specific antibodies. The ability of xenografts to capture even rare relapse-initiating LSCs offers

a new approach to characterize their therapy resistance and identify new therapeutic targets, a strategy that should be generalizable to other leukaemias and solid cancers. Finally, deployment of new methods of disease monitoring that combine genetic and cellular analysis could lead to improved clinical trials in which the drug response of genetically diverse LSC subclones is tracked.

Online Content Methods, along with any additional 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 23 December 2016; accepted 17 May 2017.
Published online 28 June 2017.

1. Sarkozy, C. *et al.* Outcome of older patients with acute myeloid leukemia in first relapse. *Am. J. Hematol.* **88,** 758–764 (2013).
2. Goldie, J. H. & Coldman, A. J. The genetic origin of drug resistance in neoplasms: implications for systemic therapy. *Cancer Res.* **44,** 3643–3653 (1984).
3. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
4. Krönke, J. *et al.* Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. *Blood* **122,** 100–108 (2013).
5. Garg, M. *et al.* Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood* **126,** 2491–2501 (2015).
6. Parkin, B. *et al.* Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia. *Blood* **121,** 369–377 (2013).
7. Bachas, C. *et al.* The role of minor subpopulations within the leukemic blast compartment of AML patients at initial diagnosis in the development of relapse. *Leukemia* **26,** 1313–1320 (2012).
8. Greaves, M. Evolutionary determinants of cancer. *Cancer Discov.* **5,** 806–820 (2015).
9. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367,** 645–648 (1994).
10. Hope, K. J., Jin, L. & Dick, J. E. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* **5,** 738–743 (2004).
11. Ishikawa, F. *et al.* Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. *Nat. Biotechnol.* **25,** 1315–1321 (2007).
12. Costello, R. T. *et al.* Human acute myeloid leukemia $CD34^+/CD38^-$ progenitor cells have decreased sensitivity to chemotherapy and Fas-induced apoptosis, reduced immunogenicity, and impaired dendritic cell transformation capacities. *Cancer Res.* **60,** 4403–4411 (2000).
13. Terpstra, W. *et al.* Fluorouracil selectively spares acute myeloid leukemia cells with long-term growth abilities in immunodeficient mice and in culture. *Blood* **88,** 1944–1950 (1996).
14. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17,** 1086–1093 (2011).
15. Ng, S. W. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540,** 433–437 (2016).
16. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506,** 328–333 (2014).
17. Wunderlich, M. *et al.* AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* **24,** 1785–1788 (2010).
18. Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl Acad. Sci. USA* **111,** 2548–2553 (2014).
19. 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).
20. Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* **47,** 330–337 (2015).
21. Klco, J. M. *et al.* Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* **25,** 379–392 (2014).
22. Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19,** 138–152 (2011).
23. Fong, C. Y. *et al.* BET inhibitor resistance emerges from leukaemia stem cells. *Nature* **525,** 538–542 (2015).
24. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
25. Qiao, W. *et al.* PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLoS Comput. Biol.* **8,** e1002838 (2012).
26. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14,** 275–291 (2014).

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

Acknowledgements This work was supported by grants from the Ontario Institute for Cancer Research with funds from the province of Ontario, the Cancer Stem Cell Consortium with funding from the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-047), and the Canadian Institutes of Health Research (CSC-105367), the Canadian Cancer Society, the Terry Fox Foundation, a Canada Research Chair to J.E.D. L.I.S. was funded by the Benjamin Pearl fellowship from the McEwen Centre for Regenerative Medicine and an American Society of Hematology Scholar Award. This research was funded in part by the Ontario Ministry of Health and Long Term Care, whose views are not expressed here. We thank all the members of the Dick laboratory for comments and P. Zandstra for supporting this work.

Author Contributions L.I.S. and A.M. developed the study and workflow, and performed all the xenografting, cell purification, and genotyping analysis. M.D.M. provided clinical annotated samples from the Princess Margaret Cancer Centre AML biobank. L.H., J.D.M., and T.J.H. performed sequencing. A.T.-G., J.J.F.M., A.R.-B., I.J.-Z., and R.M. performed genotyping. J.L.M. and M.D. supported xenografting, S.A. helped with identification of genetic variants and PDVs. S.W.K.N., V.V., C.X., and G.B. provided gene expression, and statistical, bioinformatic, and perturbation (PERT) analysis. L.I.S. and A.M. wrote the paper. L.I.S., A.M., J.C.Y.W., M.D.M., and J.E.D. revised the paper. J.E.D. provided funding and study supervision.

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. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.E.D. (jdick@uhnresearch.ca).

Reviewer Information *Nature* thanks M. Greaves, R. Levine and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Patient samples. All biological samples were collected with informed consent according to procedures approved by the Research Ethics Board of the University Health Network (REB 01-0573-C) and viably frozen in the PM Leukaemia Bank. Inclusion criteria were AML samples for which viable frozen cells were available from both diagnosis and relapse. The original cohort included 15 patients. The following samples were excluded because of insufficient material needed for all of the experiments: patients 7, 13, and 14. Patient 5 was excluded as we originally thought the patient had paired samples from diagnosis and relapse; however, later it became clear that the diagnosis sample was actually a remission sample, and no diagnosis sample was available.

T-cell isolation and expansion from primary AML patient samples. $CD3^+$ cells were isolated from peripheral blood mononuclear cells of AML patient samples using EasySep (Stem Cell Technologies) and re-suspended at a concentration of 1×10^7 cells per 2 ml in RPM1 + 10% FBS-HI + rhIL-2 (250 IU ml^{-1}, 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 harvested on day 4, re-suspended in fresh RPMI + 10% FBS-HI + rhIL2 (250 IU ml^{-1}), and re-plated 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. After expansion, the purity of $CD3^+$ T cells was checked by flow cytometry. DNA from the cultured T cells was extracted using the PureGene Cell kit (Qiagen) and used for WGS.

WGS. DNA was extracted from bulk peripheral blood mononuclear cells (after Ficoll separation) from both diagnosis and relapse AML patient samples and quantified using Qubit (Life Technologies, Carlsbad, California, USA, catalogue number Q32854). Fifty nanograms was used for library preparation using the Nextera DNA Sample Prep Kit (Illumina, San Diego, California, USA, catalogue numbers FC-121-1031 and FC-121-1011). Tagmentation of genomic DNA was performed following the manufacturer's protocol. Tagmented DNA was purified using 1.8× volume AMPure XP SPRI beads (Beckman Coulter Genomics, Danvers, Massachusetts, USA, catalogue number A63881) and then amplified using a limited-cycle PCR program according to the manufacturer's protocol. Post-PCR enriched libraries were purified using a 0.6× volume AMPure XP SPRI bead clean-up, followed by two washes with 80% EtOH. Libraries were eluted in 32.5 μl suspension buffer provided in a Nextera DNA Sample Kit, validated using an Agilent Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, California, USA, catalogue number 5067-4626), and quantified on an Illumina Eco Real-Time PCR Instrument (Illumina, San Diego, California, USA) using KAPA Illumina Library Quantification Kits (KAPA Biosciences, Woburn, Massachusetts, USA, catalogue number KK4835) according to the manufacturers' protocols. Paired-end cluster generation (Illumina, San Diego, California, USA, catalogue number PE-401-3001) sequencing of 2 × 101 cycles (Illumina, San Diego, California, USA, catalogue number FC-401-3001) was performed for all libraries on an Illumina HiSeq 2000/2500 platform (Illumina, San Diego, California, USA), and samples were sequenced with the number of lanes predicted to yield an uncollapsed coverage of 50× and 30× for AML and T cells, respectively. Lane level reads were aligned to the reference human genome build hg19 using Novoalign (version 2.07.14, Novocraft), then filtered to remove unaligned reads, secondary alignment, and alignments with a mapping quality less than 30 (samtools version 0.1.18). Aligned sequences for each sample were merged into a single file after marking and removal of duplicates (Picard version 1.72). Single nucleotide variants and indels were called using a genome analysis tool kit (GATK version 1.3.16 Unified Genotyper). Variants with low quality (genotype quality < 60) based on GATK output were filtered out.

Definitions of mutations in the PM AML cohort. Somatic mutations were defined as non-synonymous variants detected by WGS with a VAF of <0.4 in T cells at diagnosis and >0.2 difference in VAF between bulk leukaemia cells and T cells at diagnosis. To exclude germline variants, all single nucleotide variants and indels identified in T cells were searched in the University of California, Santa Cruz genome browser under the category 'ALL SNP142'. Any variant previously identified in any population with a reference mean allele frequency >0.01 was considered germline and was excluded. PDVs were defined as somatic variants (both single nucleotide variants and short indels) located in exons and predicted to be both damaging and deleterious to the protein by the PROVEAN algorithms. PDVs were defined as preL-PDV if they were present in T cells sorted from diagnosis, relapse, or remission patient samples, or in B cells sorted from xenografts, or L-PDV if absent in these populations. To estimate the false positive designation of germline variants as preL-PDVs, we assessed the VAF of several PDVs in all sorted populations (HSPCs and mature cells, human cells isolated from xenografts). Variants that were present in all isolated populations were considered germline and excluded. Across all samples in the cohort, 5% of preL-PDVs were excluded because of these criteria, suggesting that a possible false positive rate of 5% might occur owing to the lack of other germline control. Relapse variants were defined as somatic variants with VAF >20% in the bulk blasts at relapse and <5% in the diagnosis blasts.

Fluorescence-activated cell sorting of human stem/progenitor and mature cell populations. Mononuclear cells (1×10^6 per 100 μl) from peripheral blood or bone marrow of patients with AML were stained with the following antibodies (all from BD unless stated otherwise, catalogue number in parentheses): anti-CD45RA-FITC (555488), anti-CD90-APC (561971), anti-CD135-Biotin (624008), anti-CD38-PE-Cy7 (335790), anti-CD10-Alexa-700 (624040), anti-CD7-Pacific Blue (642916), anti-CD45-V500 (560777), anti-CD34-APC-Cy7 (custom made by BD, CD34 clone 581), anti-CD34-PerCP-Efluor 710 (eBioscience 46-0344-42), anti-CD33-PC5 (Beckman Coulter PNIM2647U), anti-CD19-PE (349204), anti-CD3-FITC (349201), anti-CD56-Alexafluor 647 (557711), and Streptavidin-QD605 (Invitrogen Q10101MP). Remission samples from patients 1, 10, and 11 were enriched for $CD34^+$ cells using a Miltenyi CD34 MicroBead kit according to the manufacturer's protocol before antibody staining. For all samples, cells were sorted on a FACS AriaIII to a post-sort purity of >95%: progenitor populations were all gated on $CD45^+CD33^-$ and sorted into HSC/MPP ($CD38^-CD34^+CD45RA^-$); MLP ($CD38^-CD34^+CD45RA^+$); megakaryocyte erythroid progenitors/common myeloid progenitors ($CD38^+CD34^+CD7^-CD10^-CD45RA^-$); and granulocyte monocyte progenitor ($CD38^+CD34^+CD7^-CD10^-CD45RA^+$) subsets. Mature populations were sorted into leukaemia blasts ($CD45^{dim}CD33^+$), T cells ($CD45^{high}CD3^+$), B cells ($CD45^{high}CD19^+$), and NK cells ($CD45^{high}CD56^+$). For patients 3, 6, and 12, the following additional markers were required to purify mature populations (all from BD, catalogue number in parentheses): anti-CD8-APC-H7 (560179) to sort T cells, anti-light-chain lambda-V450 (561379), and anti-light-chain kappa-V450 (561327) to sort B cells, and anti-CD57-APC (555518) to sort NK cells. DNA from all sorted populations other than leukaemic blasts was amplified by whole-genome amplification (REPLI-g Mini Kit for 16 h).

Xenotransplantation assays. Animal experiments were performed in accordance with institutional guidelines approved by the University Health Network Animal Care Committee. Eight- to 12-week-old female NSG mice[27] were sublethally irradiated (225 cGy) 6–24 h before transplantation. For patients 2, 8, and 15, AML cells were also transplanted into 8- to 12-week-old female NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl} Tg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ (NSG-SGM3) mice[17] that were sublethally irradiated (225 cGy) 6–24 h before transplantation. Variable cell numbers were transplanted into cohorts of mice using a limiting dilution analysis to ensure recipients were engrafted with single LSCs. The total number of mice transplanted per sample depended on the number of viable cells that were available, but ranged from 20 to 80 recipients across all samples. There was no randomization or blinding. Mononuclear cells from patients with AML were depleted of $CD3^+$ cells by EasySep (Stem Cell Technologies) before intrafemoral transplantation as previously described[28]. Mice were killed 8 or 16 weeks after transplantation and human engraftment in the injected right femur and non-injected bone marrow (left femur, tibias) was evaluated by flow cytometry using the following human-specific antibodies: anti-CD45-APC (BD Biosciences catalogue number 340943), anti-CD19-PE (BD Biosciences catalogue number 340364), anti-CD33-PE-Cy5 (Beckman Coulter catalogue number IM2647U), anti-CD3-FITC (BD Biosciences catalogue number 349201), anti-CD14-ECD (Beckman Coulter catalogue number IM2707U), anti-CD15-Pacific Blue (BD Biosciences catalogue number 642917), anti-CD38-PE-Cy7 (BD Biosciences catalogue number 335790), and anti-CD34-APC-Cy7 (BD Biosciences, custom order). The threshold for detection of engraftment was 0.1% human $CD45^+$ cells. All flow cytometric analysis was done on an LSR II flow cytometer (BD Biosciences). Xenografts were classified as leukaemic if >90% of the human $CD45^+$ cells were $CD33^+CD19^-$ or multilineage if >20% of the human $CD45^+$ cells were $CD19^+CD33^-$. Of the 11 diagnosis samples, 5 generated a myeloid leukaemia graft, 4 generated a multilineage graft, and 2 did not generate any graft. All of the patient samples obtained at relapse generated a leukaemic graft (Supplementary Table 7).

Limiting dilution assays. To determine the frequency of LSCs in a patient's sample, bulk mononuclear cells were transplanted into NSG mice at doses ranging from 6×10^6 to 100 cells per mouse. The proportion of engrafted mice was determined at each dose by flow cytometry and the frequency of repopulating cells was calculated using ELDA software[29].

ddPCR. For each patient, at least five PDVs and relapse variants identified by WGS were validated by ddPCR using probes designed for each variant. Amplified DNA (2 μl from a 1:20 dilution of a 16 h REPLI-g Mini Kit whole-genome amplification, Qiagen) from each sorted population was tested in a 96-well plate in

duplicate according to the manufacturer's protocol. Mutant and wild-type sequences were read using a droplet reader with a two-colour fluorescein/HEX fluorescence detector (Bio-Rad). The mutant allele frequency was calculated as the fraction of mutant-positive droplets divided by total droplets containing a target. The VAFs for all relapse variants and PDVs in each population from each patient are listed in Supplementary Table 8. 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. As reported in our previous study[16], the minimum detection level was 1:1,000 (0.1%). Variants were considered present if there were at least three dots in the mutant fluorescein channel resulting in VAF >0.1%.

Mutation order and phylogenetic analysis by median joining algorithm. To estimate the order in which PDVs were acquired during leukaemogenesis and the phylogenetic relationships between the subpopulations within each patient, we used a population genetics tool called median joining[30]. In brief, this tool can extract the phylogeny structure in closely related strata that have short genetic distances between them. Owing to small genetic differences, phylogenetic trees are not accurate in phylogeny reconstruction and a network display is a useful alternative. This network depicts evolutionary paths in the form of cycles, which present possible options for the genetic interactions and order of mutations[30]. To be able to apply median joining to our data, we had to relax the VAF information to a binary description of whether the variant was present or not in each subpopulation. PDV data from the following subpopulations at diagnosis and relapse were used in the analysis: blasts, T cells, HSCs/MPPs, MLPs, common myeloid progenitors/megakaryocyte erythroid progenitors, granulocyte monocyte progenitors, and xenografts. For each patient, a root was added representing the reference genome (no variants). The binary data for each subpopulation were uploaded to the Network software as an RDF file (Network, Fluxus Technology). Software can be downloaded at http://www.fluxus-engineering.com/sharenet.htm. Once a median joining network was reconstructed, the actual mutation order was estimated by choosing the path from the root onwards that was supported by the maximum number of subpopulations. For example, if mutations in *DNMT3A* and *NPM1* appeared on the network in two different positions and more subpopulations supported the occurrence of *DNMT3A* mutation before *NPM1* mutation, then the estimated order would be *DNMT3A* mutation followed by *NPM1* mutation.

Phylogenetic tree analysis of population relatedness by Nei genetic distance and neighbour joining. To complement the simplified genotypes used for the median joining algorithm, we used the R StAMPP package to calculate Nei's genetic distance between all subpopulations for each patient. Nei's genetic distance assumes that the input VAF is of a population (of cells in our case). All variants and subpopulations that were genotyped by ddPCR were used in the Nei's genetic distance matrix. To reconstruct the phylogeny, we used a neighbour joining algorithm via the MATLAB statistical tool box.

RNA sequencing and data processing. RNA was extracted from bulk peripheral blood mononuclear cells using an RNeasy Micro Kit (Qiagen). Libraries were constructed using SMART-Seq (Clonetec). A paired-end 50-base-pair flow-cell lane Illumina HiSeq 2000 yielded an average of 240 million sequence reads aligning to genome per sample. All data were assessed for quality before downstream analyses. The data quality was based on metrics determined by the sequences aligned to the reference genome (version Hg19/GRCh37), the coverage across sequenced regions defined by EnsEMBL gene models (version 59), and the use of virtual SAGE mapping to efficiently identify the presence of non-biological or contaminant sequences in libraries. Genes with very low counts were filtered out, and the data were then normalized using edgeR TMM (trimmed mean of M values) normalization. Genes with counts per million (CPM) >0.45 in at least two samples were retained.

Hierarchical clustering. The diagnosis samples were hierarchically clustered using the ward linkage method and Euclidean distance on the basis of the scaled $\log_2$(CPM) of the 2,188 genes with s.d. >3.5. AU (approximately unbiased) cluster probability, computed by multiscale bootstrap resampling, was calculated using the R package pvclust. RNA sequencing read count data from TCGA patients who relapsed ($n=84$) were also used for hierarchical clustering using the same method as described above. Differential gene expression between the RO_P and RO_C clusters was calculated using edgeR. Genes were ranked from top upregulated to top downregulated (RO_C versus RO_P) and this ranked list was compared using gene set enrichment analysis with the TCGA top 150 upregulated (RO_C-like TCGA) and top 150 downregulated (RO_P-like TCGA) genes.

Comparison of clusters with DMAP populations. The Gene Expression Omnibus dataset GSE24759 (DMAP)[31], containing Affymetrix GeneChip HT-HG_U133A Early Access Array gene expression data of 21 distinct haematopoietic cell states, was compared with the genes that were differentially expressed in the RO_P-like and RO_C-like clusters. GSE24759 data were background corrected using Robust Multi-Array Average (RMA), quantile normalized using the expresso function of the affy Bioconductor package (affy_1.38.1, R 3.0.1), batch corrected using the ComBat function of the sva package (sva_3.6.0), and scaled using the standard score. One population was removed from the original dataset (erythroid $CD34^-CD71^+GlyA^+$, six samples) because it showed uniformly higher gene expression than all others after normalization and batch correction. Bar graphs were created by selecting genes that were upregulated in the RO_P-like and RO_C-like clusters, and calculating the number of scaled data that were above (>0) or below (<0) the mean for each population, corrected by the number of samples per population and 1,000 random permutations.

PERT deconvolution analysis. The PERT deconvolution method[25] was run on the TMM-normalized CPM data from the diagnosis and relapse samples. The batch-corrected linear RMA-normalized data from the GSE24759 (DMAP) data were used as the reference profile[31]. The vector theta from the PERT output was used to estimate the percentage of reference populations within each diagnosis and relapse sample. In the non-aggregated form, each biological replicate of the DMAP populations was associated with a theta value. In the aggregated form, the theta percentage of each biological replicate of the same populations was summed. A high theta value for one population means a high estimated presence of the profile in the diagnosis or relapse sample: for example, a theta value of 0.8 for a myeloid population for 6Dx means that this population is estimated to represent 80% of this diagnosis sample.

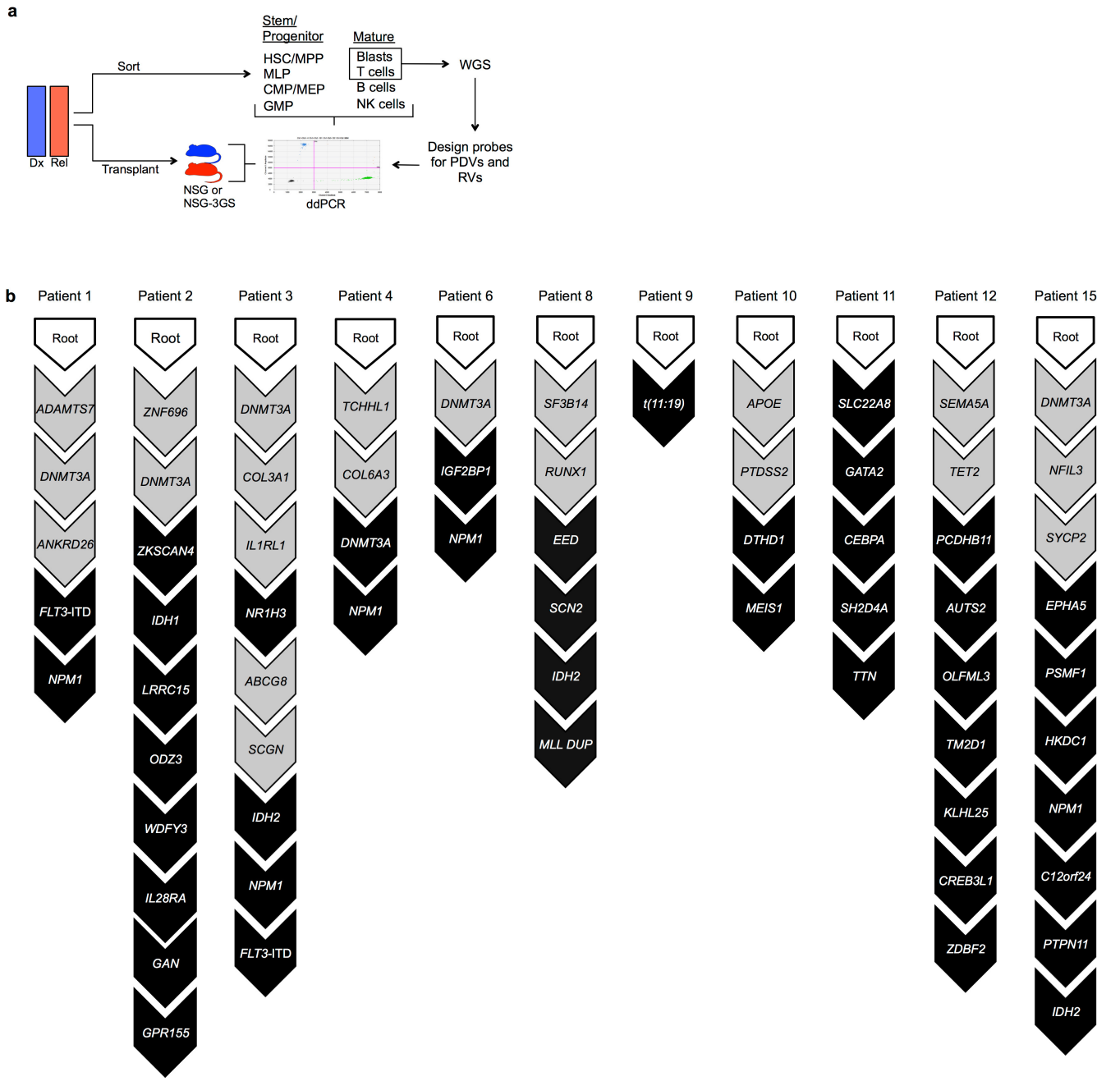
Therapy resistance score. A gene expression score that is predictive of refractoriness to primary induction chemotherapy was recently reported[15]. This score is calculated on the basis of a weighted sum of expression of six LSC genes functionally defined using xenograft transplantation assays (*MMRN1*, *KIAA0125* (also known as *FAM30A*), *CD34*, *GPR56* (also known as *ADGRG1*), *LAPTM4B*, *NYNRIN*). A high score predicts therapy resistance and a lower likelihood of achieving a complete remission after 7 + 3-based induction chemotherapy.

$$\text{Response score} = -6.58 + (MMRN1 \times 0.0442) + (KIAA0125 \times 0.0814) + (CD34 \times 0.104) + (GPR56 \times 0.208) + (LAPTM4B \times 0.168) + (NYNRIN \times 0.121).$$

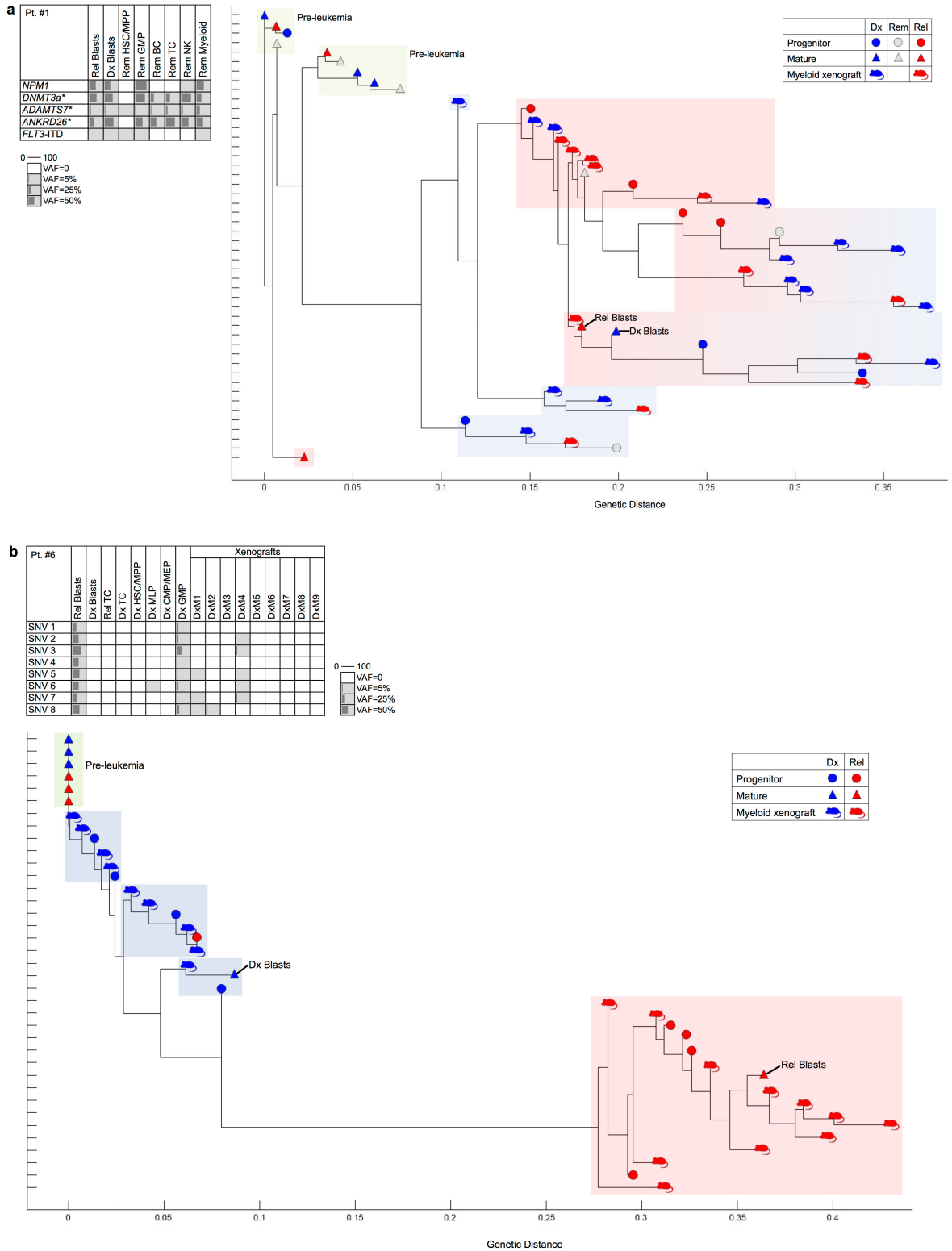
Differential gene expression between diagnosis and relapse. Gene expression between diagnosis and relapse samples was compared using two different methods. CPM-normalized data were compared using the edgeR likelihood ratio test. The paired design matrix was used to reduce the pair effects of diagnosis and relapse. False discovery rate-corrected *P* values were calculated for each gene. RNA sequencing data were normalized to produce reads per kilobase per million (RPKM) values, with the following noise filter thresholds: RPKM cutoff 0.005, read counts cutoff 25. RNA sequencing data were discrete count data and met the assumptions for using a negative binomial distribution included in the EdgeR model. The estimate of variation within each group was included in the edgeR model and found to be similar between the groups being analysed. Differential gene expression based on RPKM values was also analysed using DEfine version 0.9.1. For differential gene expression between diagnosis and relapse, statistical significance was achieved with a false discovery rate *P* value <0.05 in the CPM-normalized data and a *P* value score >2 in the RPKM-normalized data. Several genes had significantly higher expression at relapse, including *CD34*, *EDA2R*, *KCNK17*, *NPDC1*, *NTRK1*, and *RXFP1* (Supplementary Table 9). Interestingly, *EDA2R*, a transmembrane receptor protein, has been reported to be associated with resistance to Ara-C[32], one of the cornerstones of standard AML induction chemotherapy.

Data availability. WGS data have been deposited at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession number EGAS00001002225. All other data are available from the corresponding author upon reasonable request.

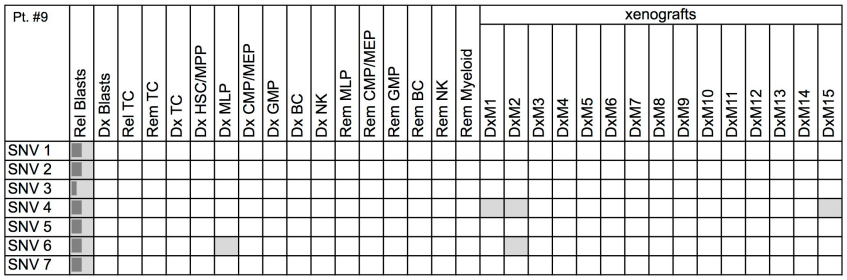
27. Notta, F., Doulatov, S. & Dick, J. E. Engraftment of human hematopoietic stem cells is more efficient in female NOD/SCID/IL-2Rgc-null recipients. *Blood* **115,** 3704–3707 (2010).
28. Mazurier, F., Doedens, M., Gan, O. I. & Dick, J. E. Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. *Nat. Med.* **9,** 959–963 (2003).
29. 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).
30. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16,** 37–48 (1999).
31. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144,** 296–309 (2011).
32. Gamazon, E. R. *et al.* Comprehensive genetic analysis of cytarabine sensitivity in a cell-based model identifies polymorphisms associated with outcome in AML patients. *Blood* **121,** 4366–4376 (2013).

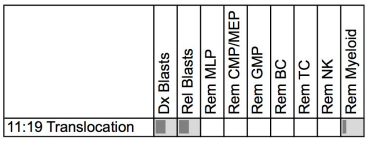


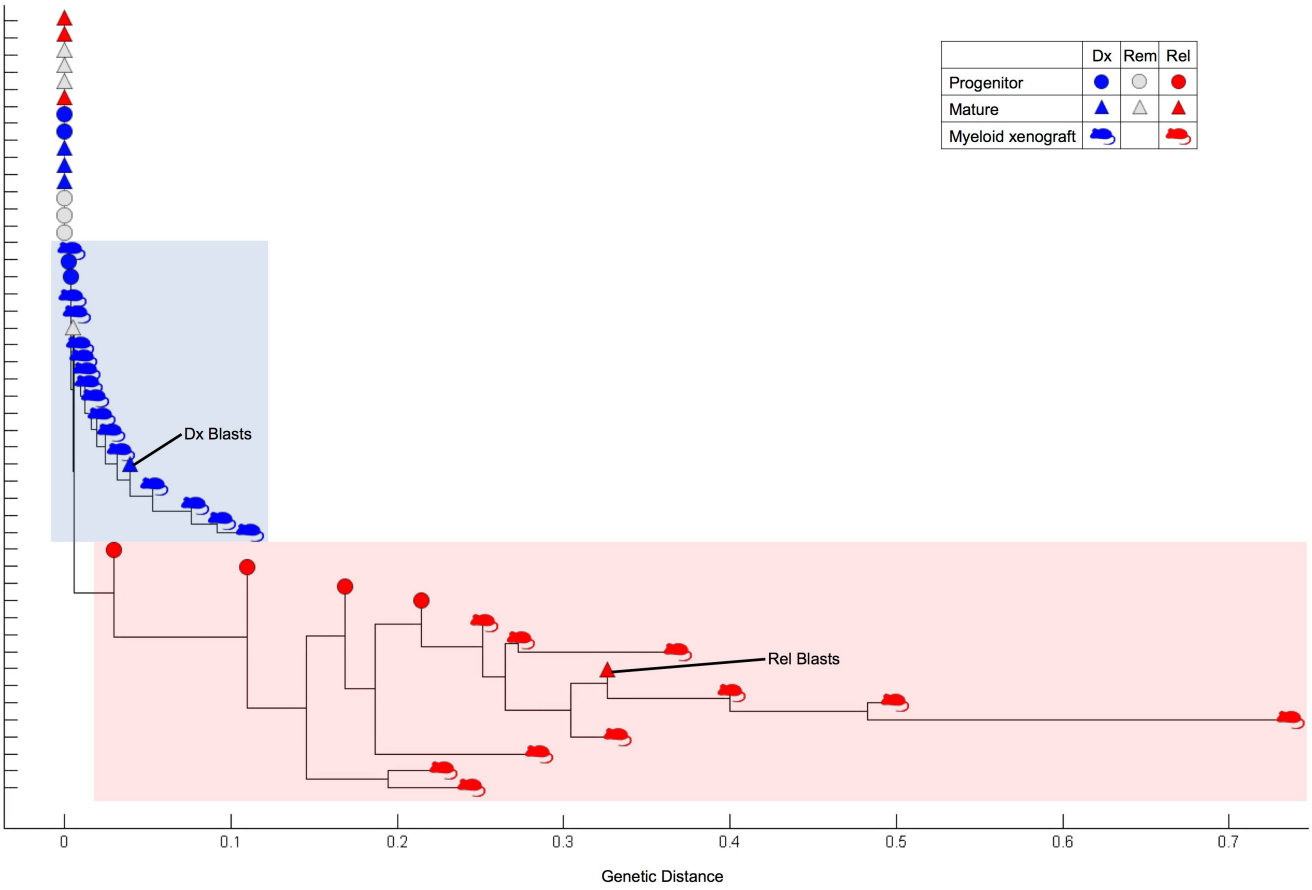
Extended Data Figure 1 | Experimental design and estimated order of mutation acquisition for patients in PM AML cohort. a, Experimental design. Patient samples obtained at diagnosis and relapse were sorted into multiple mature and progenitor populations by flow cytometry. In parallel, samples were injected into NSG or NSG-SGM3 mice, and human myeloid and T cells were sorted from the xenografts. WGS was performed on blasts and T cells, and ddPCR probes were designed for selected variants. These were then analysed by ddPCR in all the cell populations sorted from patient samples and xenografts for calculation of VAF. **b**, Estimated order of mutation acquisition for patients in PM AML cohort. Estimated order of acquisition of pre-leukaemic (grey) and leukaemic (black) PDVs, based on median joining network analysis.



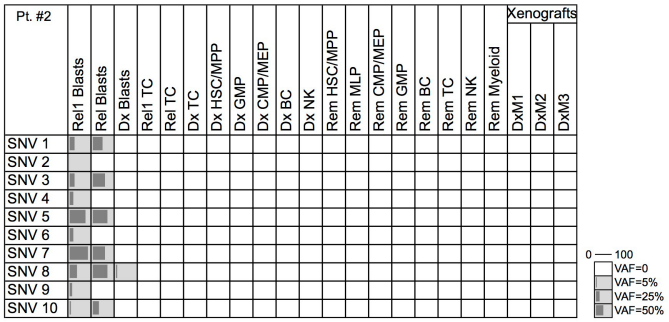
Extended Data Figure 2 | Analysis of relapse origin and clonal relationships for patients 1 and 6. a, VAFs of PDVs as detected by ddPCR in various cell populations sorted directly from the diagnosis (Dx), relapse (Rel), or remission (Rem) samples of patient 1. No relapse variants were detected by ddPCR in this patient. White squares indicate populations in which no variants were detected. Phylogenetic tree showing clonal relationships for patient 1, based on analysis of VAFs of all pre-leukaemic and leukaemic mutations. The genetic distance between any two respective symbols on the phylogenetic tree was estimated using a neighbour joining method based on the degree of genetic relatedness (Nei's genetic distance). Each coloured box contains highly related cell populations and xenografts inferred to represent a clone (green, pre-leukaemic; blue, diagnostic; red, relapse; blended blue and red, equal representation of diagnostic and relapse populations). **b**, VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic or relapse samples of patient 6, or in CD33^{+} myeloid cells isolated from xenografts generated from the diagnostic sample of the same patient. Abbreviations and display criteria are the same as in **a**. Phylogenetic tree showing clonal relationships for patient 6. The display criteria are the same as for **a**.

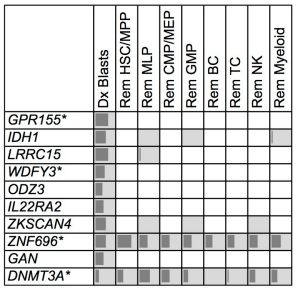


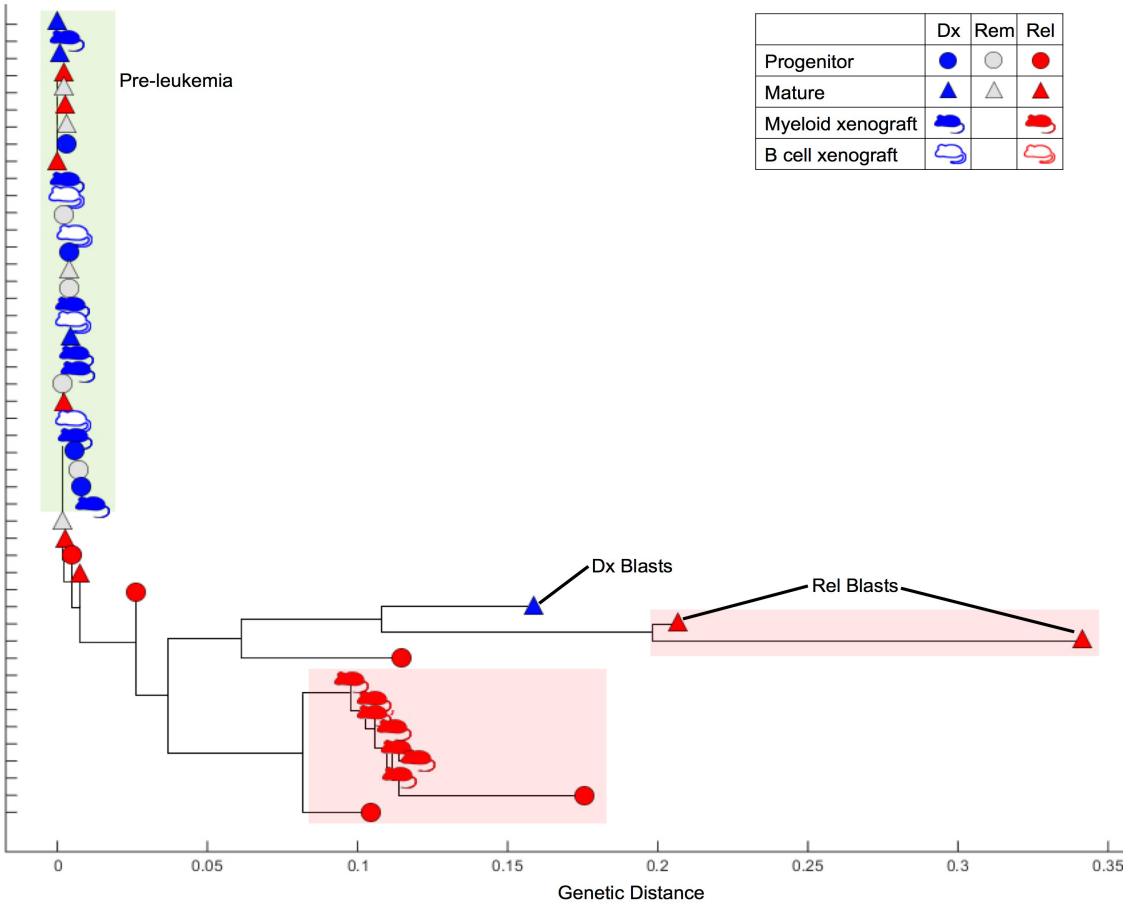




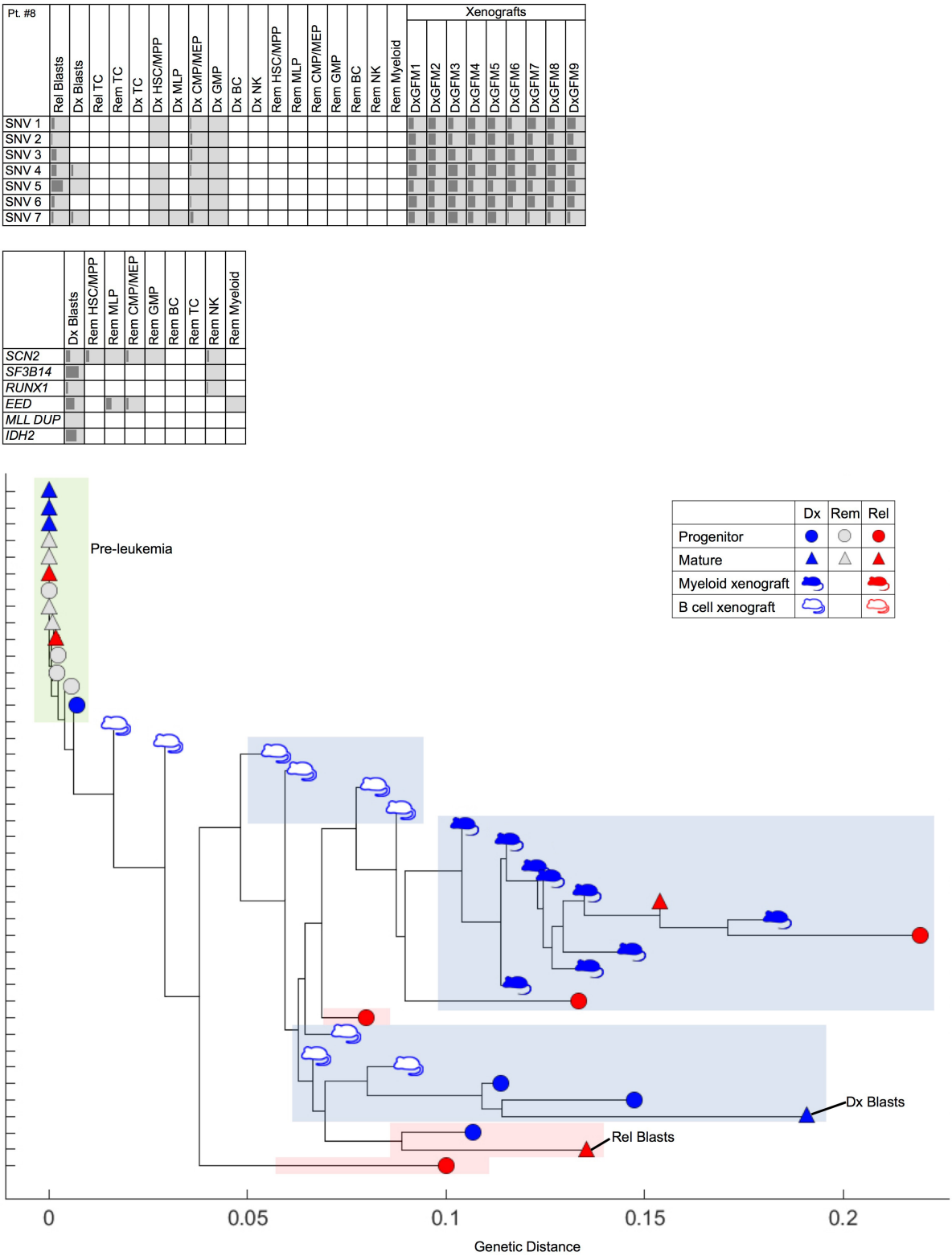
Extended Data Figure 3 | Analysis of relapse origin and clonal relationships for patient 9. VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 9, or in CD33^{+} myeloid cells isolated from xenografts generated from the diagnostic sample of the same patient. Abbreviations and display criteria are the same as in Extended Data Fig. 2a. VAFs of the only PDV detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 9. Phylogenetic tree showing clonal relationships for patient 9. The display criteria are the same as for Extended Data Fig. 2a.



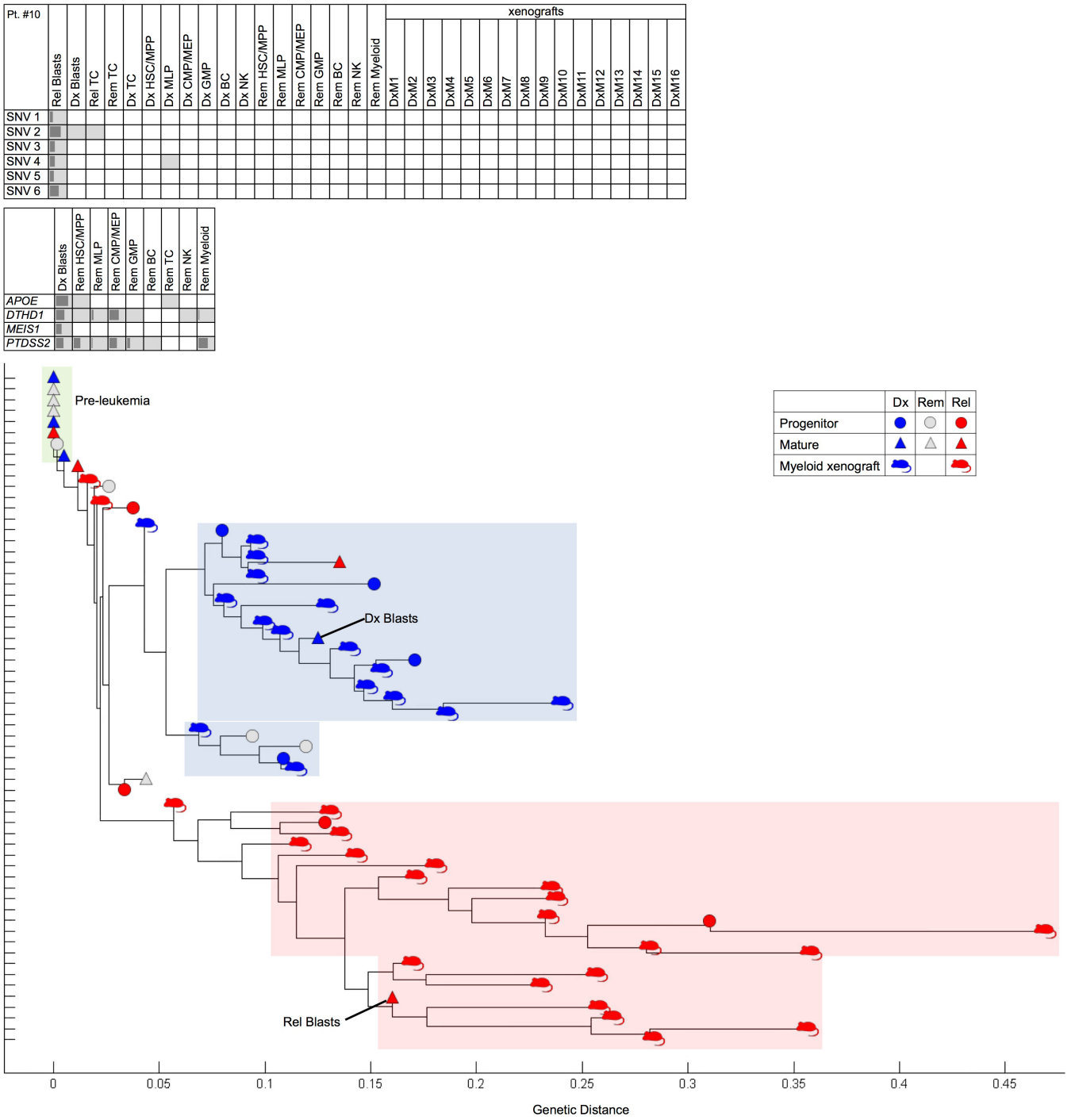




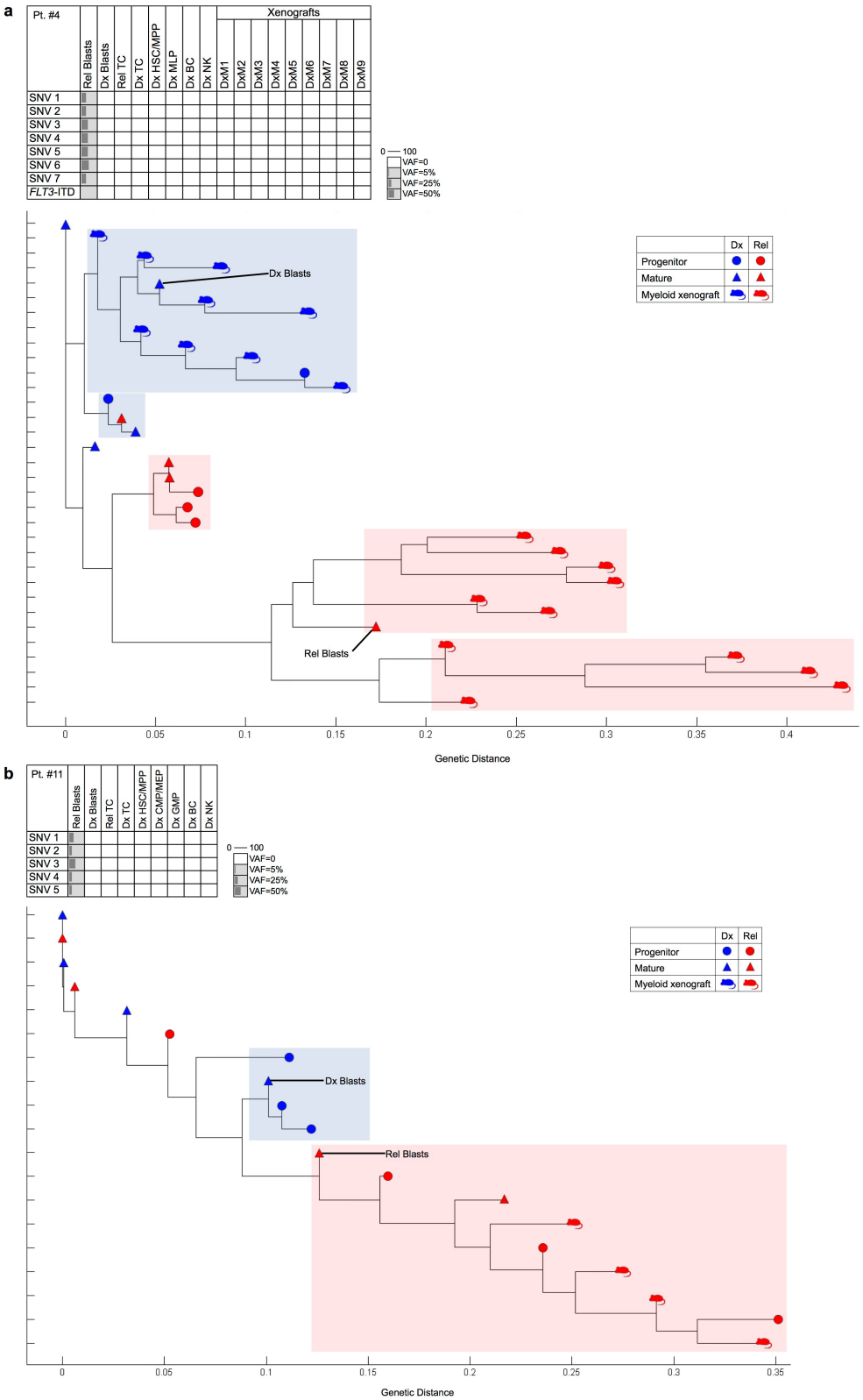
Extended Data Figure 4 | Analysis of relapse origin and clonal relationships for patient 2. VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 2, or in CD33^{+} myeloid cells isolated from xenografts generated from the diagnostic sample of the same patient. Abbreviations and display criteria as in Extended Data Fig. 2a. VAFs of PDVs as detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 2. Phylogenetic tree showing clonal relationships for patient 2, based on analysis of VAFs of all relapse variants, pre-leukaemic, and leukaemic mutations. The display criteria are the same as for Extended Data Fig. 2a.



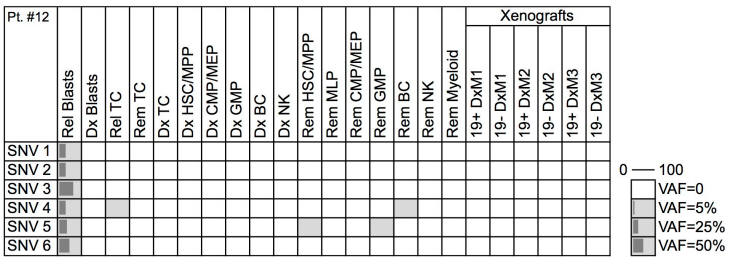
Extended Data Figure 5 | Analysis of relapse origin and clonal relationships for patient 8. VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 8, or in CD19$^+$ B or CD33$^+$ myeloid cells isolated from xenografts (DxGFM) generated from the diagnostic sample of the same patient. Abbreviations and display criteria as in Extended Data Fig. 2a. VAFs of PDVs as detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 8. Phylogenetic tree showing clonal relationships for patient 8. The display criteria are the same as for Extended Data Fig. 2a.

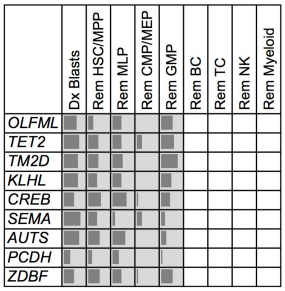


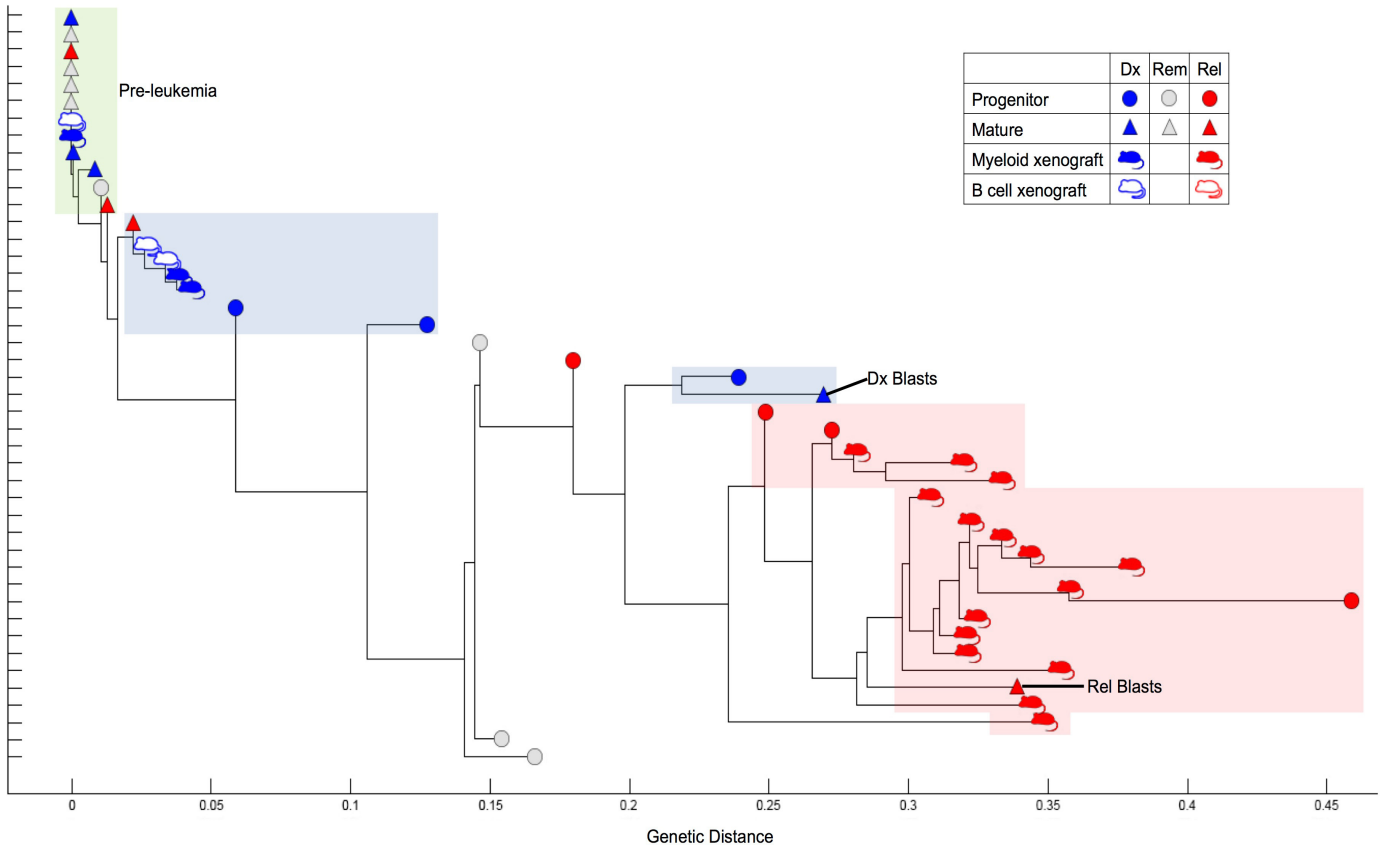
Extended Data Figure 6 | Analysis of relapse origin and clonal relationships for patient 10. VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 10, or in CD33^{+} myeloid cells isolated from xenografts generated from the diagnostic sample of the same patient. Abbreviations and display criteria are the same as in Extended Data Fig. 2a. VAFs of PDVs as detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 10. Phylogenetic tree showing clonal relationships for patient 10. The display criteria are the same as for Extended Data Fig. 2a.



Extended Data Figure 7 | Analysis of relapse origin and clonal relationships for patients 4 and 11. a, VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 4, or in CD33$^+$ myeloid cells isolated from xenografts (DxM) generated from the diagnostic sample of the same patient. Abbreviations and display criteria are the same as in Extended Data Fig. 2a. Phylogenetic tree showing clonal relationships for patient 4. The display criteria are the same as for Extended Data Fig. 2a. **b**, VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic or relapse samples of patient 11. Abbreviations and display criteria are the same as in Extended Data Fig. 2a. Phylogenetic tree showing clonal relationships for patient 11. The display criteria are the same as for Extended Data Fig. 2a.

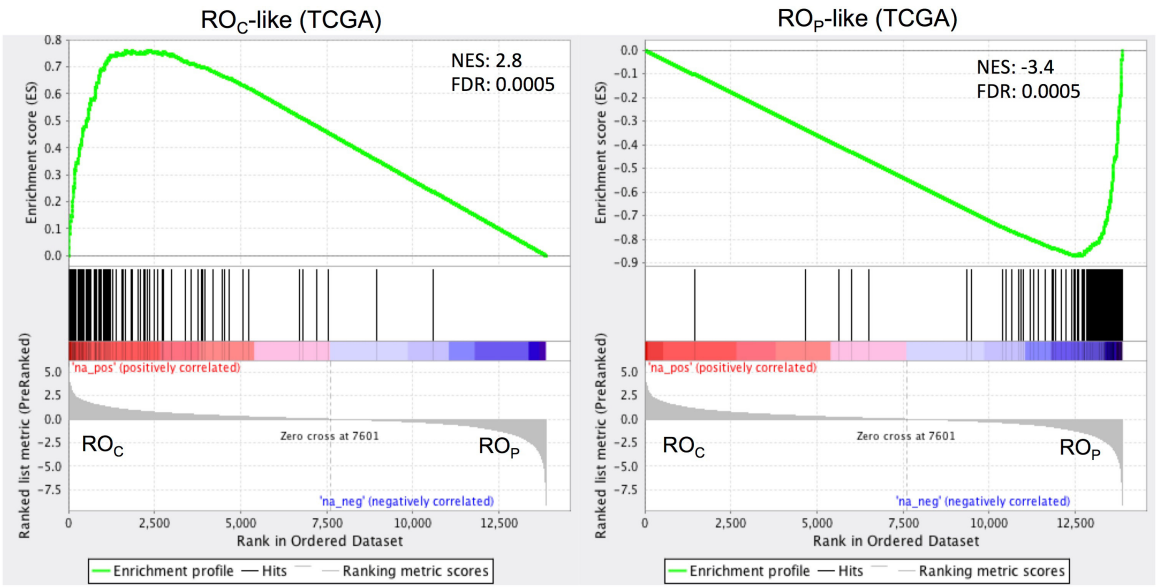






Extended Data Figure 8 | Analysis of relapse origin and clonal relationships for patient 12. VAFs of relapse variants as detected by ddPCR in various cell populations sorted directly from the diagnostic, remission, or relapse samples of patient 12, or in CD19$^+$ B or CD33$^+$ myeloid cells isolated from xenografts generated from the diagnostic sample of the same patient. Abbreviations and display criteria are the same as in Extended Data Fig. 2a. VAFs of the PDVs as detected by ddPCR in diagnostic blasts and in various cell populations sorted directly from the remission sample of patient 12. Phylogenetic tree showing clonal relationships for patient 12. The display criteria are the same as for Extended Data Fig. 2a.

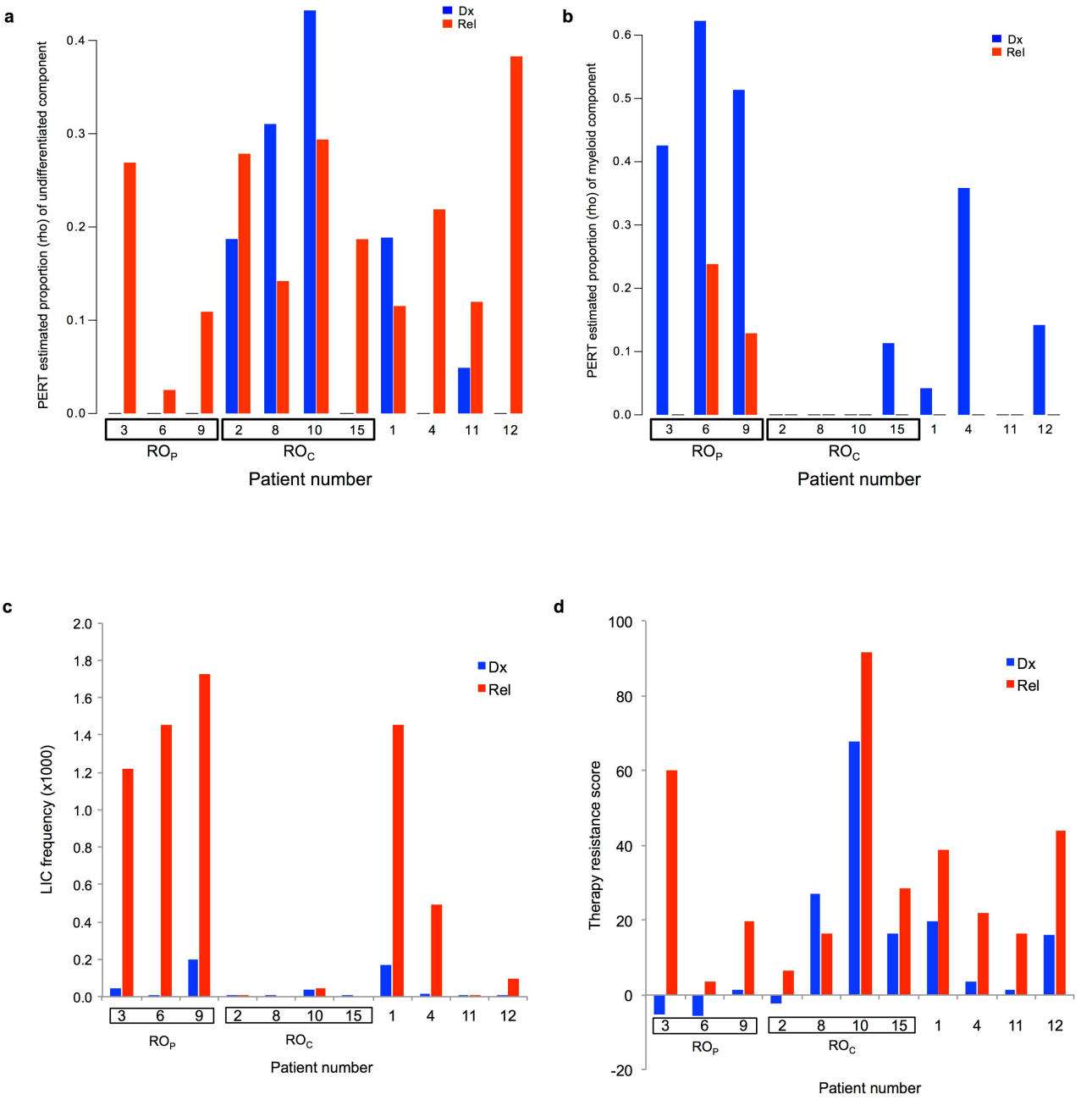
a



b

	ROP-like Patients (TCGA)	ROC-like Patients (TCGA)	P value
Total number or patients	26.0	58.0	
FAB M0/1/2	23.1	74.1	0.000026
FAB M4/M5	76.9	12.1	9.6E-09
Age	58.3	50.4	0.028
% PB Blasts	31.8	51.1	0.009
WBC	54.6	33.9	0.044
Complex Cytogenetics	0.0	13.8	0.054
Intermediate Risk Cytogenetics	76.9	15.5	<0.0001
Inv(16)	19.2	0.0	0.0021
NPM1 mutation	50.0	22.4	0.02
DNMT3A mutation	42.3	22.4	0.073

Extended Data Figure 9 | ROP and ROC correlate with TCGA gene expression clusters and demonstrate distinct molecular and clinical properties. a, Gene set enrichment analysis of ROP and ROC clusters in the Princess Margaret Cancer Centre AML and TCGA relapsed AML cohorts. Diagnosis genes were ranked from top upregulated to top downregulated (ROC versus ROP cluster), and this rank list was compared using gene set enrichment analysis with the TCGA top 150 upregulated and top 150 downregulated genes (TCGA ROC-like versus TCGA ROP-like cluster). ROC genes correlate with TCGA ROC-like genes and ROP genes correlate with TCGA ROP-like genes. **b**, Comparison of genetic and clinical parameters between the TCGA ROC-like and TCGA ROP-like clusters identified distinct genomic and clinical properties associated with relapse origin.



Extended Data Figure 10 | AML relapse is associated with increased stemness transcriptional programs and therapy resistance. **a**, **b**, Proportion of bulk leukaemia populations that possess gene expression profiles of undifferentiated (**a**) or mature myeloid (**b**) cells, as determined by the perturbation (PERT) deconvolution analysis using gene expression data from known normal haematopoietic cell subsets. **c**, Leukaemia-initiating cell (LIC) frequency as determined by limiting dilution analysis. **d**, Gene-expression-based therapy resistance score. For all panels, data are shown for diagnostic (blue) and relapse (red) samples of patients in the PM cohort. Patients for whom the relapse origin (RO_P or RO_C) was functionally established are boxed.