



Published in final edited form as:

Nature. 2018 October ; 562(7727): 373–379. doi:10.1038/s41586-018-0436-0.

The genetic basis and cell of origin of mixed phenotype acute leukaemia

A full list of authors and affiliations appears at the end of the article.

Abstract

Mixed phenotype acute leukaemia (MPAL) is a high-risk subtype of leukaemia with myeloid and lymphoid features, limited genetic characterization, and a lack of consensus regarding appropriate therapy. Here we show that the two principal subtypes of MPAL, T/myeloid (T/M) and B/myeloid (B/M), are genetically distinct. Rearrangement of *ZNF384* is common in B/M MPAL, and biallelic *WT1* alterations are common in T/M MPAL, which shares genomic features with early T-cell precursor acute lymphoblastic leukaemia. We show that the intratumoral immunophenotypic heterogeneity characteristic of MPAL is independent of somatic genetic variation, that founding lesions arise in primitive haematopoietic progenitors, and that individual phenotypic subpopulations can reconstitute the immunophenotypic diversity in vivo. These findings indicate that the cell of origin and founding lesions, rather than an accumulation of distinct genomic alterations, prime tumour cells for lineage promiscuity. Moreover, these findings position MPAL in the spectrum of immature leukaemias and provide a genetically informed framework for future clinical trials of potential treatments for MPAL.

Acute leukaemia of ambiguous lineage (ALAL) comprises a collection of high-risk leukaemias defined by immunophenotype, including MPAL and acute undifferentiated leukaemia (AUL). MPAL demonstrates features of acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML), while AUL lacks lineage-defining features. MPAL

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.M. or H.I. charles.mullighan@stjude.org; hiroto.inaba@stjude.org.

Author contributions T.B.A.: study design, flow analysis and sorting, data analysis, and manuscript writing. Z.G.: genomic data analysis. I.I.: genomic and mouse experiments, data analysis, data interpretation and manuscript preparation. K.D.: *ZNF384* modeling. J.K.C.: central review of immunophenotype. B.X.: ChIP-seq and RNA-seq data analysis. D.P.-T and H.Y.: performed experiments. M.L.L. and S.P.H.: led and contributed to Children's Oncology Group ALL studies and the ALL TARGET project. M.B. and B.W.: reviewed flow cytometry. M.D., N.A.H., and A.C.: provided clinical data. J.H., E.O., B.B., G.B., S.E., V.d.H., C.M.Z., A.Y., D.R., D.T., N.K., T.L., B.D.M., D.C., H.H., A.M., A.S.M., O.H., K.E.N., J.R.D., and J.Z.: patient samples and clinical data. S.M.: data for comparison cohort. Y.-L.Y.: flow analysis. M.A.S., T.M.D., L.C.H., P.G., M.A.M., Y.M., A.J.M., R.A.M., S.J.M.J., and J.M.G.A.: genomic sequencing, analysis, and support. M.V.: performed FISH. L.J.J.: necropsy and histology on xenograft models. J.E.R. and C.-H.P.: patient samples and clinical data. D.S.G.: support for genomic analysis and manuscript editing. L.D. and Y.L.: genomic analysis. X.C., L.S., S.P. and D.P.: statistical analysis. S.N.: somatic and germline variant analysis. H.I.: acquisition of patient samples and clinical data. C.G.M.: designed and oversaw the study, analysed data and wrote the manuscript.

Online content Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at

Reviewer information Nature thanks R. Levine and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Competing interests: The authors declare no competing interests.

Publisher's Disclaimer: Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

represents 2–3% of cases of childhood acute leukaemia, whereas AUL is rare^{1,2}. Survival rates for children and adults with MPAL are 47–75% and 20–40%, respectively, and there is no consensus regarding the optimal (AML- or ALL-directed) therapeutic regimen^{1–3}. Up to 15% of patients with MPAL have rearrangements of *KMT2A* (also known as *MLL*; rearrangements referred to as *KMT2Ar*) or a *BCR-ABL1* fusion gene, but the genetic basis of most cases of MPAL remains unknown. As the lineage ‘aberrancy’ or ‘promiscuity’ of T/M MPAL shares features with early T-cell precursor (ETP) ALL^{4,5}, we sought to define the genetic basis of MPAL, to compare its genomic landscape to those of other leukaemia subtypes, and to determine the genetic basis of the intratumoral phenotypic heterogeneity that is characteristic of this disorder.

Genomic characterization of ALAL

We performed a central review of 159 potential paediatric cases of ALAL by repeating ($n = 138$) or reviewing flow cytometry data ($n = 21$); 115 fulfilled WHO (World Health Organization) criteria for the diagnosis of ALAL⁶ (Extended Data Fig. 1). There was a male predominance of ALAL (1.6:1), which was diagnosed at similar frequency throughout childhood, except for cases with *KMT2Ar*, which were common in infants (Supplementary Tables 1, 2). The cohort included 49 cases of T/M MPAL, 35 B/M MPAL, 16 *KMT2Ar* MPAL and 2 *BCR-ABL1* MPAL, 8 MPAL not otherwise specified (NOS), and 5 AUL. There was extensive immunophenotypic heterogeneity, with bilineal patterns (multiple immunophenotypic subpopulations), biphenotypic patterns (coexpression of lymphoid and myeloid antigens), or both (Extended Data Fig. 2a–g). There was no difference in five-year overall survival between T/M MPAL and B/M MPAL (56.7% \pm 10.8% (95% confidence interval) and 59.7% \pm 11.4%, respectively); outcome for patients with *KMT2Ar* was poor (five-year overall survival 21.2% \pm 10.8%) (Extended Data Fig. 2h–o).

Genomic alterations were examined by exome ($n = 92$), transcriptome ($n = 95$), and/or whole-genome ($n = 47$) sequencing, and single nucleotide polymorphism (SNP) array analysis ($n = 95$) (Supplementary Tables 3, 4). We identified 158 recurrently altered genes, of which 81 were mutated in at least three cases. Commonly mutated genes included those recurrent in AML, such as *FLT3* ($n = 31$), *RUNX1* ($n = 15$), *CUX1* ($n = 7$) and *CEBPA* ($n = 5$); those recurrent in ALL, including *CDKN2A* or *CDKN2B* ($n = 22$), *ETV6* ($n = 23$), and *VPREB1* ($n = 15$); and those recurrent in both AML and ALL, including *WT1* ($n = 28$) and *KMT2A* ($n = 26$) (Fig. 1a, Extended Data Figs. 3, 4 and Supplementary Tables 5–13). We analysed associations between genomic alterations and age at diagnosis, sex and disease subtype, and between pathway alterations and outcome (Supplementary Tables 14, 15 and Supplementary Note). We analysed germline samples for potential pathogenic variants in recurrently somatically mutated genes, and identified few putatively deleterious variants⁷ (Supplementary Table 16 and Supplementary Note).

Distinct profiles of MPAL subtypes

The three most common subtypes of MPAL (T/M, B/M and *KMT2Ar*) had distinct patterns of genomic alterations (Fig. 1a–c, Supplementary Table 13). As in infant ALL, *KMT2Ar* MPAL had a low mutation burden (median 1 (range 0–3) copy number alterations (CNAs)

and 4 (0–12) single nucleotide variants (SNVs) or insertions/deletions (indels) per case), whereas mutation burden was higher for T/M MPAL (4.5 (0–35) CNAs, 8 (2–29) SNVs or indels) and B/M MPAL (3.5 (0–29) CNAs, 9 (0–167) SNVs or indels) (Extended Data Fig. 3b). Alterations in genes encoding transcriptional regulators were detected in 100% of cases of T/M MPAL, with mutually exclusive alterations in *WT1*, *ETV6*, *RUNX1* and *CEBPA* in 82% of cases (Fig. 1b, Extended Data Fig. 5a, b); and in 94% of cases of B/M MPAL, with the B-lineage transcriptional regulators *PAX5* and *IKZF1* altered in 40% of cases (Fig. 1b).

Alterations in signalling pathways were observed in 88% of cases of T/M MPAL, 74% of cases of B/M MPAL and 63% of cases of *KMT2Ar* MPAL. Alterations in JAK–STAT signalling were more common in T/M MPAL (57%) than B/M MPAL (23%) or *KMT2Ar* MPAL (19%) (Fig. 1c), and we observed a negative correlation between alterations in *FLT3* (43%) and the Ras pathway (33%) in T/M MPAL ($P = 0.002$) (Fig. 1c, Supplementary Table 15). Ras pathway alterations were common in B/M MPAL (63%, most commonly *NRAS* and *PTPN11*). Genes encoding epigenetic regulators were mutated in 69% of cases of T/M MPAL, including inactivating mutations in *EZH2*⁵ (16%) and *PHF6* (16%), and in 63% of cases of B/M MPAL, most commonly in *MLLT3* (17%), *KDM6A* (in one-third of *ZNF384*-rearranged cases), *EP300* and *CREBBP* (Supplementary Table 13).

Transcriptome sequencing identified chimaeric in-frame fusions in 15 of 40 cases of T/M MPAL: *ZEB2–BCL11B* ($n = 3$), *ETV6–NCOA2* ($n = 2$), *ETV6–ARNT* ($n = 2$) and single cases of *ETV6–FOXO1*, *ETV6–MAML3*, *NUP214–ABL1*, *PICALM–MLLT10* and *PCM1–FGFR1* (Supplementary Tables 17–20). *KMT2Ar* MPAL had a B/M phenotype in 15 out of 16 cases and a T/M phenotype in one case, and involved *AFF1* (also known as *AF4*) in seven cases, *MLLT3* (also known as *AF9*) in three cases and *MLLT1* (also known as *ENL*) in two cases. *KMT2Ar* was also found in two of five cases of AUL.

ZNF384 rearrangement in leukaemia

Rearrangement of *ZNF384* (*ZNF384t*) was present in 48% of cases of B/M MPAL, involving *TCF3* ($n = 8$), *EP300* ($n = 5$), *TAF15* ($n = 1$) and *CREBBP* ($n = 1$) (Extended Data Fig. 5c). The chimaeric fusions involved the entire *ZNF384* coding region, loss of the C termini of the partner genes, and translation of both wild-type *ZNF384* and chimaeric fusion proteins. The mutational burden of *ZNF384t* B/M MPAL (median of 4 (1–29) CNAs and 8 (3–39) SNVs or indels) was similar to those of other MPAL subtypes (Extended Data Fig. 3b), with no variation in mutations between immunophenotypic subpopulations in ten cases examined (Extended Data Fig. 5d). *ZNF384t*, most commonly with *TCF3*, is also observed in B cell ALL (B-ALL), in which aberrant expression of myeloid markers that do not fulfil the diagnostic criteria for B/M MPAL is common⁸. The genomic landscape of childhood *ZNF384t* B-ALL ($n = 19$, Supplementary Tables 21, 22) was similar to that of *ZNF384t* MPAL with the exception of *KDM6A* alterations, which were observed only in *ZNF384t* MPAL (Fig. 2a). Analysis of a diverse range of acute leukaemias, including AML (Supplementary Tables 23, 24), showed that the gene expression profiles (GEPs) of *ZNF384t* B/M MPAL and B-ALL were indistinguishable (Fig. 2b, Extended Data Fig. 5e and Supplementary Table 25). Patients with *ZNF384t* exhibited higher *FLT3* expression those with other types of B/M or T/M MPAL (Extended Data Fig. 5f). Cases of B/M MPAL

that exhibited genomic features of other subtypes of B-ALL, such as hyperdiploidy or a Ph-like GEP, clustered with those subtypes of B-ALL (Fig. 2b). Gene set enrichment analysis suggested that *ZNF384* B/M MPAL was arrested at a more mature stage of development than other types of B/M MPAL (Extended Data Fig 6a, Supplementary Tables 26, 27). However, compared with B-ALL, *ZNF384* leukaemia showed enrichment of stem cell pathways and genes upregulated in ETP-ALL (Extended Data Fig. 6b, Supplementary Tables 27–29). Serial sampling of a case of *ZNF384* B/M MPAL showed acquisition of a focal heterozygous *IKZF1* deletion at first relapse, and a focal homozygous deletion of *CDKN2A* and *CDKN2B* at second relapse, with a shift from a myeloid to a lymphoid immunophenotype. Thus, *ZNF384* defines a distinct subtype of acute leukaemia with a variable immunophenotype ranging from B-ALL to B/M MPAL.

To further investigate the role of *ZNF384* rearrangement in leukemogenesis, we expressed haemagglutinin (HA)-tagged ZNF384, TAF15–ZNF384 and TCF3–ZNF384 in mouse *Arf*^{-/-} pre-B cells⁹ (Extended Data Fig. 6c). Chromatin immunoprecipitation with sequencing (ChIP–seq) identified 2,298 peaks with new or increased binding of the fusion proteins compared to wild-type ZNF384, and 495 peaks with reduced binding (Extended Data Fig. 6d). Gained or increased peaks contained the core ZNF384 binding motif, and were enriched at promoters of genes important for immune system development and transcriptional regulation (Supplementary Tables 30, 31). Increased promoter binding was associated with increased gene expression (Extended Data Fig. 6e and Supplementary Table 32), with similarity between the GEPs of mouse pre-B cells expressing ZNF384 fusions and human *ZNF384* leukaemia cells (Extended Data Fig. 6f and Supplementary Table 28). Thus, chimaeric ZNF384 oncoproteins exhibit perturbed binding and drive transcriptional deregulation in human *ZNF384* leukaemia.

The driver alterations and GEPs of non-*ZNF384* MPAL and AUL were heterogeneous (Supplementary Table 33). Three cases were Ph-like (*EBF1–PDGFRB*, *IGH–CRLF2* and a case lacking an identified kinase lesion), and two were hyperdiploid. Eight cases were *KMT2A*-like with *HOXA9* deregulation, and six of these had genetic alterations associated with *HOXA* overexpression: *MLLT10* rearrangement ($n = 2$), *SET–NUP214* ($n = 2$), *KMT2A* partial tandem duplication and *MNX1–ETV6* ($n = 1$ each).

Similarity between T/M MPAL and ETP-ALL

ETP-ALL exhibits aberrant expression of stem cell and myeloid markers (with the exception of myeloperoxidase, which would classify the disease as AML or MPAL)¹⁰. ETP-ALL is characterized by mutations in regulators of haematopoietic development, signalling, and chromatin remodelling, and a GEP suggesting the cell of origin to be a haematopoietic stem cell (HSC) or progenitor, rather than a T-cell precursor⁵. Because T/M MPAL and ETP-ALL are defined by a phenotype that includes lymphoid and myeloid features^{6,10}, we hypothesized that they might share molecular features. We compared the genomic features of T/M MPAL with those of childhood T cell ALL (T-ALL; $n = 245$)¹¹, ETP-ALL ($n = 19$)¹¹ and AML ($n = 197$)¹² (Supplementary Tables 34, 35). Transcription factor gene alterations were common in each but varied between subtypes (Extended Data Fig. 6g). The core transcription factors driving T-ALL (*TAL1*, *TAL2*, *TLX1*, *TLX3*, *LMO1*, *LMO2*,

NKX2-1, *HOXA10* and *LYL1*) were less frequently altered in T/M MPAL and ETP-ALL (63% versus 16% and 26%, respectively; $P < 0.001$). Alterations that deregulated *TALI*, which were present in 31% of cases of T-ALL, were never observed in T/M MPAL, including the 15 cases for which whole-genome sequencing (WGS) was examined for noncoding enhancer mutations¹³. Other alterations that are common in T-ALL, such as *MYB* amplification, *LEF1* deletion, *CDKN2A* and *CDKN2B* deletions, and amplification of the NOTCH1-driven *MYC* enhancer, were rare in T/M MPAL and ETP-ALL. By contrast, *WT1* alterations were common in T/M MPAL (41%) and ETP-ALL (42%), but not in T-ALL (9%; $P < 0.001$). Alterations of *CEBPA* and *CUX1* were common in T/M MPAL but not ETP-ALL or T-ALL. Conversely, *NOTCH1* mutations were uncommon in T/M MPAL and AML. Signalling pathway mutations were also associated with specific subtypes, with Ras and JAK–STAT pathway mutations being common in T/M MPAL and ETP-ALL, and phosphatidylinositol 3-kinase (PI3K) signalling pathway mutations being common in T-ALL (Extended Data Fig. 6h). Several genes were mutated at similar frequencies in T/M MPAL and ETP-ALL, including *ETV6*, *EZH2*, *WT1* and *FLT3* (Fig. 2c), and the GEPs of T/M MPAL and ETP-ALL were similar (Fig. 2b). Thus, T/M MPAL and ETP-ALL are similar entities in the spectrum of immature leukaemias.

Analysis of intratumoral variegation

Elucidating whether the intra-sample immunophenotypic heterogeneity is determined by genetic variegation or by genomic priming of a haematopoietic progenitor has important implications for therapy. Accordingly, we sequenced 2–4 subpopulations from 50 cases of MPAL (Supplementary Table 36). In 41 cases, the non-silent mutations were present in each separate population (Fig. 3a, b and Extended Data Fig. 7a). In nine cases, multiple mutations were detected in a single gene (*WT1* in five cases) with at least one of the mutations detected in all subpopulations in all cases. In two cases, the second mutation called from the same gene was not present in each subpopulation sequenced (*WHSC1* in T/M case SJMPAL016447 and *CREBBP* in T/M case SJMPAL017976). In five cases, a subpopulation-restricted mutation occurred in a signalling pathway, either as gain of function (*PTPN11*, *FLT3*) or loss of function (*NFI*, *CBL*) (Supplementary Table 36), consistent with previous studies of diagnosis and relapse pairs showing frequent subclonal signalling alterations¹⁴. By contrast, mutations in the most commonly altered transcription factor in T/M MPAL, *WT1*, were consistently present in the major clone in each case. These observations support the notion that transcription factor gene alterations arise early in leukemogenesis, and alterations that drive signalling alterations are secondary events.

Similarly, analysis of the DNA methylation profiles of 27 cases of MPAL (11 with multiple subpopulations), 74 non-MPAL leukaemias and 17 normal progenitor samples showed distinct methylation profiles between leukaemia subtypes, but not between MPAL subclones (Extended Data Fig 7b–e, Supplementary Table 37). Thus, cytosine methylation does not drive immunophenotypic heterogeneity in MPAL.

Phenotypic plasticity of MPAL

To further examine the basis of lineage plasticity in MPAL, we used xenograft models in which immunophenotypic subpopulations were purified and transplanted into immunocompromised NOD-SCID IL2R γ -null-3/GM/SF (NSG-SGM3) mice. Sorted subpopulations of cells from a patient with T/M MPAL (Fig. 3c, Extended Data Fig. 8a), the *ZNF384* B/M JIH-5 cell line¹⁵ (Extended Data Fig. 8b, c), and a patient with *KMT2Ar* MPAL (Extended Data Fig. 8d), when transplanted into multiple independent NSG-SGM3 mice, propagated the immunophenotypic diversity of the primary samples. Moreover, we observed a phenotype shift in a sample from a patient with T/M MPAL during passaging of the bulk tumour sample, with engraftment of either a B/M or T/M leukemia phenotype (Extended Data Fig. 8e–h). These data demonstrate the multilineage potential of phenotypic subpopulations in MPAL, and phenotypic evolution even in the absence of therapeutic pressure.

Collectively, our genomic data and in vivo lineage plasticity data suggest that intra-sample lineage diversification in MPAL is driven by constellations of genomic alterations acquired in a haematopoietic stem or progenitor cell with multilineage potential. To test this idea, we purified progenitor cell and blast populations and normal mature lymphocytes from samples from a patient with *ZNF384* B/M MPAL and two patients with *WT1*-altered T/M MPAL (Fig. 4a and Extended Data Fig. 9a, b). Alterations identified in the unfractionated samples (for example, *TCF3*–*ZNF384* and mutations in *MYCN*, *NTSD2* and *DNAH17* in the *ZNF384* sample) were identified in the purified blast populations but not in non-leukaemic T or natural killer (NK) cells. Each alteration was also present in multiple haematopoietic progenitor populations with myeloid and lymphoid potential, and a subset of HSCs (Fig. 4b and Extended Data Fig. 9c). Analogous results were detected in two cases of T/M MPAL with *WT1* alterations (data not shown); these contrast with Ph-like B-ALL, in which founding lesions are detectable in a primitive progenitor with the capacity for myelo-lymphoid differentiation, but not in HSCs¹⁶. These data support the notion that mutations are acquired in a haematopoietic stem cell that is primed for lineage aberrancy.

To gain further insight into the relative roles of founding genomic lesions, acquired genetic alterations and the role of therapy in dictating MPAL phenotype, we analysed sequential samples obtained at initial diagnosis and disease recurrence in nine patients. The immunophenotypes of five cases (three T/M MPAL, one B/M MPAL, and one MPAL NOS with T/B phenotype) were stable from diagnosis and relapse, but changed in four cases. Two were ALL (one B-ALL, one ETP-ALL) at diagnosis and relapsed as MPAL, and two were MPAL at diagnosis (one T/M, one B/M) and subsequently relapsed as AML and ALL, respectively (Extended Data Fig. 10). In the five cases with immunophenotypic stability, mutations in the predominant clone were lost (*PTPN11*, *CCND3*, *NOTCH1*, and *RPL22*) or emerged (*TP53*, *IKZF1*, *NF1*, *NCOR1*, and *SUZ12*). Despite this genomic evolution, the lineage ambiguity remained, further supporting the notion that MPAL leukaemia-initiating cells are primed for multi-lineage potential. In all four cases with phenotype shifts, the initial therapy correlated with the type of phenotype shift: patients who received ALL-directed therapy relapsed with myeloid leukaemia and one patient who received AML-directed therapy relapsed with lymphoid leukaemia. In two cases, immunophenotype at relapse was

also correlated with a mutation characteristic of leukaemia subtype: *CEBPA* for AML and *CDKN2A* or *CDKN2B* for B-ALL. Together, these nine cases with serial samples support the theory that early genomic lesions prime progenitors for lineage aberrancy, which may remain stable or change over time, and that phenotype is influenced by therapeutic pressure and/or genomic evolution.

Discussion

This study provides a comprehensive genomic analysis of paediatric MPAL, providing insights into the genomic relationships between immunophenotypically defined subtypes of acute leukaemia. We propose an update to the WHO classification of acute leukaemia that includes new subtypes of *ZNF384*-rearranged acute leukaemia (either B-ALL or MPAL), *WT1*-mutant T/M MPAL, and Ph-like B/M MPAL (Extended Data Fig. 1c).

The ALL-like genomic landscape of B/M MPAL and the similarity in genomic alterations between *ZNF384* B/M MPAL and B-ALL supports the use of ALL-directed therapy for patients with B/M MPAL. Furthermore, the overexpression of *FLT3* and responsiveness to *FLT3* inhibition in *ZNF384* leukaemia¹⁷ suggest that such targeted therapy should be considered in this form of leukaemia. Non-*ZNF384* cases of B/M MPAL should be carefully evaluated for other kinase-activating alterations that may be amenable to kinase inhibition, as shown in Ph-like ALL¹⁸.

Our data show that ETP-ALL⁵ and T/M MPAL are genomically and epigenomically similar, and suggest that *FLT3* and/or *JAK* inhibition should be evaluated further⁴. T/M MPAL exhibits infrequent alteration of core T-ALL transcription factor genes and few mutations in *CDKN2A*, *CDKN2B*, *NOTCH1* and *FBXW7*; frequent *FLT3*-activating mutations; and a GEP that overlaps with that of AML, consistent with the notion that the pathogenesis of T/M MPAL is distinct from that of T-ALL. However, contemporary paediatric ALL trials have demonstrated remarkable success in treating ETP-ALL, which is similar to T/M MPAL, so ALL-directed therapy may also be appropriate for T/M MPAL¹⁹.

In contrast to the notion that subclonal genomic variation drives clonal evolution during disease progression in ALL¹⁴, our analysis of phenotypically distinct subpopulations within individual patients with MPAL revealed that mutational variegation did not determine phenotypic diversification. Rather, the common genomic features of *ZNF384* B-ALL and MPAL, limited mutational variegation between subclones, multi-lineage potential of subclones in xenograft models, lineage plasticity in serial patient samples, and identification of leukaemia-initiating alterations in early haematopoietic progenitors indicate that the ambiguous phenotype of MPAL is the result of the acquisition of alterations in immature haematopoietic progenitors (Fig. 4c, d). These data also support a model of haematopoiesis in which progenitors retaining multilineage potential undergo terminal differentiation into a single lineage only relatively late in haematopoiesis²⁰.

By demonstrating the genomic similarity of phenotypically distinct malignant populations, and by identifying the potential clinical importance of *ZNF384* fusions, these results emphasize the limitations of morphology and immunophenotype alone in diagnostic

evaluation. As has been demonstrated in AML, ETP-ALL, MDS, and Ph-like ALL^{5,11,18,21,22}, accurate MPAL sub-classification requires careful genomic analysis to optimally guide diagnosis, risk-stratification and tailoring of therapy. Together, these findings have implications for disease classification and therapeutic decisions, while also clarifying the pathogenesis of this high-risk subtype of acute leukaemia.

METHODS

Patients and samples

Diagnosis and remission samples were obtained from St. Jude Children's Research Hospital (SJCRH), the Children's Oncology Group, the European Organization for Research and Treatment of Cancer—Children's Leukaemia Group, the Belgian Society for Paediatric Hematology—Oncology, the Dutch Children's Oncology Group, the Italian Association of Paediatric Hematology and Oncology, the Japanese Association of Childhood Leukaemia Study, the Tokyo Children's Cancer Study Group, the I-BFM Study Group, the Queensland Children's Tumour Bank, The Children's Hospital at Westmead, Schneider Children's Medical Center, Yong Loo Lin School of Medicine in Singapore, and the United Kingdom Childhood Leukaemia Cell Bank. After central review of pathology and immunophenotyping of 159 cases, 115 patients diagnosed with ALAL were included in this analysis, including 80 with germline samples. We examined leukaemia samples from 115 patients with ALAL (Supplementary Table 2–4) using whole-exome sequencing (WES) or WGS, transcriptome sequencing (RNA-seq), SNP microarray, and methylation array analysis. Samples collected on tumour banking protocols were used. Samples were not prospectively collected. The study was approved by the SJCRH Institutional Review Board.

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

Tissue

Non-tumour DNA was extracted from remission bone marrow or peripheral blood samples, flow-sorted normal lymphocytes, or cultured fibroblasts using phenol-chloroform organic extraction. Tumour DNA was extracted using phenol-chloroform organic extraction. Tumour RNA was extracted using a TRIzol (Life Technologies).

Whole genome/exome and transcriptome sequencing

WGS for 44 cases and RNA-seq for 45 cases were performed by the British Columbia Cancer Agency's Michael Smith Genome Sciences Centre (BCGSC); WGS for 3 cases, WES for 92 cases and RNA-seq for 77 cases were performed at SJCRH. For WGS at BCGSC, methods for DNA preparation, sequencing, and quality control are available at <https://ocg.cancer.gov/programs/target/target-methods>. For WES at SJCRH, library construction used DNA tagmentation (fragmentation and adaptor attachment) performed using the reagent provided in the Illumina Nextera rapid exome kit, and was performed using the Caliper Biosciences (Perking Elmer) Sciclone G3. First-round PCR (10 cycles) was performed using Illumina Nextera kit reagents, and clean-up steps employ BC/

Agencourt AMPure XP beads. Target capture used Illumina Nextera rapid capture exome kit and supplied hybridization and associated reagents. The pre-hybridization pool size was 12 samples, and second round PCR (10 cycles) performed with Nextera kit reagents. Library quality control was performed using a Victor fluorescence plate reader with Quant-it dsDNA reagents for pre-pool quantitation, and Agilent Bio-analyzer 2200 for final library quantitation. Paired-end sequencing was performed using Illumina HiSeq 2500 with read length 100 bp.

Methods for RNA preparation, sequencing, and quality control at BCGSC are available at <https://ocg.cancer.gov/programs/target/target-methods>. At SJCRH, total RNA quality and quantity were assessed on Agilent RNA6000 chips (Agilent Technologies) and Qubit (Life Technologies). RNA-seq libraries were prepared from 500 ng of total RNA for each sample following Illumina RNA-seq protocols, including DNase treatment and phenol purification, cDNA conversion, fragmentation by Covaris Ultrasonicator, end repair, deoxyadenosine tailing, adaptor ligation and PCR amplification (ten cycles). Libraries with a 10 pM concentration were clustered on an Illumina cBot, and each flow cell was loaded onto a HiSeq instrument for sequencing using the Illumina 2×100 bp sequencing kit. RNA-seq was not performed on flow sorted subpopulations due to the deleterious effects on RNA integrity of cellular fixation/permeabilization performed to enable staining for intracellular markers.

Sequencing read alignment

Paired-end WGS and WES data were aligned to the human reference genome GRCh37 by BWA²³ (version 0.7.12). Samtools²⁴ (version 1.3.1) was used to generate chromosomal coordinate-sorted and indexed bam files, and then Picard (<http://broadinstitute.github.io/picard/>, version 1.129) MarkDuplicates module was used for marking PCR duplication. Afterwards, the reads were realigned around potential indel regions by GATK²⁵ (version 3.5) IndelRealigner module following the recommended pipeline. Sequencing depth and coverage was evaluated based on coding regions defined by refSeq genes from UCSC, with the length around 34 Mb.

SNV/indel calling and filter workflow

The GATK UnifiedGenotyper module was used to identify SNVs and indels from leukaemia and germline samples, which were filtered by a homemade pipeline, excluding: 1) reported common SNPs/indels from UCSC dbSNP v142; 2) germline mutations detected from matched germline control samples. All the non-silent SNVs/indels yield from the filtering pipeline were manually reviewed and only the highly reliable somatic ones were reported. Meanwhile, adjacent nucleotide changes on the same allele were merged into a single mutation. For patients with flow sorted subpopulations of leukaemia cells sequenced, the mutation calling for each population was performed de novo. Mutations detected from some/one of the samples were checked across the other samples from the same patient. In these cases, we applied a threshold of at least 3 mutant allele reads and variant allele frequency of at least 1% to report a mutation. For cases without germline samples, a germline sample was picked with highest sequence depth as a pseudo-germline sample to run through the filtering pipeline. In cases in which flow sorted subpopulations were sequenced, WES or WGS of the unfractionated samples were not performed.

Structure variant detection

Structural variants in the tumours were identified by CREST²⁶ using tumour vs germline mode, with pseudo-germline data applied for tumours without germline samples. Candidate variants were manually reviewed and the mapping uniqueness was re-evaluated by running BLAT²⁷ mapping and the confident calls were considered as the final structural variant set.

RNA-seq data analysis for patient samples

Paired-end reads were mapped to the GRCh37 human genome reference by STAR²⁸ (version 2.5.1b) through the recommended two pass mapping pipeline with default parameters and the Picard MarkDuplicates module was used to mark the duplication rate. Gene annotation files were downloaded from Ensembl (<http://www.ensembl.org/>) and used for STAR mapping and subsequent gene expression level evaluation. CICERO¹⁸ and FusionCatcher²⁹ were used to detect fusions from mapped BAM files and raw FASTQ files, respectively. The reported fusion contigs were remapped by BLAT to check the reliability of mapping quality, the breakpoints were manually reviewed from the aligned reads and the highly confident fusions were reported. To evaluate GEP, reads count for annotated genes was called by HTSeq³⁰ (version 0.6.0) and processed by DESeq2 R package³¹ to normalize gene expression into regularized log₂ values (rlog). Six cases without DNA sequence data were screened for SNVs/indels by following the GATK Best Practices for Variant Calling on RNAseq (<https://gatkforums.broadinstitute.org/gatk/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail>). The filtering process is the same as germline variant analysis described below.

Gene set enrichment and pathway analysis

Read counts from RNA-seq data were imported to DESeq2³² R package for differential gene expression analysis. To perform gene set enrichment analysis (GSEA)³³, all the genes were ranked according to the fold-change and significance from differential analysis. GSEA was performed using mSigDB C2 genes and curated gene sets from in house analyses.

Cell line transcriptome analysis

Total RNA was isolated from green fluorescent protein (GFP)-positive, sorted cells using the RNeasy Mini Kit (Qiagen). RNA quality was checked using 2100 Bioanalyzer RNA 6000 Nanoassay (Agilent) or LabChip RNA Pico Sensitivity assay (PerkinElmer) before library generation. Libraries were prepared from total RNA with the TruSeq Stranded Total RNA Library Prep Kit (Illumina). Libraries were quantified using the Quant-iT PicoGreen dsDNA assay (Life Technologies) Kapa Library Quantification kit (Kapa Biosystems) or low pass sequencing on a MiSeq Nano v2 run (Illumina). One hundred cycle paired end sequencing was performed on an Illumina HiSeq 2500, HiSeq 4000, or NovaSeq 6000. RNA isolation, library preparation, and sequencing were performed on three biological replicates. RNA-seq data were mapped as described previously¹⁸ and HTSeq³⁰ (version 0.6.1p1) were used to get gene-level count and estimated FPKM based on GENCODE (vM9)³⁴. Voom³⁵ was used for gene differential expression analysis after trimmed mean normalization.

CNA and loss of heterozygosity (LOH)

DNA from leukaemia and matched germline samples was prepared for hybridization to Illumina Infinium Omni2.5 Exome-8 SNP arrays according to the manufacturer's protocol. The raw intensity data (*.idat files) were analysed by the Genotyping Module of Illumina Genome Studio software version 2.0.3. Normalized log R ratio (LRR) and B allele frequency (BAF) for all the available probes in each sample were extracted. For *ZNF384* B-ALL cases, data acquired from Affymetrix Genome-Wide Human SNP Array 6.0 was also converted to LRR and BAF value following the pipeline described by PennCNV³⁶ (<http://penncnv.openbioinformatics.org/en/latest/user-guide/affy/>). With the input of LRR and BAF, somatic genomic alterations in paired or unpaired samples were called by OncoSNP version 2.1³⁷. To verify the reliability of CNAs and LOHs, all the reported alterations were plotted based on LRR and BAF in ShinyCNV (<https://github.com/gzhmat/ShinyCNV>) and visually checked³⁸. Only somatic alterations meeting the criteria proposed by OncoSNP and PennCNV were kept for further analysis.

DNA methylation assay and data analysis

We examined DNA methylation profiles in 27 MPAL cases (11 with 2–4 subpopulations), 15 AML, 29 B-ALL, 30 T-ALL, and 17 normal lymphocyte samples from 4 healthy donors. Raw data from the Infinium MethylationEPIC BeadChip Kit (Illumina Inc.) were analysed using the ChAMP³⁹ R package. In general, the raw *.idat files were imported through 'minfi' method⁴⁰ and then the following filters were applied to exclude the probes: 1) with detection *P* value above 0.01 in one or more samples; 2) with beadcount <3 in at least 5% of samples; 3) as non-CpG probes; 4) identified as SNPs⁴¹; 5) aligned to multiple locations⁴²; and 6) on the X or Y chromosome. After filtering, 'BMIQ' normalization from ChAMP package was used as the author suggested to calculate methylation beta values. Batch effect was observed by the singular value decomposition method⁴³ and adjusted by ComBat normalization method⁴⁴. The 5,000 probes with the highest median absolute deviations (MAD) were used to perform clustering with a two-dimensional *t*-distributed stochastic neighbour embedding (t-SNE) plot and heatmap⁴⁵.

Fusion validation

Fluorescence *in situ* hybridization (FISH) was performed to confirm fusions in 22 cases (Supplementary Table 19) using the probes listed, in Carnoy's fixative as previously described⁴⁶. BAC clones (Supplementary Table 20) were labelled with rhodamine or fluorescein isothiocyanate. At least 100 interphase nuclei were scored per case.

Flow cytometric analysis and flow cytometric assisted cell sorting

Flow cytometric analysis and sorting were performed on an 18 colour Aria cell sorter (BD Biosciences). When available, cryopreserved samples were analysed by flow cytometry using CD45-APC-H7 (BD Catalog #560178), cytoplasmic CD3-PE (BD #347347), CD34-PerCP Cy5.5 (BD #347203), CD19-APC (BD #340437), cytoplasmic MPO-FITC (Dako #F071401-1), and CD33-PE-Cy7 (BD #333946). Depending on the phenotypes reported from the outside institutions, samples were additionally analysed using cytoplasmic CD79a-APC (BD #551134), CD22-BV421 (BD #563940), CD64-PerCP-Cy5.5 (BD #561194),

CD14-PE-Cy7 (BD #560919), cytoplasmic lysosome-FITC (Life Technologies #GIC207), and CD11c-APC (BD #560895). For 50 cases, leukaemic cells in the CD45 and side scatter-defined blast gate were sorted into subpopulations based upon cytoplasmic MPO and either CD19 or cytoplasmic CD3. When feasible, normal lymphocytes were sorted using side scatter, CD45 lymphocyte gate and secondarily using CD19 and cytoplasmic CD3 to collect normal B-cells in T/M MPAL cases and normal T-cells in B/M, *KMT2Ar*, AUL, or NOS cases.

Fibroblast cultures

Bone marrow cells were cultured in change medium (Irvine scientific, T105), which was changed every 5 days. Cells were collected for DNA extraction when the fibroblasts became at least 70% confluent.

Comparison cohorts

Comparison cohorts of AML, ETP-ALL, non-ETP T-ALL, and B-ALL were examined. A cohort of 197 paediatric AML patients from the COG with WGS performed through the NCI TARGET initiative was used as comparison (Supplementary Table 34) and publicly available data can be found at <https://ocg.cancer.gov/programs/target12>. Nineteen ETP-ALL and 245 non-ETP T-ALL cases from the COG were sequenced through the NCI TARGET project using WES and total stranded RNA-seq for SNV, indel, CNA and SV calls, fusion detection and GEP comparison¹¹ (Supplementary Table 35). A cohort of AML, B-ALL^{16,18,47-49} ($n = 161$), T-ALL¹¹ ($n = 50$), and ETP-ALL¹¹ ($n = 19$), and 12 normal lymphocyte samples was used for GEP comparison (Supplementary Table 23). The AML samples were sequenced at SJCRH and had stranded total RNA-seq for GEP comparisons. This cohort consists of five cases with core binding factor translocations (three with *RUNX1-RUNX1T1*, two with *CBFB-MYH11*), five cases with normal karyotype, and five cases with *KMT2Ar*).

B-ALL subtyping based on GEP

RNA-seq data analysis for patient samples is described above. As many B-ALL subtypes defined by single chromosomal aneuploidy or rearrangement may be clustered based on their GEP^{48,50-54}, a subtype prediction model was trained by prediction analysis of microarrays (PAM)⁵⁵ using a cohort of 322 B-ALL samples from our previous studies^{18,48,49}, which consists of eight canonical B-ALL subtypes: *DUX4* rearrangement ($n = 40$), *ETV6-RUNX1* ($n = 42$), high hyperdiploidy ($n = 45$), *MEF2D* rearrangement ($n = 29$), *KMT2Ar* ($n = 44$), *TCF3-PBX1* ($n = 40$), *BCR-ABL1* ($n = 42$) and *ZNF384* rearrangement ($n = 40$). The PAM model was trained on 200 different thresholds with tenfold cross-validation. Based on the trained model and cross-validation result, 100 thresholds (control the selected feature genes from 5,000 to 50) were tested on the training data set to determine the optimal threshold range for each subtype. Then the trained model was applied to the MPAL samples to determine their similarity to each B-ALL subtype for 100 rounds, using evenly distributed thresholds across the optimal threshold range for each B-ALL subtype, and the average score was taken as the consensus likelihood score for that subtype (Supplementary Table 33).

Germline variant analysis

Germline variants were called by GATK⁵⁶ UnifiedGenotyper from the bam files of all the germline samples, and then the following filters were applied to identify potential pathogenic germline variants: 1) exclusion of variants with fewer than five mutant reads support or a VAF below 20%; 2) exclusion of variants in common SNP database (VAF greater than 0.1% in population according to dbSNP 142); 3) exclusion of SNPs with at least ten occurrences observed in dbSNP 142 but not reported as somatic mutations in the COSMIC V80 database; 4) exclusion of variants in genes with fewer than three somatic mutations in MPAL cohort; 5) annotation of variants using the Variant Effect Predictor (VEP; <https://useast.ensembl.org/Tools/VEP>) and then exclusion of variants predicted as benign by any of the predictors (SIFT, PolyPhen, Condel). The remaining mutations were manually reviewed and obvious mapping artefacts were excluded. Mutations were then assessed according to ACMG recommendations⁷.

Lentiviral transduction of cells

cDNAs encoding ZNF384 (XM_017018949), TAF15 (NM_139215)–ZNF384 (exon 6–exon 3), and TCF3 (NM_003200)–ZNF384 (exon 13–exon 5) were amplified from human leukaemic cell RNA and cloned with a C-terminal HA epitope tag (added using the QuikChange II XL Site-Directed Mutagenesis Kit, Agilent) into the CL20c-MSCV-IRES-GFP vector. Vectors were packaged into lentiviral particles by transient transfection of HEK293T cells with a triple plasmid (pHDMG, pCAG HIV, pCAG RTR) system. Lentiviral supernatants were used to infect interleukin-7 (IL-7)-dependent *Arf*^{-/-} pre-B cells on RetroNectin (Takara Bio) for 48 h before sorting for GFP⁺ cells (BD FACSAria, BD Biosciences).

Chromatin immunoprecipitation and sequencing

ChIP assays were carried out as described previously⁴⁹. In brief, 2×10^7 GFP positive cells were incubated for 10 min in 1% formaldehyde in phosphate buffered saline (PBS) at room temperature, quenched by the addition of 1/10 volume of 2 M glycine. Cells were then washed three times with cold PBS containing proteinase inhibitors and lysed on ice for 10 min in lysis buffer (50 mM HEPES, pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100). Chromatin was washed twice in washing buffer (10 mM Tris-HCl, pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and then twice in shearing buffer (0.1% SDS, 10 mM Tris-HCl, pH 8, 1 mM EDTA) before resuspension in 1 ml shearing buffer. Chromatin was sonicated in 1-ml AFA millitubes using a Covaris E210 instrument for 15 min at 5% duty cycle, intensity 4, 200 cycles per burst at 4 °C. Sheared chromatin was spun down for 10 min at 13,200g at 4 °C, and the supernatant was mixed with an equal amount of ChIP dilution buffer (0.1% SDS, 30 mM Tris-HCl, pH 8, 1 mM EDTA, 300 mM NaCl, 2% Triton X-100) before ChIP experiments. Immunoprecipitation was performed with an antibody to HA (ab9110, Abcam) and a normal rabbit IgG control (Santa Cruz Biotechnology) using 2 µg antibody per ChIP. This experiment was performed with three biological replicates.

To prepare ChIP-seq libraries, 10 ng of ChIP DNA was end repaired and adaptor ligation was performed using the Next ChIP-Seq Library Prep Reagent Set for Illumina (New

England BioLabs). Libraries were purified after 14 rounds of PCR amplification with Q5 DNA Hot-Start polymerase (New England BioLabs). Each ChIP-seq library underwent 50-cycle single-end sequencing using TruSeq SBS kit v3 on an Illumina HiSeq 2000.

Alignment and quality control were performed as described⁵⁷. Fifty base pair single-end reads were mapped to mouse genome mm9 (MGSCv37) with BWA²³ (version 0.7.12-r1039), duplicated reads were marked with Picard and only unique mapped reads extracted by Samtools²⁴ (version 1.2) were kept for analysis. We extended each read to estimated fragment size by SPP⁵⁸ (version 1.1) and generated bigwig files, scaling the track by normalizing to 15 million unique mapped reads.

For differential binding analysis, peaks were called with MACS2⁵⁹ (version 2.0.10.20131216, parameter ‘-nomodel-extsize FRAGMENT SIZE’ and fragment size was estimated as described above by SPP⁵⁸ (version 1.1) twice for each sample. High confidence peaks used a cutoff of FDR corrected *P* value of 0.05 and low confidence peaks used a cutoff of FDR-corrected *P* < 0.5. Peaks from replicates were merged only if called as high confidence peaks in one sample and called as at least low confidence peaks in other replicates. Finally, peaks from WT *ZNF384* and *ZNF384* fusions were merged as a reference peak set. For each sample, we first extend read to the estimated fragment size, then we counted the extended reads number overlapping the reference peaks by BEDTools (version 2.24.0)⁶⁰. Following PCA analysis, which showed a clear separation of WT and fusion ChIP-seq data, Voom³⁵ was used to examine differences in strength of binding between WT and fusion after trimmed mean normalization. Common differential binding sites (*q* value less than 0.05 and fold change greater than 1) between TAF15-ZNF384 vs WT and TCF3-ZNF384 vs WT were used for visualization. Real-time PCR (Ct method) was employed to validate ChIP-seq results. Differential binding sites were annotated to genes if their promoter (transcription start site ± 2kb) overlapped the binding sites. Gene set enrichment analysis³³ was used to compare ChIP-seq peak lists to the GEP of cell lines expressing ZNF384 fusions.

Statistical analysis

The correlation between sex, disease subtype (WHO 2016 criteria, our proposed update to classification of ALAL, or fusion presence/absence) and single gene mutation or pathway mutations was assessed using the two-sided Fisher exact test. The correlation between subtypes and age categories was assessed using the two-sided Fisher exact test. The correlation between age as a continuous variable and single gene mutation or pathway mutations was assessed using the non-parametric Wilcoxon rank-sum test. The Kaplan-Meier method was used to estimate the survival function and overall survival distributions were compared with log-rank tests. GraphPad Prism (version 7.04) and SAS (version 9.4) were used for statistical analysis.

Fluorescence-activated cell sorting (FACS) of human stem/progenitor and mature cell populations

For sorting of HSC and progenitor cells, mononuclear cells from diagnosis bone marrow samples from patient SJMPAL040028 were stained with the following human-specific

antibodies (all from BD Biosciences unless stated otherwise, catalogue number in parentheses): anti-CD45RA-FITC (555488), anti-CD90-PE (Biolegend, 328109), anti-CD135-BV711 (563908), anti-CD38-PE-Cy7 (335790), anti-CD10-BV421 (562902), anti-CD7-V450 (642916), anti-CD45-Alexa 700 (Thermo Fisher Scientific MHCD4529), anti-CD34-APC-Cy7 (Biolegend, custom made, CD34 clone 581), anti-CD33-APC (340680) and anti-CD19-BV605 (562653). For sorting of mature cells and leukaemic blasts, mononuclear cells from bone marrow of patient SJMPAL040028 were stained with the following antibodies: anti-CD45-Alexa 700 (Thermo Fisher Scientific MHCD4529), anti-CD19-BV605 (562653), anti-CD10-BV421 (562902), anti-CD33-PE-Cy7 (333946), CD3-PE (347347) and anti-CD56-Alexafluor 647 (557711). For all samples, cells (from 5 to 1,000) per fraction were sorted on a BD FACS Aria in a 96-well plate. As previously published¹⁶ and as described⁶¹, progenitor populations were all gated on CD45⁺CD33⁻CD19⁻ and sorted into haematopoietic stem cells (HSC; CD38⁻CD34⁺CD90⁺CD45RA⁻); multipotent progenitor fraction (MPP; CD38⁻CD34⁺CD90⁻CD45RA⁻); multilymphoid progenitor fraction (MLP; CD38⁻CD34⁺CD45RA⁺); megakaryocyte erythroid progenitors (MEP)/common myeloid progenitors (CMP; CD38⁺CD34⁺CD7⁻CD10⁻CD45RA⁻); and granulocyte monocyte progenitor (GMP; CD38⁺CD34⁺CD7⁻CD10⁻CD45RA⁺) subsets. Leukaemia blasts were gated on CD45^{dim} expression and sorted into the following fractions: CD45^{dim}CD33⁺CD19⁺CD10⁻; CD45^{dim}CD33⁺CD19^{moderate}CD10⁻; CD45^{dim}CD33⁺CD19⁻CD10⁻; and CD45^{dim}CD33⁻CD19⁻. Normal mature populations were gated on CD45^{high} expression and sorted into T cells (CD45^{high}CD3⁺) and NK cells (CD45^{high}CD56⁺). The following numbers of cells were sorted in a single well of a 96-well plate (each number in the parenthesis is a replicate): HSC (6 and 21); MPP (387); MLP (12); MEP/CMP (500); GMP (21 and 18); CD45^{dim}CD33⁺CD19⁺CD10⁻ (1,000; 5 replicates); CD45^{dim}CD33⁺CD19^{moderate}CD10⁻ (1,000; 6 replicates); CD45^{dim}CD33⁺CD19⁻CD10⁻ (1,000; 6 replicates); CD45^{dim}CD33⁻CD19⁻ (82 and 36); T cells (1,000; 5 replicates) and NK cells (100 and 40). DNA from all sorted populations was amplified by whole-genome amplification (WGA) by REPLI-g Single Cell Kit (150345, Qiagen) according to the manufacturer's protocol.

Genomic analysis of sorted subpopulations

Upon completion of WGA, DNA was subjected to PCR amplification using primers specific for the *TCF3-ZNF384* fusion or for additional genetic alterations, including SNVs/indels in *NDST2*, *DNAH17* and *MYCN*, identified from analysis of whole exome sequencing data. Primers were designed to flank the fusion breakpoint or the identified variants using Primer3 (*TCF3-ZNF384_F*: 5'-GAGGAGGACCAGGAGAGATGG-3' and *TCF3-ZNF384_R*: 5'-ATCAGGCAAGGCTTCCTAAAAG-3'; *NDST2_F*: 5'-ATAGGTACACTCCCTGCCTTTCC-3' and *NDST2_R*: 5'-ACCCCAAACCTTGACCCTTTT-3'; *DNAH17_F*: 5'-CTCCTCTTTGGGAACCCTCTG-3' and *DNAH17_R*: 5'-GAAAAGGCTTGTGCTGACATCTT-3'; *MYCN_F*: 5'-GTGCTGTCTCGGTTGCAGTGTT-3' and *MYCN_R*: 5'-AGTCGTTCTCAAGCAGCATCT-3'). PCR was performed using KAPA2G Fast HotStart Ready Mix (#07961260001, Kapa Biosystems) according to the manufacturer's instructions with 10 µM each primer and 2 µl diluted (1:100) WGA DNA. All amplicons were quality

checked on a 1.5% agarose gel and purified using Wizard SV Gel and PCR Clean-Up System (A9282, Preprotech). Sequence was verified by Sanger sequencing. The sequenced amplicon was aligned to a reference fusion sequence generated from National Center for Biotechnology Information and to the contigs obtained from RNA-seq in case of *TCF3-ZNF384* fusion. The results were analysed using CLC Main Workbench (Qiagen).

Xenografts

Mice were housed in an American Association of Laboratory Animal Care (AALAC)-accredited facility and were treated according to Institutional Animal Care and Use Committee (IACUC) protocols approved by SJCRH in accordance with NIH guidelines.

JIH-5-derived xenograft

The JIH-5 cell line¹⁵ was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). Cells were thawed and cultured according to DSMZ's instructions (<https://www.dsmz.de/catalogues/details/culture/ACC-788.html>). Immunophenotypic and genomic analyses (RNA-seq) were performed prior transplant assays. Short Tandem Repeat (STR) DNA analysis was performed for cell line authentication (Supplementary Table 38), showing concordance with DSMZ STR analysis. STR analysis was performed using the The PowerPlex 16 HS System (Promega) which allows co-amplification and three-colour (blue or fluorescein-labelled, black or TMR-labelled, and green or JOE-labelled) detection of sixteen loci (fifteen STR loci and Amelogenin), including Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818. All sixteen loci were amplified simultaneously in a single tube and analysed in a single injection. Cells were transduced with a lentiviral vector (vCL20SF2-Luc2a-YFP) expressing luciferase and yellow fluorescent protein (YFP) and FACS sorted for YFP. YFP-positive (YFP⁺) cells were stained with anti-CD19-APC (BD, 340437), anti-CD34 PerCP-Cy5.5 (BD, 347203) and anti-CD33 PE-Cy7 (BD, 333946) and sorted in the following subpopulations: YFP⁺CD34⁺; YFP⁺CD34⁻CD19⁺CD33⁺; and YFP⁺CD34⁻CD19⁺CD33⁻. FACS-sorted leukaemia subpopulations or YFP⁺ bulk cells were intravenously injected in 8- to 10-week-old female NSG-SGM3 mice⁶². The number of cells that was transplanted and the total number of mice transplanted per subpopulation depended on the number of viable cells that were available, but ranged from 0.2 to 0.6 million cells and 1 to 5 mice, respectively. Engraftment was monitored by weekly measurement of bioluminescence (Region of Interest, ROI) at Xenogen IVIS-200 (PerkinElmer). ROI measurements and total fluxes (photons/second, p/s) were recorded and analysed by the Living Imaging v.4.4 software (Caliper Life Sciences). When total fluxes were at least 1×10^8 in all animals, mice were euthanized, and blood, bone marrow, and spleen samples were analysed to determine the leukaemia phenotype, using morphology, flow cytometry, and histopathologic analysis.

MPAL patient-derived xenografts (PDX)

MPAL PDX were established from three patients (SJMAPL011911, SJMAPL014124 and SJMAPL040036). Frozen mononuclear cells from bone marrow at diagnosis were thawed and used as bulk (SJMAPL040036) or flow-sorted in transplantation assays. Cells from SJMAPL011911 were stained with the following human-specific antibodies: anti-CD45-

APC-H7 (BD, 641399), CD34-PerCP-Cy5.5 (BD, 347203), anti-CD33 PE-Cy7 (BD, 333946) and anti-CD7-PE-Cy-7 (BD, 544019). Blast cells were gated on CD45^{dim} expression and sorted into CD45^{dim}CD7⁺CD33⁻, CD45^{dim}CD7⁻CD33⁺ and CD45^{dim}CD7⁻CD33⁻. Cells from SJMAPL014124 were stained with the following human-specific antibodies: anti-CD45-APC-H7 (BD, 641399), anti-CD33 PE-Cy7 (BD, 333946) and anti-CD19-APC (BD, 340437) and sorted into CD45^{dim}CD19⁺CD33⁺, CD45^{dim}CD19⁺CD33⁻ and CD45^{dim}CD19⁻CD33⁺. Bulk or sorted cells were intravenously injected into 8- to 10-week-old female NSG-SGM3⁶² mice that were sublethally irradiated (250 RAD) 6–24 h before transplantation. The number of cells that was transplanted and the total number of mice transplanted per sample depended on the number of viable cells that were available, but ranged from 0.2 to 0.6 million cells and 1 to 5 mice, respectively. Human leukaemia engraftment was monitored in peripheral blood by performing serial retro-orbital bleeds one month after injections and monthly thereafter. Peripheral blood samples were analysed by flow cytometry for human CD45⁺ cells and when CD45⁺ cells were >5%, mice were euthanized, and blood, bone marrow, and spleen samples were analysed to determine the leukaemia phenotype, using morphology, flow cytometry, and histopathologic analysis. Immunohistochemistry (IHC) was performed on formalin-fixed paraffin-embedded tissues sectioned at 4 µm. Assays for CD19 (AbDserotec, MCA2454T; 1:100), CD34 (Ventana, 790–2927; ready to use), CD45 (Ventana, 760–2505; ready to use) and myeloperoxidase (MPO, DAKO A398; 1:500) were performed on the Ventana Benchmark. The assay for CD33 (Leica Biosystems, NCL-L-CD33; 1:200) was performed on the Dako Omnis.

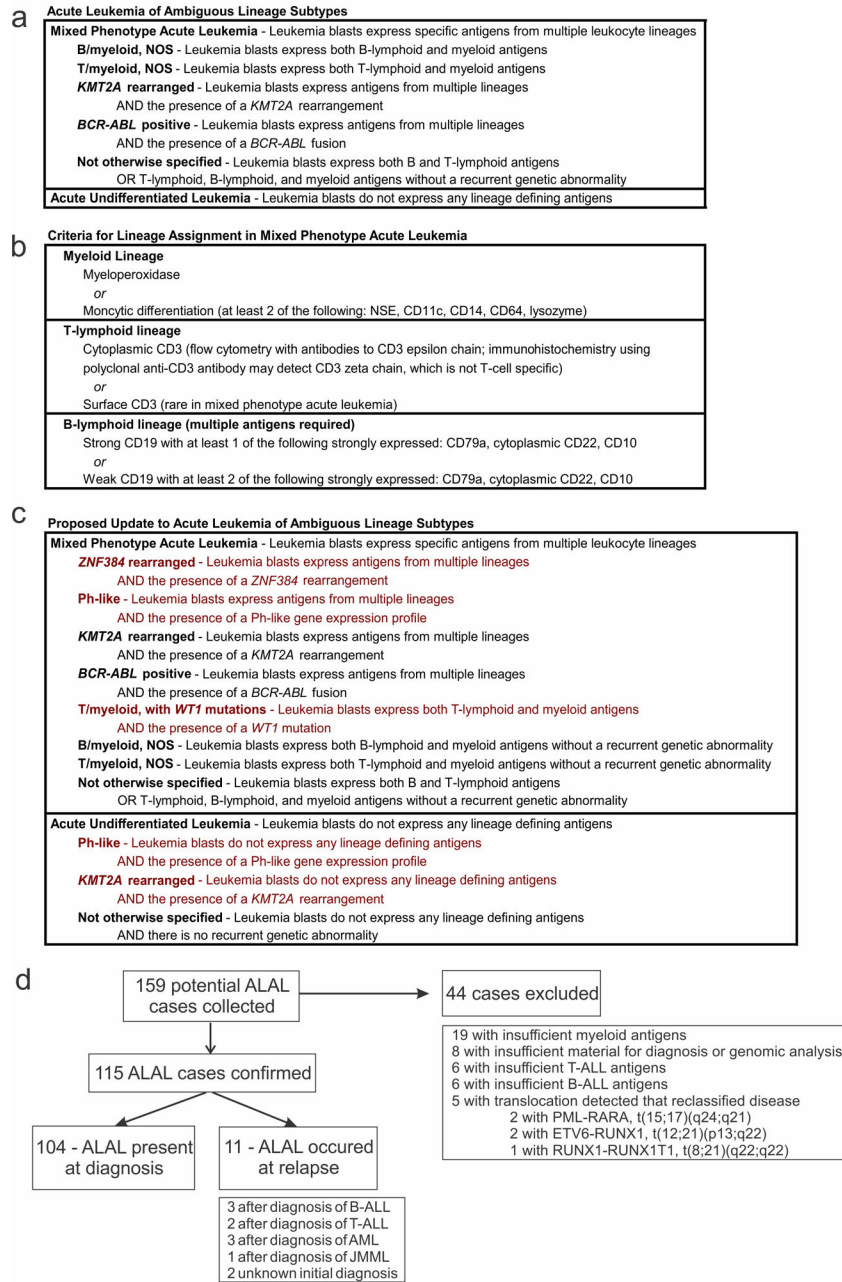
Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequencing, SNP, and methylation data are available at the NCI Genomics Data Commons (GDC, gdc.cancer.gov) and analysed data may be accessed at the TARGET website at <https://ocg.cancer.gov/programs/target/data-matrix> or <https://gdc.cancer.gov/about-data/publications/TARGET-ALAL-2018>. Murine RNA-seq and ChIP-seq data have been deposited in the GEO database under accession ID GSE112561. For T-ALL and ETP-ALL, RNA sequencing for comparison comprised previously published data¹¹. B-ALL RNA-sequencing data for comparison comprised previously published data and recently sequenced samples that will be made available through St Jude's Children's Research Hospital^{11,18,48,49,63}. T-ALL, ETP-ALL, and AML data for mutation comparison comprised previously published data^{11,12}. The genomic landscape reported in this study can be explored at the St. Jude PeCan Data Portal, <http://pecan.stjude.org/proteinpaint/study/pediatric-mpal>.

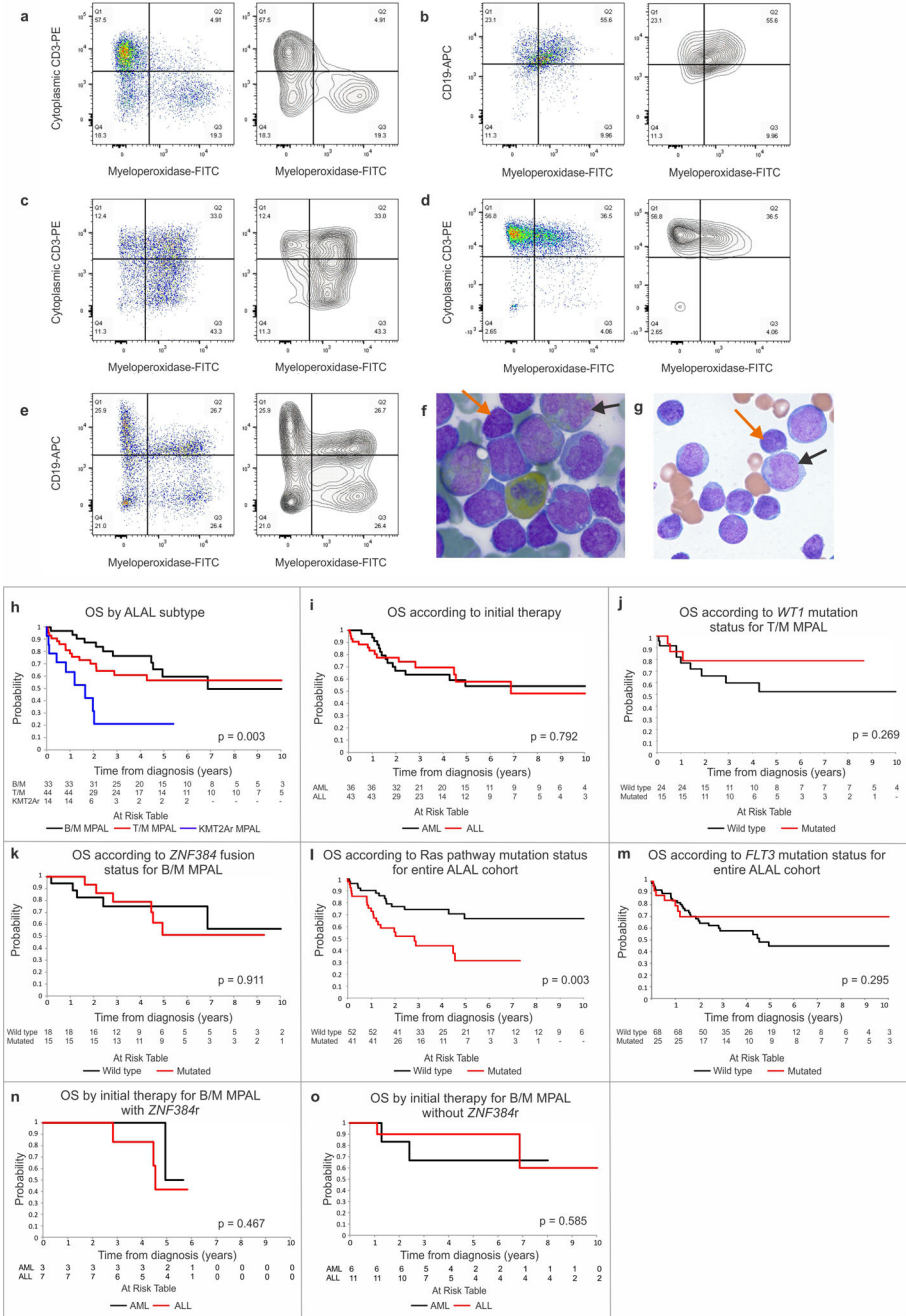
Extended Data



Extended Data Fig. 1 |. Criteria for diagnosis of ALAL.

a, Subtypes of ALAL according to the WHO 2008 criteria and consistent with minor revisions of WHO 2016 criteria⁶. **b**, Antigen requirements for lineage assignment for MPAL according to WHO 2008 criteria. The 2016 revisions to the WHO classification for ALAL did not change the above categories or requirements. Rather, the revision emphasized that care should be taken before making a diagnosis of B/M MPAL when low-intensity myeloperoxidase is the only myeloid-associated feature. Additionally, the revision emphasized that in cases in which it is possible to resolve two distinct blast populations, it is

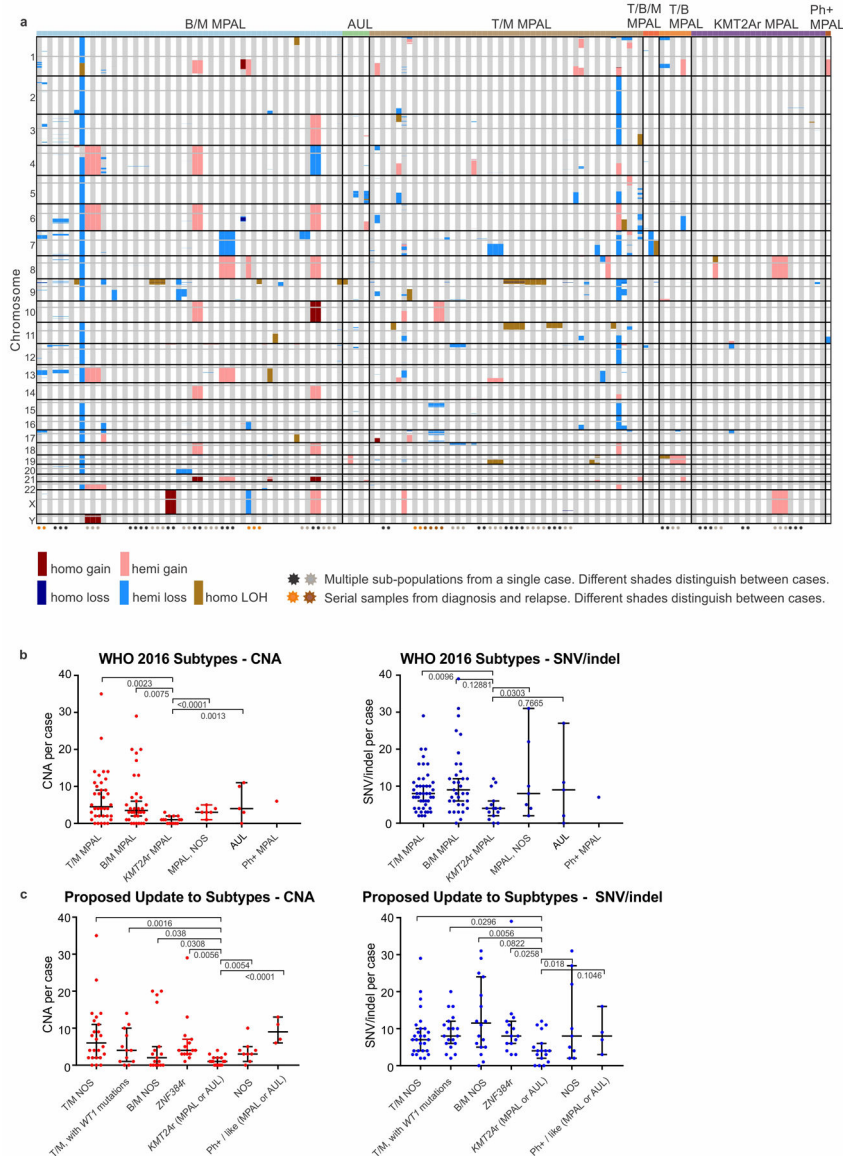
not necessary that the specific markers be present, but only that each population would meet the criteria for B, T, or myeloid leukaemia⁶⁴. **c**, Proposed update to WHO ALAL subtypes incorporating critical newer genomic information (new subtypes in red). **d**, Flow chart of ALAL cohort showing reasons for exclusion and initial diagnosis in cases for which initial ALAL diagnosis occurring at relapse.



Extended Data Fig. 2 | Illustrative immunophenotype and overall survival.

a–e, Representative flow cytometry pseudocolour dot plots and contour plots for five different MPAL cases gated on blast area from CD45 and side scatter area (SSC-A). There

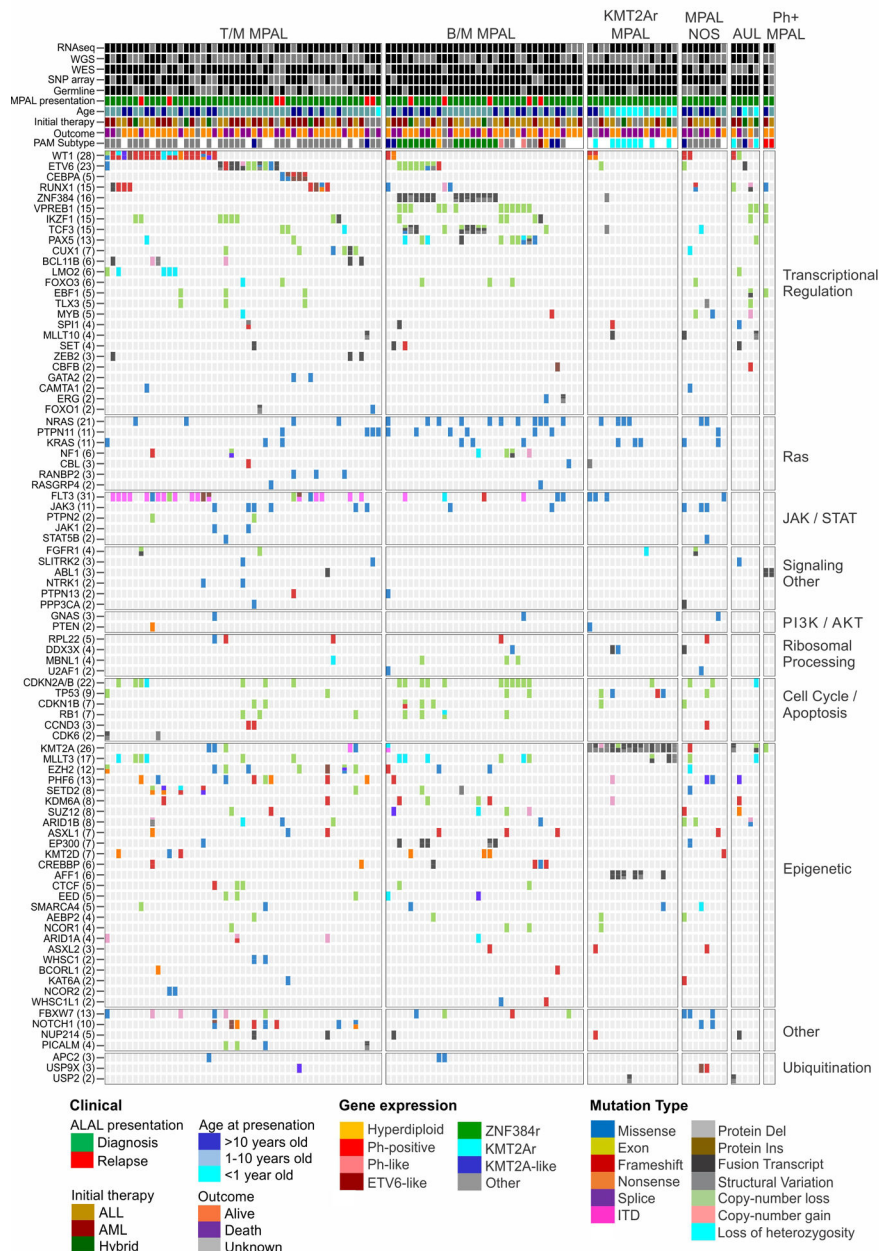
are a wide variety of immunophenotypic patterns, including classic bilineal phenotype (**a**), classic biphenotypic case (**b**), myeloid predominance (**c**), lymphoid predominance (**d**) and complex phenotype with more than two immunophenotypic clones (**e**). **f, g**, Morphology of cells from two patients with MPAL showing both lymphoid (orange arrow) and myeloid (black arrow) morphology. **f**, Bone marrow aspirate stained with myeloperoxidase from a patient with T/M MPAL showing multiple blasts with moderate MPO positivity along with one normal granulocyte. **g**, Peripheral blood haematoxylin and eosin stain from a patient with B/M MPAL. **h–o**, Kaplan–Meier survival curves with overall survival (OS) distributions of patients whose initial diagnosis was MPAL or AUL compared using log-rank tests. At risk numbers for each analysis are provided in the figures. Outcome associations were analysed with the log-rank test. OS according to WHO 2016 subtype (**h**), initial therapy (**i**), *WT1* status within the T/M MPAL cohort (**j**), *ZNF384* status within the B/M MPAL cohort (**k**), RAS pathway alteration within the entire cohort (**l**) and *FLT3* alteration within the entire cohort (**m**). **n**, OS according to initial therapy for patients with B/M MPAL with *ZNF384*. **o**, OS according to initial therapy for patients with B/M MPAL without *ZNF384*. Patients included in this cohort were collected from a range of treatment eras, treatment locations, treatment regimens, and include a range of ages and genomic subtype, limiting the conclusions that may be drawn from these analyses.



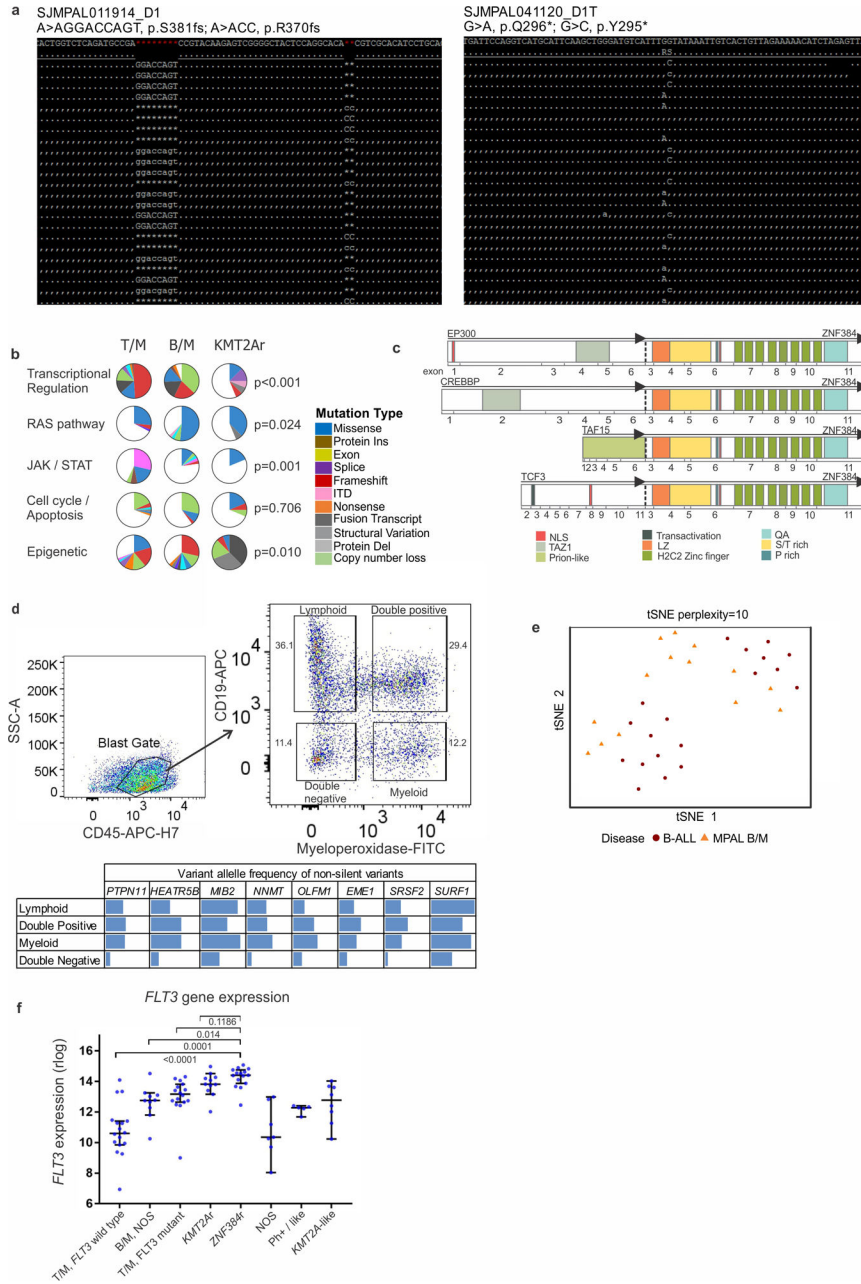
Extended Data Fig. 3 |. Copy number alterations and mutation burden in ALAL.

a. Map showing spectrum of CNAs, visually recapitulating the data shown in Supplementary Table 10. Twenty-seven patients had SNP arrays for multiple subpopulations, annotated by stars. **b.** CNA and non-silent SNVs or indels in ALAL subtypes according the WHO 2016 classification. (CNA, T/M MPAL $n = 36$, B/M MPAL $n = 34$, *KMT2Ar* MPAL $n = 15$, MPAL NOS $n = 7$, AUL $n = 5$, Ph+ MPAL $n = 1$; SNV/indel, T/M MPAL $n = 46$, B/M MPAL $n = 35$, *KMT2Ar* MPAL $n = 15$, MPAL NOS $n = 7$, AUL $n = 5$, Ph+ MPAL $n = 1$) Patients with *KMT2Ar* MPAL have a lower mutation burden than those with T/M MPAL or B/M MPAL. **c.** CNAs and non-silent SNVs or indels in our proposed updated classification system. (CNA, T/M MPAL NOS $n = 24$, T/M MPAL with *WT1* alteration, $n = 12$, B/M MPAL NOS $n = 17$, B/M MPAL with *ZNF384* $n = 15$, *KMT2Ar* MPAL/AUL $n = 17$, MPAL/AUL NOS $n = 9$, Ph+/Ph-like MPAL/AUL $n = 4$; SNV/indel, T/M MPAL NOS $n = 27$, T/M MPAL with *WT1* alteration, $n = 19$, B/M MPAL NOS $n = 18$, B/M MPAL with

ZNF384r *n* = 15, *KMT2Ar* MPAL/AUL *n* = 17, MPAL/AUL NOS *n* = 9, Ph+/Ph-like MPAL/AUL *n* = 4) Data shown as median ± 95% confidence interval. Comparisons assessed by two-sided unpaired *t*-test. One data point is outside the SNV/indel graph for the B/M NOS subtype (1 patient with 167 SNV/indels). SNV/indels per case shown for cases with DNA sequencing completed.



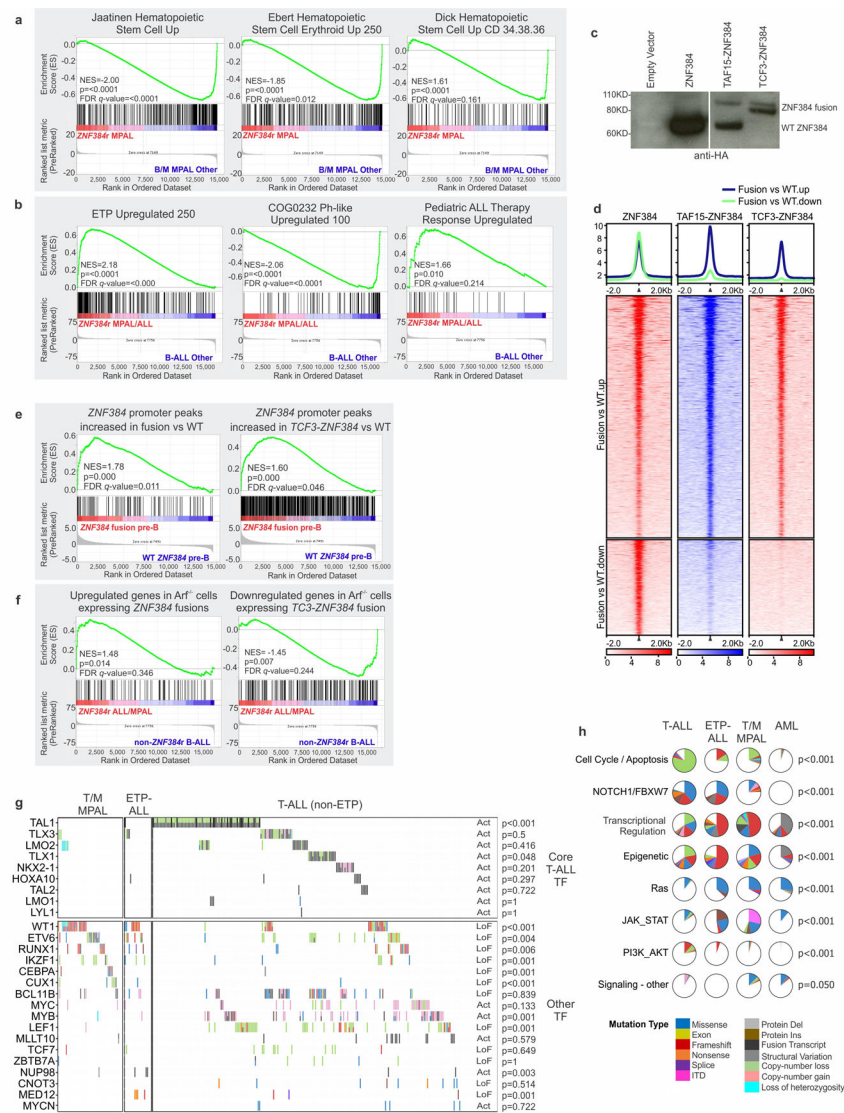
Extended Data Fig. 4 | Complete ALAL mutation oncoprint. Mutation spectrum of ALAL.



Extended Data Fig. 5 |. Features of MPAL genomic analysis.

a, *WT1* alterations were observed in 28 patients, commonly as frameshift mutations (31/47 mutations) in exon 7 (29/47 mutations) and were frequently biallelic. In 16 patients, two clonal alterations were detected, and in 9 patients the locations of the alteration were encompassed by the same sequencing read, providing definitive demonstration that the mutations were in *trans*. Additionally, one patient (SJMPAL043773) had a frameshift mutation and copy number loss of the second allele, while another had a frameshift mutation with copy-neutral loss of heterozygosity (SJMPAL040036). Data are shown for two representative patients with MPAL, showing double-hit mutations on *WT1*. The read alignment view was generated by Samtools²⁴. The reference human genome is on the first

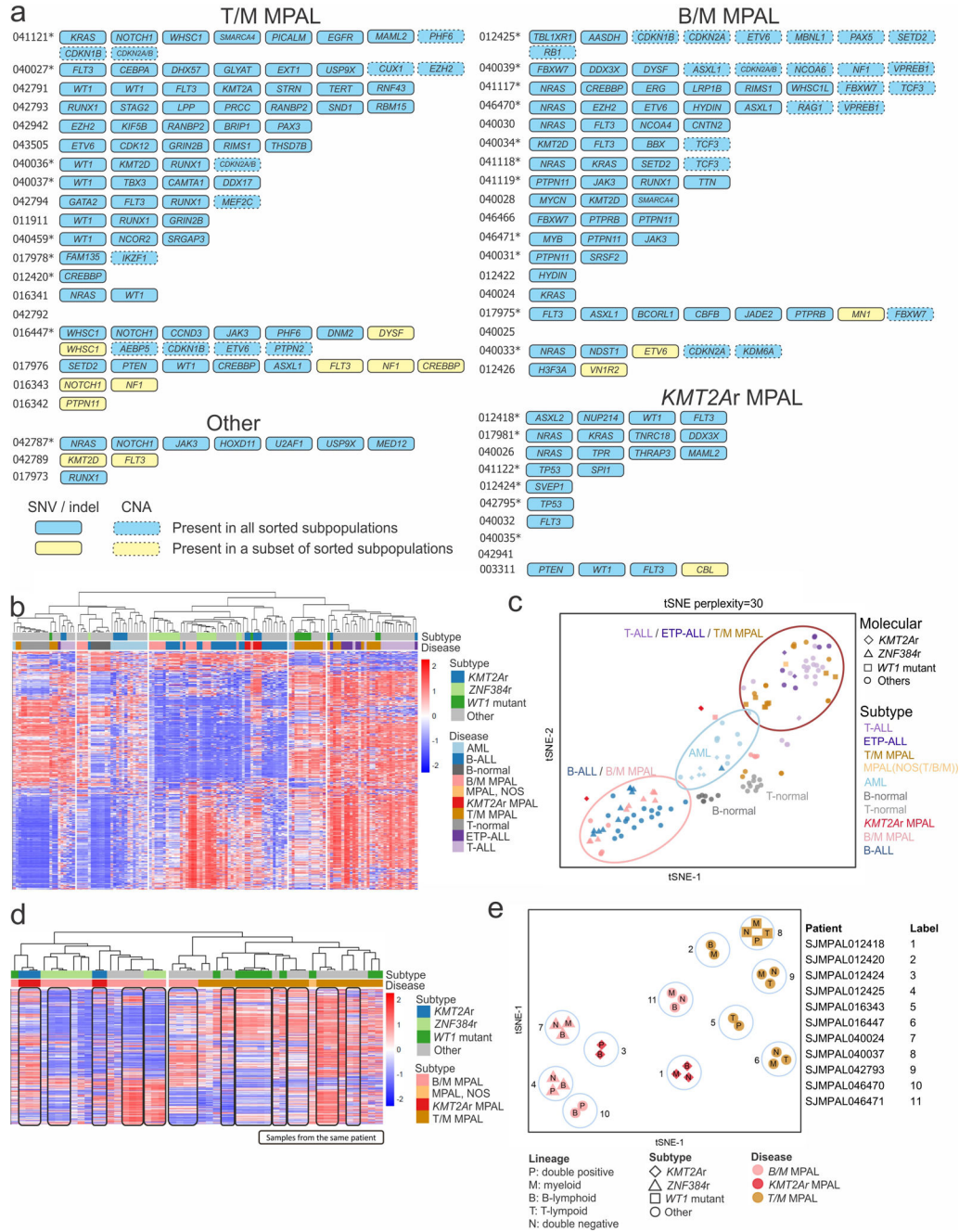
row and sequence reads are aligned below, with matched nucleotides as dots (forward strand match) and commas (reverse strand match) and mismatched ones showing the differences. Alignment gaps are shown as asterisks. Adjacent mutations are shown on different sequence reads, indicating that the mutations are on different alleles. **b**, Frequency of alteration by pathway analysis and MPAL subtype. The similarity of somatic alteration prevalence in different leukaemia subtypes was evaluated by two-sided Fisher's exact test ($n = 100$ biologically independent cases). See also Supplementary Tables 12, 13 for numbers and P values for each gene and pathway. **c**, Schematic representation of *ZNF384* observed in B/M MPAL. NLS, nuclear localization signal; TAZ1, transcriptional adaptor zinc-binding; LZ, leucine rich domain; QA, glycine/alanine repeat. **d**, Fluorescence-activated sorting schema in a representative case with a *ZNF384*, and variant allele frequency of SNVs/indels present in the respective sorted subpopulations, demonstrating genomic similarity of the sorted populations. **e**, tSNE plot of RNA-seq gene expression of all patients with *ZNF384* show no clear segregation of B/M MPAL and B-ALL cases. **f**, *FLT3* gene expression in subtypes of ALAL showing that patients with *ZNF384* B/M MPAL have high levels of *FLT3* expression. As in patients with *KMT2Ar*, this occurs in the absence of *FLT3* alteration in most cases. By contrast, high levels of *FLT3* expression in T/M MPAL appears to be driven by *FLT3* alterations. Data shown as median \pm 95% confidence interval. Comparisons assessed by unpaired t -test, two sided. T/M MPAL *FLT3* wild type $n = 18$, B/M MPAL NOS $n = 10$, T/M MPAL with *FLT3* alteration $n = 16$, B/M MPAL $n = 17$, B/M MPAL with *ZNF384* $n = 15$, *KMT2Ar* MPAL/AUL $n = 11$, MPAL/AUL NOS $n = 7$, Ph+/Ph-like MPAL/AUL $n = 5$, *KMT2A*-like MPAL/AUL $n = 8$.



Extended Data Fig. 6 | *ZNF384r* leukaemia analysis and T/M MPAL mutation comparisons.

a, GSEA of *ZNF384r* B/M MPAL versus non-*ZNF384r* B/M MPAL. HSC gene sets are negatively enriched, supporting the proposed update to MPAL subtypes in which *ZNF384r* leukaemia has distinct biology compared with other B/M MPAL cases^{20,65,66}. **b**, GSEA of all *ZNF384r* cases versus other B-ALL cases indicates immaturity of this subtype compared to B-ALL, with positive enrichment for genes upregulated in ETP-ALL (a stem cell leukaemia), and negative enrichment for genes upregulated in Ph-like ALL in other B-ALL cases. *ZNF384r* acute leukaemia is also enriched for genes upregulated in patients with detectable minimal residual disease at end of induction^{10,51,67}. **c**, Western blot analysis to validate expression of *ZNF384*, *TAF15-ZNF384*, and *TCF3-ZNF384* in transduced *Arf*^{-/-} pre-B cells. Proteins contain an HA epitope tag and are detected by anti-HA antibody. **d**, Heatmap showing the ChIP-seq signal, centred on *ZNF384* peaks, of wild-type (WT) *ZNF384* compared to *TAF15-ZNF384* and *TCF3-ZNF384*. Middle, peaks with increased binding of fusion proteins compared to wild-type. Bottom, peaks with decreased binding of the fusion proteins compared to wild-type. **e**, GSEA showing enrichment of genes whose

promoters exhibit increased binding by ZNF384 fusions in the GEP of ZNF384r versus WT pre-B cells. **f**, GSEA showing similarity of the GEP of mouse pre-B cells expressing ZNF384r to the GEP of human *ZNF384r* leukaemia cells, supporting the notion that perturbation of ZNF384 binding contributes to deregulated gene expression in human *ZNF384r* leukaemia. **g**, Oncoprint of mutations in transcription factor genes across T/M MPAL ($n = 49$), ETP-ALL ($n = 19$) and T-ALL (other) ($n = 245$), showing lack of *TAL1* alterations in T/M MPAL and few core T-ALL transcription factor alterations in T/M MPAL or ETP-ALL. The association of leukemia subtype with individual transcription factor alterations was evaluated using two-sided Fisher exact test. Act, activating mutation; LoF, loss-of-function mutation. **h**, Gene pathway analyses showing similarity of ETP-ALL and T/M MPAL, specifically in frequency of mutations in pathways regulating cell cycle or apoptosis, transcriptional regulation, and signalling pathways. The similarity of somatic alteration prevalence in different leukaemia subtypes was evaluated by two sided Fisher's exact tests in these four subtypes (T/M MPAL $n = 49$, ETP-ALL $n = 19$, non-ETP T-ALL $n = 245$, AML $n = 197$).



Extended Data Fig. 7 |. MPAL subpopulation analysis and methylation analysis.

a. Results of genomic analysis of the 50 patients with sorted subpopulations with WGS or WES results. Listed here are all genes with mutations that were either recurrent in the ALAL cohort or were in known cancer consensus genes⁶⁸. *CNA results also available for sorted subpopulations in these cases. **b–d.** Methylation analysis of MPAL, comparison with acute leukaemia and normal lymphocytes. The top 5,000 probes with highest mean absolute deviation were used to assess the clustering through a 2D t-SNE plot and heatmap with Pearson correlation clustering. See Supplementary Table 37 for sample details. **b.** Heatmap of all samples for methylation analysis showing the general alignment of samples by

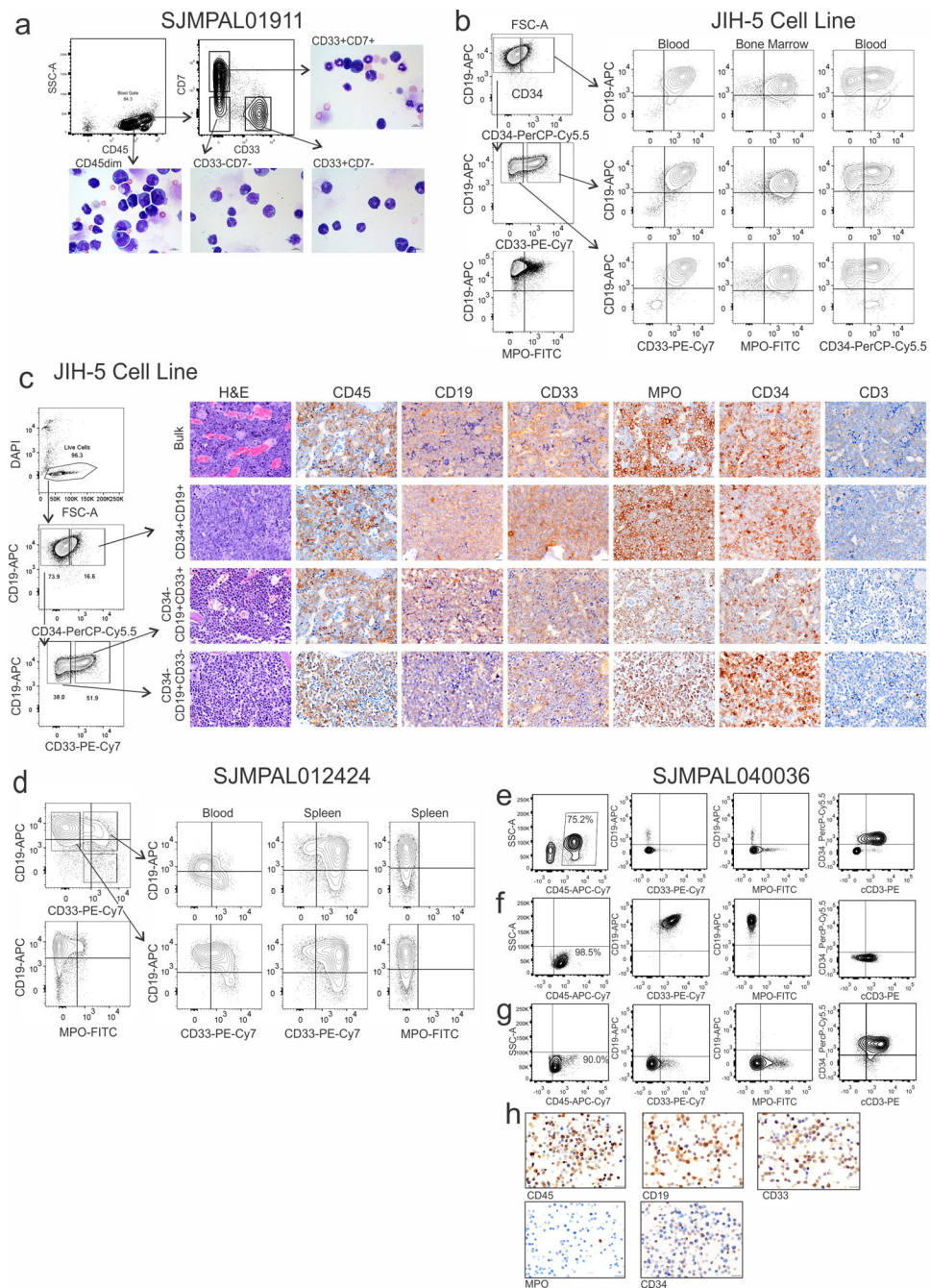
leukaemia phenotype with B/M cases clustering with B-ALL, T/M MPAL, ETP-ALL cases together, and AML cases clustering separately. **c**, tSNE analysis of the same samples as in the top heatmap, showing general alignment by leukaemia phenotype with B/M cases clustering with B-ALL, T/M MPAL, ETP-ALL cases together, and AML cases clustering separately. **d**, Heatmap of all MPAL cases, again showing some clustering by phenotype between B/M and T/M cases. Subpopulations sorted by distinct immunophenotype in MPAL cases clustered tightly with samples from the same patient, rather than with samples with similar phenotype from a different patient. **e**, Methylation analysis of sorted subpopulations from 11 patients with MPAL, demonstrating that methylation profiles cluster by patient and not by immunophenotype lineage.

Author Manuscript

Author Manuscript

Author Manuscript

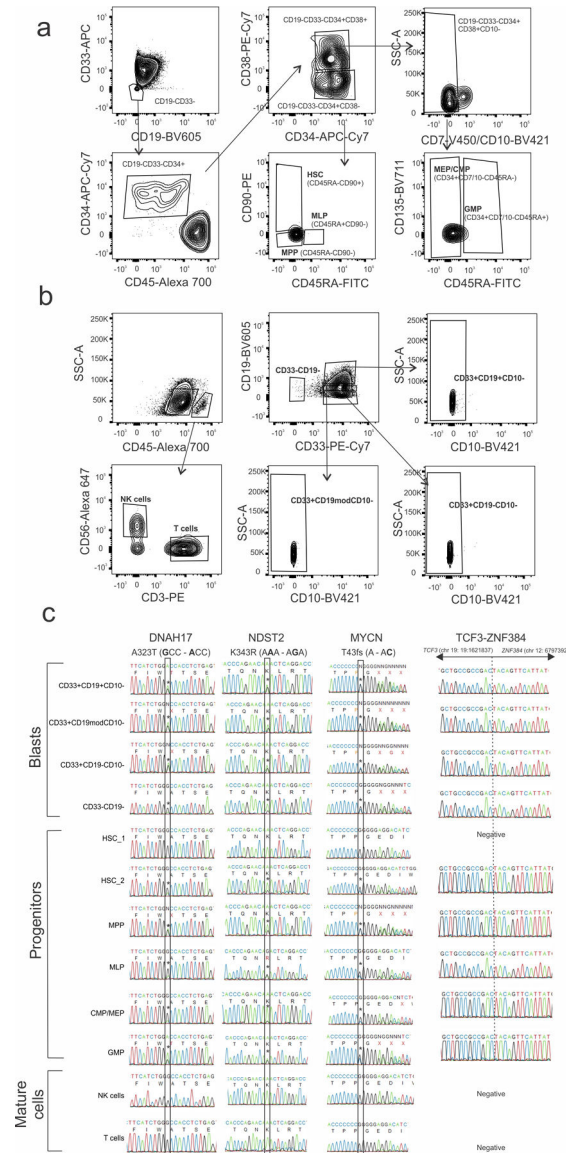
Author Manuscript



Extended Data Fig. 8 | Xenograft analysis.

a, Flow cytometry analysis of bulk leukaemic cells from patient SJMPAL01911 before sorting, and cytopspins from bone marrow samples from representative primary recipient mice transplanted with different leukaemia subpopulations or bulk, confirming the presence of leukaemic blasts from each engrafted population. Scale bars, 10 μ m. **b**, Phenotypic subpopulations from JIH-5 cells in the first column were sorted and injected into NSG-SGM3 mice. Remaining plots show the immunophenotypes of engrafted leukaemia propagated from each sorted subpopulation, demonstrating recapitulation of biphenotypic leukaemia from each. **c**, Flow cytometry analysis of bulk JIH-5 cells prior to sorting (left)

and haematoxylin and eosin staining and IHC labelling for human CD45, CD19, CD33, MPO, CD34 and CD3 in sternum samples from representative primary recipient mice transplanted with different leukaemia subpopulations or bulk. Scale bars, 20 μm . **d**, Phenotypic subpopulations from patient SJMPAL012424 were sorted (left) and injected into irradiated NSG-SGM3 mice. Remaining plots show the immunophenotypes of engrafted leukaemia from each starting subpopulation, demonstrating recapitulation of mixed phenotype leukaemia from two sorted subpopulations. **e**, Flow cytometry analyses of bone marrow cells from an engrafted primary mouse transplanted with leukaemia cells from a patient with T/M MPAL (SJMPAL040036). **f, g**, Flow cytometry analyses of representative engrafted secondary recipient mice transplanted with leukaemia cells from the mouse in **e** showing lineage plasticity with mice developing an emerging CD19⁺CD33⁺ population (**f**) and other mice recapitulating the immunophenotype in the primary recipient (**g**). **h**, IHC labelling for human CD45, CD19, CD33, MPO and CD34 from harvested and fixed spleen cells from a representative secondary recipient mouse showing high expression of CD19 and CD33 and thus confirming the leukaemic lineage plasticity. Scale bars are 20 μm .



Extended Data Fig. 9 | Haematopoietic progenitor cell analysis.

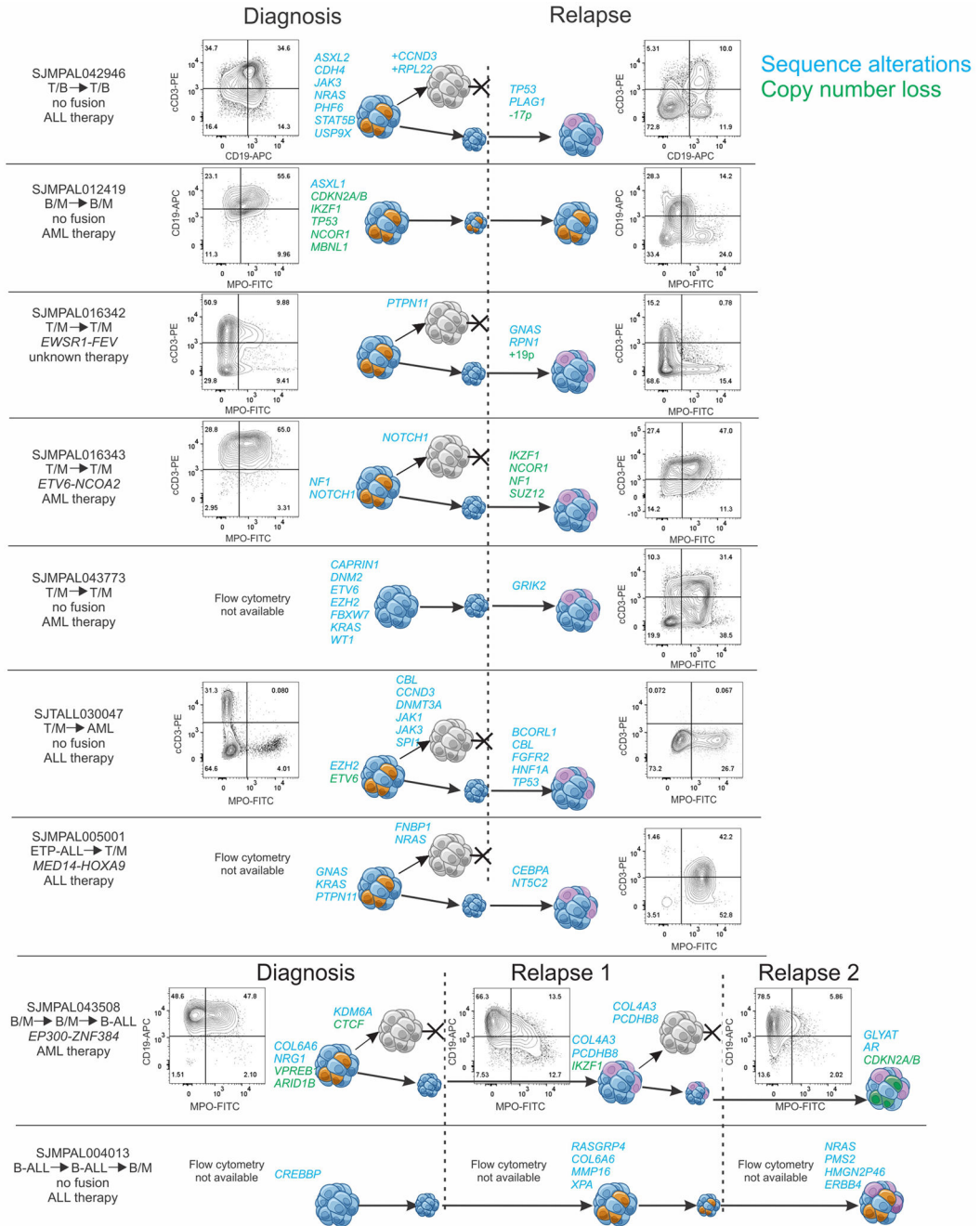
a. Progenitor cell sorting scheme for diagnosis sample from patient SJMPAL040028.

Progenitor populations were all gated on $CD19^-CD33^-CD34^+$ and sorted into HSC ($CD38^-CD34^+CD90^+CD45RA^-$; 2 replicates: HSC_1 and HSC_2); MPP ($CD38^-CD34^+CD90^-CD45RA^-$); MLP ($CD38^-CD34^+CD45RA^+$); megakaryocyte erythroid progenitors/common myeloid progenitors ($CD38^+CD34^+CD7^-CD10^-CD45RA^-$); and granulocyte monocyte progenitor ($CD38^+CD34^+CD7^-CD10^-CD45RA^+$) populations.

b. Blast cell sorting scheme for diagnosis sample from patient SJMPAL040028. Cells were gated on $CD45^{dim}$ and sorted into four different immunophenotypic populations ($CD33^+CD19^+CD10^-$; $CD33^+CD19^{mod}CD10^-$; $CD33^+CD19^-CD10^-$; and $CD33^-CD19^-$).

c. Sanger sequencing electropherograms for the mutational status of *DNAH17*, *NDST2* and *MYCN* and for the fusion *TCF3-ZNF384* in isolated progenitor and blast populations from patient SJMPAL040028 at diagnosis. The identification of somatic missense mutations and

TCF3-ZNF384 fusion in early haematopoietic progenitors indicate that the ambiguous phenotype of MPAL is the result of the acquisition of alterations within an immature haematopoietic progenitor cells.



Extended Data Fig. 10 | Phenotypic and genotypic evolution from diagnosis to relapse. Patients for which diagnosis and relapse pairs with matching non-tumour controls are available show recapitulation of the diagnostic multilineage phenotype in some cases and phenotype plasticity in others. The first column shows the case ID, the leukaemia subtype at diagnosis and then subsequent relapse, the in-frame fusion if present, and initial therapy

received by the patient. Flow plots are shown of cells gated on CD45^{dim} versus SSC-A^{low}. The diagram depicts the inferred clonal evolution based on WES and/or WGS and SNP array data (where available). Mutated genes (either recurrent in ALAL cohort or known cancer consensus genes⁶⁸) are listed. The genes beside the initial diagnostic cell cluster remained present at relapse. The grey cells represent clones that were extinguished with therapy. The genes in the relapse column represent mutations that were gained at relapse.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Thomas B. Alexander^{1,2,37}, Zhaohui Gu^{3,37}, Ilaria Iacobucci^{3,37}, Kirsten Dickerson³, John K. Choi³, Beisi Xu⁴, Debbie Payne-Turner³, Hiroki Yoshihara³, Mignon L. Loh⁵, John Horan⁶, Barbara Buldini⁷, Giuseppe Basso⁷, Sarah Elitzur⁸, Valerie de Haas⁹, C. Michel Zwaan¹⁰, Allen Yeoh¹¹, Dirk Reinhardt¹², Daisuke Tomizawa¹³, Nobutaka Kiyokawa¹⁴, Tim Lammens¹⁵, Barbara De Moerloose¹⁵, Daniel Catchpoole¹⁶, Hiroki Hori¹⁷, Anthony Moorman¹⁸, Andrew S. Moore¹⁹, Ondrej Hrusak²⁰, Soheil Meshinchi^{21,22}, Etan Orgel²³, Meenakshi Devidas²⁴, Michael Borowitz²⁵, Brent Wood²⁶, Nyla A. Heerema²⁷, Andrew Carroll²⁸, Yung-Li Yang²⁹, Malcolm A. Smith³⁰, Tanja M. Davidsen³¹, Leandro C. Hermida³², Patee Gesuwan³², Marco A. Marra³³, Yussanne Ma³³, Andrew J. Mungall³³, Richard A. Moore³³, Steven J. M. Jones³³, Marcus Valentine³⁴, Laura J. Janke³, Jeffrey E. Rubnitz¹, Ching-Hon Pui¹, Liang Ding⁴, Yu Liu⁴, Jinghui Zhang⁴, Kim E. Nichols¹, James R. Downing³, Xueyuan Cao³⁵, Lei Shi³⁵, Stanley Pounds³⁵, Scott Newman⁴, Deqing Pei⁴, Jaime M. Guidry Auvil³², Daniela S. Gerhard³², Stephen P. Hunger³⁶, Hiroto Inaba^{1,*}, and Charles G. Mullighan^{3,*}

Affiliations

¹Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA. ²Department of Pediatrics, University of North Carolina, Chapel Hill, NC, USA. ³Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁴Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA. ⁵Department of Pediatrics, Benioff Children's Hospital and the Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA. ⁶Aflac Cancer and Blood Disorders Center, Children's Healthcare of Atlanta and Emory University School of Medicine, Department of Pediatrics, Atlanta, GA, USA. ⁷Department of Women and Child Health, Hemato-Oncology Division, University of Padova, Padova, Italy. ⁸Pediatric Hematology-Oncology, Schneider Children's Medical Center, Sackler Faculty of Medicine, Tel Aviv University, Israel. ⁹Prinses Maxima Centre, Dutch Childhood Oncology Group Laboratory, Utrecht, The Netherlands. ¹⁰Department of Pediatric Oncology, Erasmus MC-Sophia, Rotterdam and Princess Máxima Center, Utrecht, The Netherlands. ¹¹Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹²Universitäts-Klinikum,

Essen, Germany. ¹³Division of Leukemia and Lymphoma, Children's Cancer Center, National Center for Child Health and Development, Tokyo, Japan. ¹⁴Department of Pediatric Hematology and Oncology Research, National Research Institute for Child Health and Development, Tokyo, Japan. ¹⁵Department of Pediatric Hematology-Oncology and Stem Cell Transplantation, Ghent University Hospital, Ghent, Belgium. ¹⁶The Tumour Bank CCRU, The Kids Research Institute, The Children's Hospital at Westmead, Westmead, NSW, Australia ¹⁷Department of Pediatrics, Mie University, Tsu, Japan. ¹⁸Wolfson Childhood Cancer Centre, Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, UK. ¹⁹The University of Queensland Diamantina Institute & Children's Health, Brisbane, Queensland, Australia. ²⁰Department of Paediatric Haematology and Oncology, 2nd Faculty of Medicine, Charles University and University Hospital Motol, Prague, Czech Republic. ²¹Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA, USA. ²²Children's Oncology Group, Arcadia, CA, USA. ²³Children's Center for Cancer and Blood Disease, Children's Hospital Los Angeles, Los Angeles, CA, USA. ²⁴University of Florida, Gainesville, FL, USA. ²⁵Johns Hopkins Medical Institutions, Baltimore, MD, USA. ²⁶University of Washington, Seattle, WA, USA. ²⁷The Ohio State University School of Medicine, Columbus, OH, USA. ²⁸University of Alabama at Birmingham, Birmingham, AL, USA. ²⁹Department of Laboratory Medicine and Pediatrics, National Taiwan University Hospital, College of Medicine, National Taiwan University, Taipei, Taiwan. ³⁰Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, MD, USA. ³¹Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD, USA. ³²Office of Cancer Genomics, National Cancer Institute, Bethesda, MD, USA. ³³Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. ³⁴Cytogenetics Shared Resource, St. Jude Children's Research Hospital, Memphis, TN, USA. ³⁵Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA. ³⁶Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. ³⁷These authors contributed equally: Thomas B. Alexander, Zhaohui Gu, Ilaria Iacobucci.

Acknowledgements

We thank the Biorepository, the Genome Sequencing Facility of the Hartwell Center for Bioinformatics and Biotechnology, and the Flow Cytometry and Cell Sorting core facility and Cytogenetics core facility of St. Jude Children's Research Hospital (SJCRH). This work was supported in part by the American Lebanese Syrian Associated Charities of SJCRH, Cookies for Kids Cancer (to H.I.), St. Baldrick's Foundation Robert J. Arceci Innovation Award and Henry Schueler 41&9 Foundation (to C.G.M.), SJCRH Physician Scientist Training Program Fellowship (to T.B.A.), the National Cancer Institute grants P30 CA021765 (SJCRH Cancer Center Support Grant), Chair's grant and supplement to support the COG ALL TARGET project), U10 CA98413 (to the COG Statistical Center), U24 CA114766 (to COG; Specimen Banking), and Outstanding Investigator Award R35 CA197695 (to C.G.M.). The results published here are in part based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments initiative of the NCI (<http://ocg.cancer.gov/programs/target>). This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract No. HHSN261200800001E (to C.G.M. and Michael Smith Genome Sciences Centre). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human

Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. We acknowledge Canada's Michael Smith Genome Sciences Centre, Vancouver, Canada for library construction and sequencing. A full list of funders of infrastructure and research supporting the services accessed is available at www.bcgsc.ca/about/funding_support.

References

- Gerr H et al. Acute leukaemias of ambiguous lineage in children: characterization, prognosis and therapy recommendations. *Br. J. Haematol* 149, 84–92 (2010). [PubMed: 20085575]
- Rubnitz JE et al. Acute mixed lineage leukemia in children: the experience of St Jude Children's Research Hospital. *Blood* 113, 5083–5089 (2009). [PubMed: 19131545]
- Matutes E et al. Mixed-phenotype acute leukemia: clinical and laboratory features and outcome in 100 patients defined according to the WHO 2008 classification. *Blood* 117, 3163–3171 (2011). [PubMed: 21228332]
- Maude SL et al. Efficacy of JAK/STAT pathway inhibition in murine xenograft models of early T-cell precursor (ETP) acute lymphoblastic leukemia. *Blood* 125, 1759–1767 (2015). [PubMed: 25645356]
- Zhang J et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 481, 157–163 (2012). [PubMed: 22237106]
- Swerdlow SH et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Revised 4th Edition) (IARC, Lyon, 2017).
- Richards S et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med* 17, 405–424 (2015). [PubMed: 25741868]
- Yasuda T et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat. Genet* 48, 569–574 (2016). [PubMed: 27019113]
- Williams RT, Roussel MF & Sherr CJ Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* 103, 6688–6693 (2006). [PubMed: 16618932]
- Coustan-Smith E et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* 10, 147–156 (2009). [PubMed: 19147408]
- Liu Y et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet* 49, 1211–1218 (2017). [PubMed: 28671688]
- Bolouri H et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med* 24, 103–112 (2018). [PubMed: 29227476]
- Mansour MR et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* 346, 1373–1377 (2014). [PubMed: 25394790]
- Ma X et al. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun* 6, 6604 (2015). [PubMed: 25790293]
- Ping N et al. Establishment and genetic characterization of a novel mixed-phenotype acute leukemia cell line with EP300-ZNF384 fusion. *J. Hematol. Oncol* 8, 100 (2015). [PubMed: 26293203]
- Iacobucci I et al. Truncating erythropoietin receptor rearrangements in acute lymphoblastic leukemia. *Cancer Cell* 29, 186–200 (2016). [PubMed: 26859458]
- Griffith M et al. Comprehensive genomic analysis reveals FLT3 activation and a therapeutic strategy for a patient with relapsed adult B-lymphoblastic leukemia. *Exp. Hematol* 44, 603–613 (2016). [PubMed: 27181063]
- Roberts KG et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N. Engl. J. Med* 371, 1005–1015 (2014). [PubMed: 25207766]
- Conter V et al. Early T-cell precursor acute lymphoblastic leukaemia in children treated in AIEOP centres with AIEOP-BFM protocols: a retrospective analysis. *Lancet Haematol* 3, e80–e86 (2016). [PubMed: 26853647]

20. Notta F et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* 351, aab2116 (2016). [PubMed: 26541609]
21. Lindsley RC et al. Prognostic mutations in myelodysplastic syndrome after stem-cell transplantation. *N. Engl. J. Med* 376, 536–547 (2017). [PubMed: 28177873]
22. Papaemmanuil E et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med* 374, 2209–2221 (2016). [PubMed: 27276561]
23. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009). [PubMed: 19451168]
24. Li H et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009). [PubMed: 19505943]
25. DePristo MA et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet* 43, 491–498 (2011). [PubMed: 21478889]
26. Wang J et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* 8, 652–654 (2011). [PubMed: 21666668]
27. Kent WJ BLAT—the BLAST-like alignment tool. *Genome Res* 12, 656–664 (2002). [PubMed: 11932250]
28. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013). [PubMed: 23104886]
29. Edgren H et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* 12, R6 (2011). [PubMed: 21247443]
30. Anders S, Pyl PT & Huber W HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169 (2015). [PubMed: 25260700]
31. Anders S & Huber W Differential expression analysis for sequence count data. *Genome Biol* 11, R106 (2010). [PubMed: 20979621]
32. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550 (2014). [PubMed: 25516281]
33. 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). [PubMed: 16199517]
34. Harrow J et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22, 1760–1774 (2012). [PubMed: 22955987]
35. Law CW, Chen Y, Shi W & Smyth GK voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15, R29 (2014). [PubMed: 24485249]
36. Wang K et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17, 1665–1674 (2007). [PubMed: 17921354]
37. Yau C et al. A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biol* 11, R92 (2010). [PubMed: 20858232]
38. Gu Z & Mullighan CG ShinyCNV: a Shiny/R application to view and annotate DNA copy number variations. *Bioinformatics* (2018). 10.1093/bioinformatics/bty546
39. Morris TJ et al. ChAMP: 450k chip analysis methylation pipeline. *Bioinformatics* 30, 428–430 (2014). [PubMed: 24336642]
40. Aryee MJ et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369 (2014). [PubMed: 24478339]
41. Zhou W, Laird PW & Shen H Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 45, e22 (2017). [PubMed: 27924034]
42. Nordlund J et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol* 14, r105 (2013). [PubMed: 24063430]
43. Teschendorff AE et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* 4, e8274 (2009). [PubMed: 20019873]

44. Johnson WE, Li C & Rabinovic A Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127 (2007). [PubMed: 16632515]
45. van der Maaten L & Hinton G Visualizing data using t-SNE. *J. Mach. Learn. Res* 9, 2579–2605 (2008).
46. Mullighan CG et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446, 758–764 (2007). [PubMed: 17344859]
47. Andersson AK et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet* 47, 330–337 (2015). [PubMed: 25730765]
48. Gu Z et al. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nat. Commun* 7, 13331 (2016). [PubMed: 27824051]
49. Zhang J et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat. Genet* 48, 1481–1489 (2016). [PubMed: 27776115]
50. Den Boer ML et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol* 10, 125–134 (2009). [PubMed: 19138562]
51. Mullighan CG et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med* 360, 470–480 (2009). [PubMed: 19129520]
52. Lilljebjörn H et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nat. Commun* 7, 11790 (2016). [PubMed: 27265895]
53. Harvey RC et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood* 116, 4874–4884 (2010). [PubMed: 20699438]
54. Yeoh EJ et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133–143 (2002). [PubMed: 12086872]
55. Tibshirani R, Hastie T, Narasimhan B & Chu G Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl Acad. Sci. USA* 99, 6567–6572 (2002). [PubMed: 12011421]
56. McKenna A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297–1303 (2010). [PubMed: 20644199]
57. Aldiri I et al. The dynamic epigenetic landscape of the retina during development, reprogramming, and tumorigenesis. *Neuron* 94, 550–568.e10 (2017). [PubMed: 28472656]
58. Kharchenko PV, Tolstorukov MY & Park PJ Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol* 26, 1351–1359 (2008). [PubMed: 19029915]
59. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137 (2008). [PubMed: 18798982]
60. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010). [PubMed: 20110278]
61. Shlush LI et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* 547, 104–108 (2017). [PubMed: 28658204]
62. 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). [PubMed: 20686503]
63. Roberts KG et al. High frequency and poor outcome of Philadelphia chromosome-like acute lymphoblastic leukemia in adults. *J. Clin. Oncol* 35, 394–401 (2017). [PubMed: 27870571]
64. Arber DA et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127, 2391–2405 (2016). [PubMed: 27069254]
65. Jaatinen T et al. Global gene expression profile of human cord blood-derived CD133+ cells. *Stem Cells* 24, 631–641 (2006). [PubMed: 16210406]
66. Novershtern N et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309 (2011). [PubMed: 21241896]

67. Flotho C et al. Genes contributing to minimal residual disease in childhood acute lymphoblastic leukemia: prognostic significance of CASP8AP2. *Blood* 108, 1050–1057 (2006). [PubMed: 16627760]
68. Futreal PA et al. A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183 (2004). [PubMed: 14993899]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

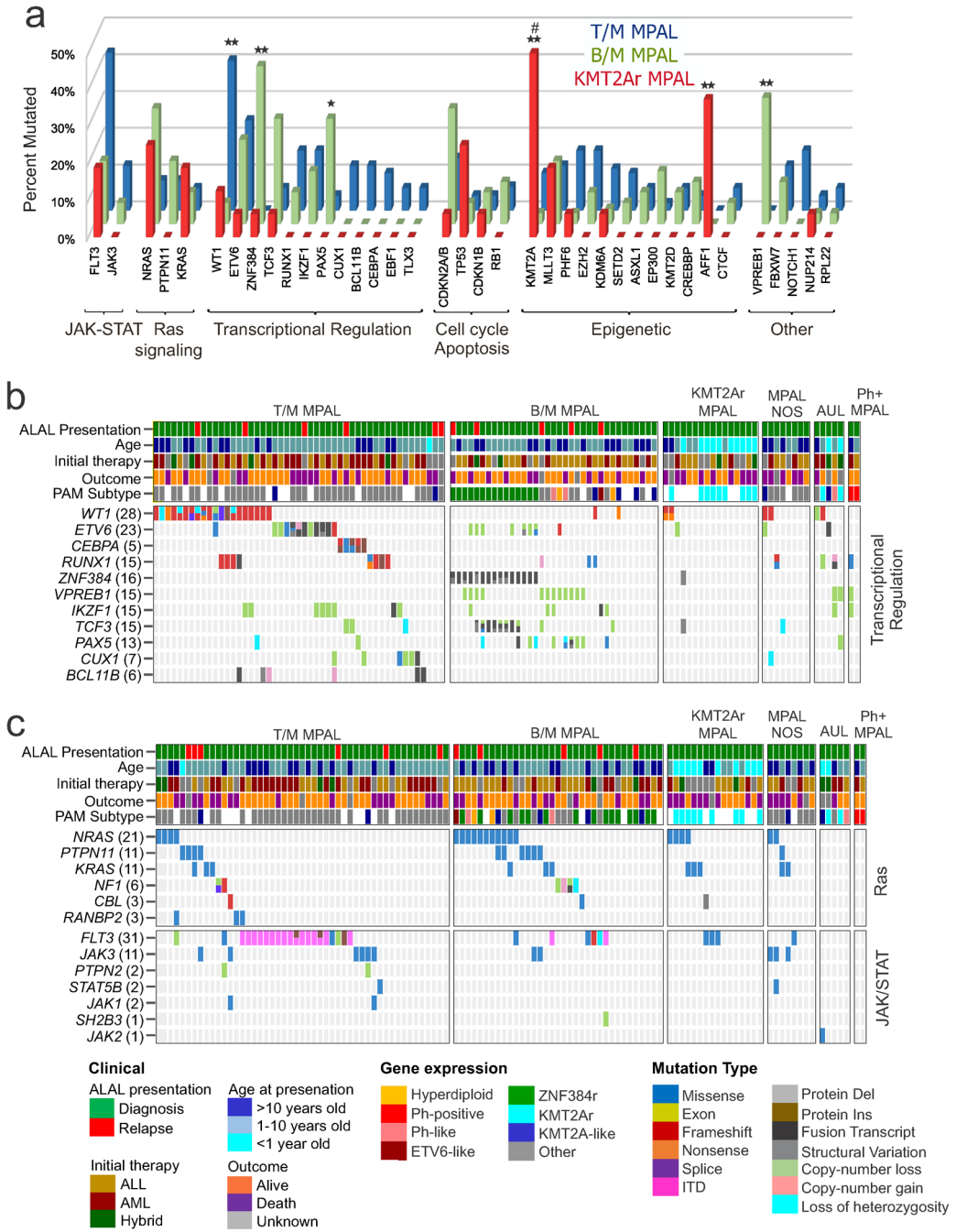


Fig. 1 | Genomic overview of ALAL.

a, Distribution of the most frequently altered genes by MPAL subtype. Frequency of mutations in the different MPAL subtypes were compared by two-sided Fisher exact tests; $**P < 0.001$, $*0.001 < P < 0.01$ (see Supplementary Table 13 for numbers for each group and P values for each gene). #*KMT2A* alterations were present in all cases in the *KMT2Ar* subgroup. **b**, Oncoprint of mutations in transcriptional regulation and cell cycle/apoptosis pathways. **c**, Oncoprint of mutations in signalling pathways. Mutations altering genes involved in transcription and signalling pathways in these subtypes are distinct.

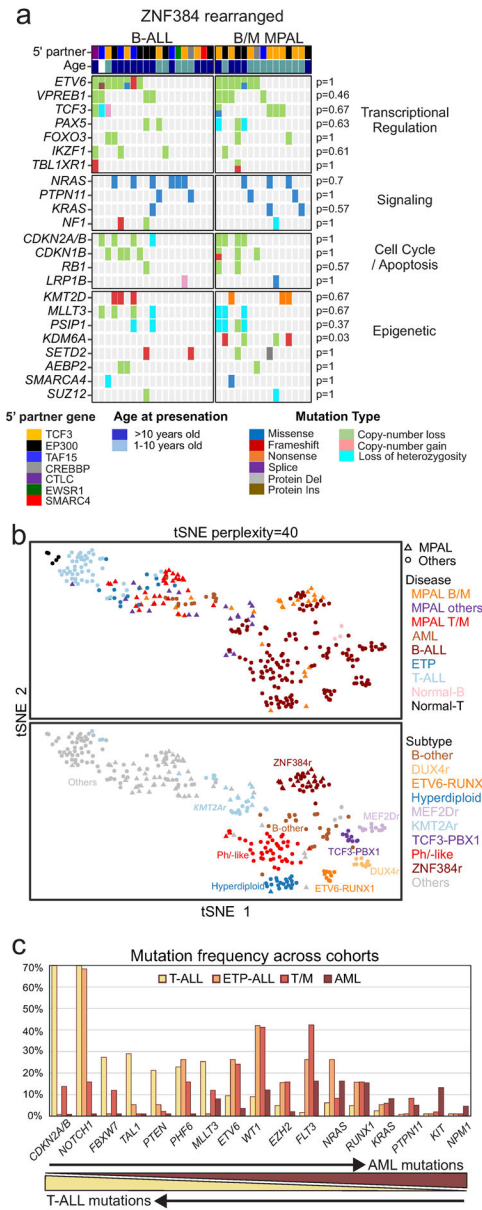


Fig. 2 | Genomic comparisons across leukaemia subtypes.

a, Mutations observed in *ZNF384* B-ALL ($n = 19$) and *ZNF384* B/M MPAL ($n = 15$), showing similar mutational profile between the two phenotypically defined subtypes. **b**, tSNE plot of top 1,000 variably expressed genes of ALAL, B-ALL, T-ALL, ETP-ALL, AML, and normal lymphocytes, showing that B/M MPAL has a GEP more similar to B-ALL than AML, and T/M MPAL more similar to ETP-ALL than AML. *ZNF384* cases cluster together, without separation based upon B/M MPAL or B-ALL phenotype. Cases in the ALAL-other category, including *KMT2Ar* MPAL, AUL, and MPAL NOS, are intermixed across the transcriptional continuum, primarily between AML and B-ALL clusters. **c**, Depiction of the frequency of mutations of the five most frequently altered genes from each disease cohort, demonstrating that T/M MPAL ($n = 49$) and ETP-ALL ($n = 19$) share a high

frequency of mutations in *ETV6*, *WT1*, *EZH2* and *FLT3*, while lacking the most characteristically mutated genes in T-ALL ($n = 245$) and AML ($n = 197$).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

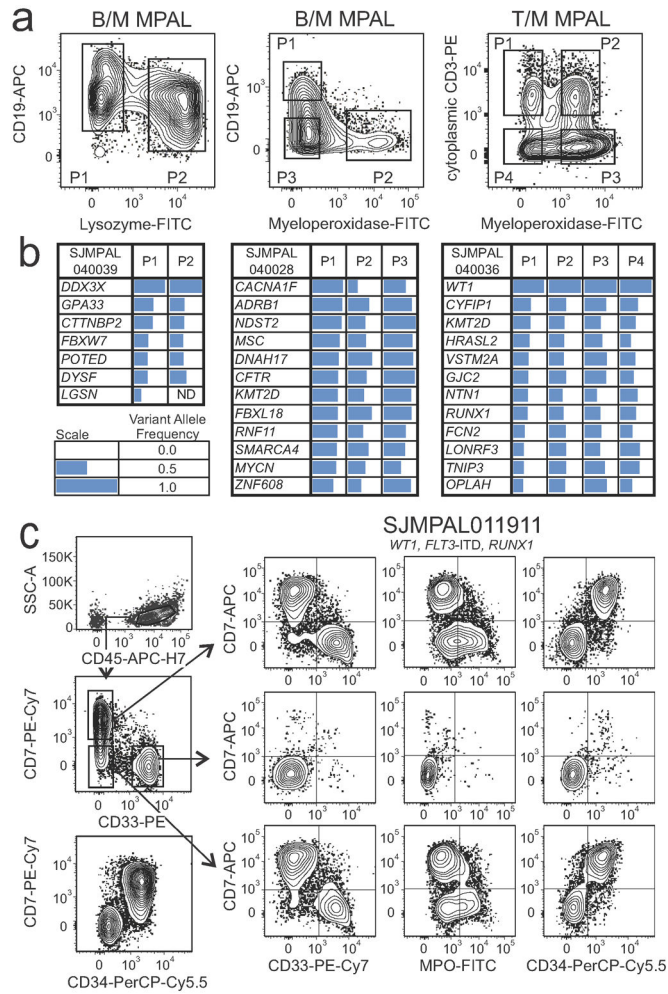


Fig. 3 |. Plasticity is independent of mutation.

a, Flow cytometric scatter plots of representative cases of MPAL, showing the primary lymphoid marker (CD19 or cytoplasmic CD3) and myeloid marker (myeloperoxidase or lysozyme) used for sorting subpopulations. P1–P4 represent sorted subpopulations subjected to DNA sequencing. **b**, Variant allele frequency (VAF, represented by length of blue bar) from each of the purified populations in **a** demonstrating concordance of mutational VAF of SNV or indel between distinct immunophenotypically defined subpopulations. **c**, Phenotypic subpopulations from case SJMPAL011911 were sorted (first column) and injected into irradiated NSG-SGM3 mice. Key gene alterations are shown above the flow scatter plots (ITD, internal tandem duplication). Remaining plots show the immunophenotype of harvested bone marrow of engrafted leukaemia from each starting subpopulation, demonstrating recapitulation of mixed phenotype leukaemia from two sorted subpopulations. The third subpopulation (CD7⁻, CD33⁺) also engrafted with hCD45⁺ cells and morphologic leukaemia, but with an undifferentiated immunophenotype. ND=not detected

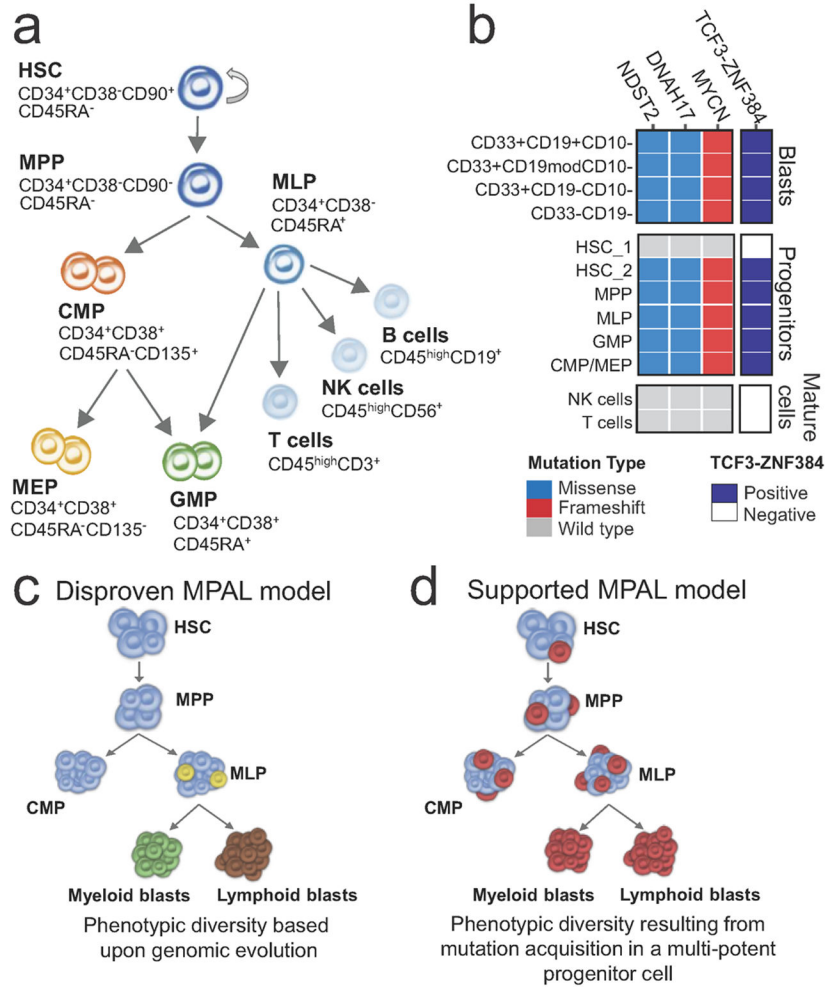


Fig. 4 | Model of MPAL leukaemogenesis.

a, Schematic and simplified representation of human haematopoietic hierarchy showing HSCs, multipotent progenitors (MPPs), multilymphoid progenitors (MLPs), megakaryocyte erythroid progenitors (MEPs), common myeloid progenitors (CMPs), granulocyte monocyte progenitors (GMPs), and mature lymphocytes: B cells, T cells, and NK cells. **b**, Summary of the presence of *ZNF384* and additional somatic alterations in isolated stem/progenitor, mature and blast cell populations showing the presence of each alteration throughout haematopoietic development. **c, d**, Potential models of bilineal MPAL leukaemogenesis. Different colours represent clones with different genomic alterations. **c**, A model of MPAL in which phenotypic divergence is driven by acquisition of secondary genomic alterations (yellow and green cells), which is inconsistent with the results of the current study. **d**, A model of MPAL showing that necessary and sufficient mutations are acquired in an early haematopoietic progenitor that retains myeloid and lymphoid potential, thus propagating similar mutation profiles in the different phenotypes. The results of the current study support this model of leukaemogenesis.