Epigenetic regulator genes direct lineage switching in *MLL-AF4* leukaemia

Ricky Tirtakusuma[1], Katarzyna Szoltysek[1,2,3], Paul Milne[4], Vasily V Grinev[5], Anetta Ptasinska[6], Claus Meyer[7], Sirintra Nakjang[1], Jayne Y Hehir-Kwa[2], Daniel Williamson[1], Pierre Cauchy[6], Salam A Assi[6], Maria R Imperato[6], Fotini Vogiatzi[9], Shan Lin[10], Mark Wunderlich[10], Janine Stutterheim[2], Alexander Komkov[8], Elena Zerkalenkova[8], Paul Evans[11], Hesta McNeill[1], Alex Elder[1], Natalia Martinez-Soria[1], Sarah E Fordham[1], Yuzhe Shi[1], Lisa J Russell[1], Deepali Pal[1], Alex Smith[12], Zoya Kingsbury[13], Jennifer Becq[13], Cornelia Eckert[14], Oskar A Haas[15], Peter Carey[16], Simon Bailey[1,16], Roderick Skinner[1,16], Natalia Miakova[8], Matthew Collin[4], Venetia Bigley[4], Muzlifa Haniffa[17,18,19], Rolf Marschalek[7], Christine J Harrison[1], Catherine A Cargo[11], Denis Schewe[9], Yulia Olshanskaya[8], Michael J Thirman[20], Peter N Cockerill[6], James C Mulloy[10], Helen J Blair[1], Josef Vormoor[1,2], James M Allan[1], Constanze Bonifer[6*], Olaf Heidenreich[1,2*†], Simon Bomken[1,16*†]

Author Affiliations

[1]Wolfson Childhood Cancer Research Centre, Centre for Cancer, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK

[2]Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands

[3]Maria Sklodowska-Curie Institute - Oncology Center, Gliwice Branch, Gliwice, Poland

[4]Translational and Clinical Research Institute, Newcastle University, Framlington Place, Newcastle upon Tyne, UK

[5]Department of Genetics, the Faculty of Biology, Belarusian State University, 220030 Minsk, Republic of Belarus.

[6]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

[7]Institute of Pharmaceutical Biology/DCAL, Goethe-University, Frankfurt/Main, Germany

[8]Dmitry Rogachev National Research Center of Pediatric Hematology, Oncology, and Immunology; Moscow, Russia

[9]Pediatric Hematology/Oncology, ALL-BFM Study Group, Christian Albrechts University Kiel and University Hospital Schleswig-Holstein, Campus Kiel, Germany

[10]Experimental Hematology and Cancer Biology, Cancer and Blood Disease Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

[11]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, UK

[12]Epidemiology and Cancer Statistics Group, University of York, York, United Kingdom

[13]Illumina Cambridge Ltd., Great Abington, UK

[14]Department of Pediatric Oncology/Hematology, Charité Universitätsmedizin Berlin, Berlin, Germany

[15]Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna, Austria

[16]Department of Paediatric Haematology and Oncology, The Great North Children's Hospital, Newcastle upon Tyne, UK

[17]Biosciences Institute, Newcastle University, Framlington Place, Newcastle upon Tyne, UK

[18]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton UK

[19]Department of Dermatology and Newcastle NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne

[20]Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago, USA

***Co-senior authors**

†**Co-corresponding Authors.** Correspondance to: **s.n.bomken@ncl.ac.uk; O.T.Heidenreich@prinsesmaximacentrum.nl**

Summary

The fusion gene *MLL-AF4* defines a high-risk subtype of pro-B acute lymphoblastic leukaemia. However, relapse can be associated with a switch from acute lymphoblastic to acute myeloid leukaemia. Here we show that these myeloid relapses share oncogene fusion breakpoints with their matched lymphoid presentations and can originate in either early, multipotent progenitors or committed B-cell precursors. Lineage switching is linked to substantial changes in chromatin accessibility and rewiring of transcriptional programmes indicating that the execution and maintenance of lymphoid lineage differentiation is impaired. We show that this subversion is recurrently associated with the dysregulation of repressive chromatin modifiers, notably the nucleosome remodelling and deacetylation complex, NuRD. In addition to mutations, we show differential expression or alternative splicing of NuRD members and other genes is able to reprogram the B lymphoid into a myeloid gene regulatory network. Lineage switching in *MLL-AF4* leukaemia is therefore driven and maintained by defunct epigenetic regulation.

Introduction

Translocation of Mixed Lineage Leukaemia (*MLL*) with one of over 130 alternative partner genes is a recurrent cytogenetic finding in both acute myeloid and lymphoblastic leukaemias and is generally associated with poor prognosis (Pieters *et al.*, 2007; Moorman *et al.*, 2010; Meyer *et al.*, 2018). Amongst the most common translocations is t(4;11)(q21;q23), forming the *MLL-AF4* (also known as *KMT2A-AFF1*) fusion gene. Uniquely amongst *MLL* rearrangements (*MLL*r), *MLL-AF4* is almost exclusively associated with pro-B cell acute lymphoblastic leukaemia and is prototypical of infant acute lymphoblastic leukaemia (ALL) where it carries a very poor prognosis (Meyer *et al.*, 2018). However, despite this general lymphoid presentation, *MLL-AF4* leukaemias have an intriguing characteristic - that of lineage switched relapses. Lineage switch acute leukaemias (LSALs) lose their lymphoid specific features and gain myeloid phenotype upon relapse (Jiang *et al.*, 2005; Germano *et al.*, 2006; Park *et al.*, 2011; Carulli *et al.*, 2012; Rossi *et al.*, 2012; Ivanov *et al.*, 2013). Alternatively, *MLL-AF4* leukaemias may harbour distinct lymphoid and myeloid populations at the same time, thus classifying as mixed phenotype acute leukaemias (MPALs) of the bilineage subtype.

In order to understand the molecular basis of lineage promiscuity and switching, we examined a unique cohort of *MLL-AF4*-positive LSAL presentation/relapse pairs and MPALs. We demonstrate that disruption of the epigenetic machinery, including the nucleosome remodelling and deacetylation complex (NuRD), is associated with the loss of lymphoid restriction. Lineage switch is then enacted through redistribution of transcription factor binding and chromatin reorganisation. Whilst identified here within this rare clinical context, our findings bare relevance for our understanding of the transforming capacity of *MLL-AF4*, and how this oncoprotein imposes lineage determination on haematopoietic precursor cells. Furthermore, given the high-risk nature of this disease, we provide a novel insight into factors which may prove critical to the effective implementation of lineage specific, epitope-directed therapies such as chimeric antigen receptor T-cell (CAR-T) cell or bi-specific T-cell engaging antibody (BiTE) approaches.

Results

Characterisation of *MLL-AF4* acute leukaemias with lineage switch

To characterize lineage promiscuity in *MLL-AF4* leukaemia and the underlying molecular mechanisms, we collected a cohort of ten cases of *MLL-AF4* ALL comprising 6 infant, 2 paediatric and 2 adult patients who had relapsed with acute myeloid leukaemia (AML). Amongst these, one infant patient (LS10) had relapsed following B-lineage directed blinatumomab treatment (Table S1). The time to relapse ranged from 3 to 48 months. Seven patients within the cohort subsequently died. Lineage switch was defined as loss of expression of B lymphoid antigens (CD19, CD22, CD79A) with concomitant gain of expression of myeloid antigens (CD33, CD117/KIT, CD64/FCGR1A) and/or an unequivocal change in morphology to AML (Figure 1A and Table S1). In addition, we studied two *MLL-AF4* infant mixed phenotype acute leukaemias (MPALs), comprising distinct lymphoid and myeloid populations (MPAL1, MPAL2; Table S1).

All matched samples displayed identical oncogene fusion breakpoints at diagnosis and relapse as shown by DNA (n=14) and/or RNA (n=10) sequencing, confirming a common clonal origin and proving that the relapses are not *de novo* or therapy-associated AMLs (Figures 1B, C, S1, Tables S1). Breakpoints of LSALs and MPALs show a similar distribution as *MLL-AF4* ALL cases, clustering in *MLL* introns 9-11 and AF4 introns 3 and 4 (Meyer *et al.*, 2009; Jung *et al.*, 2010) (Figure 1D, Table S1) thus excluding that distinct, “non-canonical” chromosomal breakpoints are causative for *MLL-AF4*-positive AML. These data raised the question of the cellular origin of relapse and the nature of events secondary to *MLL-AF4* that affect lineage commitment.

Cellular origin of lineage switched relapse

We hypothesised that myeloid leukaemias may not have undergone substantial B-cell receptor (BCR) rearrangements. We used this feature to further interrogate the developmental

stage at which the relapse arose. To that end we examined BCR rearrangements within RNA-seq and whole exome-seq (WES) derived data, using MiXCR software (Bolotin *et al.*, 2017). All ALL cases showed classical oligoclonal rearrangements of BCR loci, supporting the lymphoid lineage decision (Figure S2, Table S2). However, we observed three distinct patterns for AML relapses (Figure 2A). Pattern 1 comprises AML relapse cells with no BCR rearrangements implying presence of a relapse-initiating cell residing in a primitive precursor population prior to early DJ recombination (Figure 2A, cases LS01, LS02, LS04). As a second pattern, we found unrelated BCR rearrangements, which may indicate either aberrant rearrangement in a myeloid cell or relapse initiating from either a B-lymphoid cell committed to undergo rearrangement, or a transdifferentiated minor ALL clone with an alternative rearrangement (Figure 2A, cases LS03, LS06, LS07, LS08, MPAL2). Interestingly, this pattern is also found in a relapse after blinatumomab treatment (LS10). Pattern 3 comprises shared BCR rearrangements between diagnostic and relapse material, which suggest a direct transdifferentiated myeloid relapse from the ALL (Figure 2A, cases LS05, LS09, MPAL1). These data demonstrate that AML relapses can originate from different stages of lymphoid leukaemogenesis and suggest, at least for a subset, a common precursor preceding the pro-B cell stage.

To functionally assess the plasticity of immunophenotypically defined diagnostic ALL and relapsed AML, we transplanted NSG or MISTRG mice with either ALL or AML cells from patient LS01 (Figure 2B). Because of the expression of several human myeloid growth factors, the MISTRG strain more strongly supports AML engraftment than the NSG strain, which shows a stronger lymphoid bias. Diagnostic ALL transplants rapidly produced representative CD19+CD33- lymphoid leukaemias in both mouse strains. In contrast, transplantation of relapsed AML engrafted only in the more myeloid-permissive MISTRG strain as a CD34+/-CD19-CD33+ AML (Figure 2C). Thus, in contrast to the immunophenotypic plasticity seen in MPAL leukaemias with wildtype *MLL* (Alexander *et al.*, 2018), transplantation of a relapsed, fully switched AML was only capable of generating AML.

Lineage switch relapse can originate in HSPC compartments

To further investigate a potential origin of relapse in an early progenitor or stem cell, we purified haematopoietic stem/progenitor cell (HSPC) populations from diagnostic ALL and relapsed AML and tested purified populations for the presence of *MLL-AF4* targeted sequencing. The *MLL-AF4* translocation was found in the lymphoid-primed multipotent progenitor-like population (LMPP, CD34+CD38-CD45RA+; lymphoid and myeloid potential, but not megakaryocyte-erythroid potential), in the multipotential progenitor population (MPP, CD34+CD38-CD45RA-CD90-; no lineage restriction) and for MPAL1 even in the haematopoietic stem cell-like population (HSC, CD34+CD38-Lin-CD90+) (Figures 3A-C, S3; Table S3). In line with these findings, serial xenotransplantation of LS01P identified a persistent human CD34+CD38-CD45RA-CD90+ HSC compartment across four generations of mice with maintenance of the *MLL-AF4* fusion gene within purified human CD34+ cells (Figures 3D-F). These findings suggest maintenance of a (pre-)leukaemic clone with high malignant self-renewal potential in HSPC populations and support our findings from BCR analysis that, at least in a subgroup of cases, an early multipotent progenitor or HSC can act as the cell of origin for the AML relapse.

In concordance with the translocation being present within the early HSPC compartment, we sorted viable differentiated leukocytes and were able to detect the *MLL-AF4* fusion in myeloid and lymphoid lineages including CD34-CD19/3-HLA-DR+CD14/11c+ monocytes, NK, B and mature T cells (Figure 3C, S3A, B; Table S3). These findings imply the existence of a pre-leukaemic progenitor cell, in which *MLL-AF4* does not impose a complete block on haematopoietic differentiation but is compatible with myeloid and lymphoid differentiation. These findings raise the question of which factors and molecular mechanisms affect the ALL and AML lineage choice in *MLL-AF4* leukaemia.

Lineage switch leukaemia is associated with transcriptional reprogramming

We next investigated the underlying molecular events associated with a lineage switch. To this end we analysed differential gene expression across eight cases, including six LSALs for which RNA was available at both presentation and relapse, and sorted lymphoid and myeloid blast populations of two MPALs. This analysis identified in total 1374 up- (adj. p<0.01, Log Fold change >2) and 1323 down-regulated genes in the AML lineage switches and the myeloid populations of MPAL patients (Table S4). The most substantially down-regulated genes include lymphoid genes such as *PAX5* and *RAG2*, but also the polycomb PRC1 like complex component *AUTS2* and SWI/SNF complex component *BCL7A,* while up-regulated genes comprise several myeloid genes such as cathepsins, cystatins, *PRAM1* and *CSF3R* (Figure 4A). These fingings were consistent irrespective of which cellular origin of relapse the BCR rearrangement analysis supported (pattern 1, 2 or 3),

Both gene set enrichment analysis (GSEA) and non-negative matrix factorisation (NMF) showed that presentation and relapse cases with lineage switch have expression signatures similar to previously published *MLL-AF4* ALL and *MLL*r AML cases as well as normal lymphoid and myeloid cell types, respectively (Figures 4B, S4A, B) (Zangrando *et al.*, 2009; Novershtern *et al.*, 2011; Andersson *et al.*, 2015). More specifically, lineage switch included increased expression of factors controlling myeloid differentiation (e.g., *CSF3R, KIT)* and changes in haematopoietic surface marker expression (e.g., *CD19, CD22*, *CD33, CD14, FCGR1A/CD64*), loss of immunoglobulin recombination machinery genes (e.g., *RAG1*, *RAG2, DNTT*) and reduced expression of genes encoding heavy and light immunoglobulin chains (Figures 4C-E, S4C). Notably, GSEA also indicated impaired DNA repair and cell cycle progression of the AML relapse when compared with diagnostic ALL (Figures S4D). In particular, the reduced self-renewal potential might reflect the transition from ALL with a high incidence of leukaemic stem cells (LSCs) to AML with fewer LSCs.

MLL fusion proteins including MLL-AF4 have previously been shown to directly regulate multiple genes linked to haematopoietic and leukaemic stemness (Ayton and Cleary, 2003;

Somervaille *et al.*, 2009; Gessner *et al.*, 2010; Wilkinson *et al.*, 2013; Kerry *et al.*, 2017). For instance, significant changes in gene expression across the *HOXA* cluster, notably with a 50-fold reduction in *HOXA7* expression, represent a major additional disruption of *MLL*r leukaemogenic transcriptional regulation (Figures S5A-C) (Ayton and Cleary, 2003; Somervaille *et al.*, 2009; Gessner *et al.*, 2010). Furthermore, a very significant portion of target genes including *PROM1*, *IKZF2* and chromatin modifying factors are highly enriched amongst genes that show lower expression in myeloid lineages (Figures S5D, E, Table S4). In total, 996 out of 5208 bona fide direct target genes of *MLL-AF4* changed expression in the AML relapse (Wilkinson *et al.*, 2013). These data suggest that the process of lineage switch is associated with a major reorganisation of the *MLL-AF4* transcriptional network and pose the question of which epigenetic regulators are involved in the lineage determination associated with this fusion gene.

Reorganisation of chromatin accessibility and transcription factor binding upon lineage switch

For case LS01 we had sufficient diagnostic material to perform DNase hypersensitivity site (DHS) analysis and thus link transcriptional changes to altered genome-wide chromatin accessibility. Many differentially expressed genes showed altered chromatin accessibility in proximity to the transcriptional start site (TSS), including genes encoding key hematopoietic surface markers CD33 and CD19, transcription factors and proteins related to differentiation (Figures 5A-C, S6A).

Changes in chromatin accessibility were linked with an altered pattern of transcription factor binding. High resolution DHS-seq (digital footprinting) of presentation ALL and relapse AML, respectively, showed marked genome-wide alterations at sites distal to the TSS, from which several AML and ALL specific *de novo* occupied transcription factor binding motifs were identified (Figures 5C-E, S6B). Lineage switch from ALL to AML was associated with a loss of occupancy of motifs binding lymphoid transcription factors such as EBF or PAX5 and a gain

of occupancy of motifs bound by C/EBP, IRF and NF-κB family members (Figure 5E-F). The gain in C/EBP motif binding was associated with the expression of a C/EBPA regulated myeloid transcriptional gene set (Figure 5G). We also observed a redistribution of occupancy of transcription factors controlling both lymphoid and myeloid maturation such as RUNX, AP-1 and ETS members to alternative cognate motifs (Figures 5E and S6B) (Hohaus *et al.*, 1995; Nerlov and Graf, 1998; Leddin *et al.*, 2011). This finding is exemplified by decreased accessibility of a region located 1 kb upstream of the *CD19* TSS with concomitant loss of EBF binding at this element (Figure 5C). Differential motif enrichments were associated with changes in RNA expression and chromatin accessibility at the genes encoding the corresponding cognate transcription factors *EBF1*, *PAX5*, *LEF1* (B lymphoid determinants), *NFKB2* and *CEBPB/D/E* (myeloid determinants), particularly in regions flanking TSSs (Figures 5H, S6B).

In conclusion, the transition from lymphoid to myeloid immunophenotype is associated with global lineage specific transcriptional reprogramming and genome-wide alteration in chromatin accessibility and transcription factor binding.

Lineage switch changes alternative splicing patterns

Lineage fidelity and determination are not only linked to differential gene expression but may also include co- or post-transcriptional mechanisms. It has been previously demonstrated that lineage commitment during haematopoiesis leads to substantial changes in alternative mRNA splicing patterns (Chen *et al.*, 2014). Furthermore, we recently showed that the *AML-ETO* (RUNX1-RUNX1T1) fusion protein controls leukaemic self-renewal by both differential gene transcription and alternative splicing (Grinev *et al.*, 2021). To complement the transcriptional analysis we therefore sought to define the different composition of RNA isoforms in lymphoid and myeloid populations from lineage switch and MPAL cases. Here we focussed on three lineage switch patients and the two MPAL patients whose RNA-seq data provided sufficient read depth for the analysis of exon-exon junctions, exon usage and intron retention (Figures

S7A-C). We detected in total 2630 retained introns (RIs) shared amongst the three lineage switches with 653 and 343 RIs exclusively found in the diagnostic ALL or the AML relapses, respectively. This was complemented by 97 exons (DEUs) and 193 exon-exon junctions (DEEjs) differentially used between diagnostic ALL or the AML relapses. In contrast, this analysis identified only 43 RIs present in both MPAL cases with 18 and 6 introns specifically retained in either the lymphoid or myeloid subpopulation, respectively (Figure S7D, Table S5). Intersection of the affected genes identified 21 shared genes out of 166 DEEjs and 74 DEUs. MPALs had 420 DEUs and 155 DEEjs affecting 257 and 103 genes with 30 genes having both DEUs and DEEjs (Figure 6A, Table S6). While more than 80% of the non-differential exon-exon linkages were canonical, this was true for only 15% of the DEEjs. Here, non-canonical exon skipping and complex splicing events contributed more than 30% each, most prominently to differential alternative splicing (Figure 6B).

Pathway analysis revealed an enrichment of alternatively spliced genes in immune pathways including antigen processing, membrane trafficking and FCGR-dependent phagocytosis reflecting the change from a lymphoid to a myeloid state (Figures 6C, S7E). Furthermore, it highlighted RNA processing and maturation including mRNA splicing, processing of capped intron-containing pre-mRNAs and rRNA processing. Indeed, myeloid populations expressed 4-6-fold higher alternatively spliced *SF3B1* and *SRSF5* levels than their matched lymphoid populations (Figure 6D, E). In addition, we noted a significant number of genes encoding epigenetic modulators including *KDM5C, HDAC2* and several *CHD* members being differentially spliced in AML relapse or myeloid subpopulations of MPALs (Figure 6D, E).

Analysing the occurrence of *MLL-AF4* isoforms showed that the fusion site of *MLL-AF4* itself was subject to alternative splicing in lineage switch and MPAL. Three cases shared the breakpoints in introns 10 and 4 of *MLL* and *AF4*, respectively (Figure 6F). By examining RNA-seq and competitive RT-PCR data, we identified two co-occurring isoforms with either *MLL* exon9 or exon10 joined to *AF4* exon4 (Figure 6F). Interestingly, lymphoid populations expressed a higher ratio of exon10/exon4 over exon9/exon4 than the myeloid populations in

all three cases examined. Furthermore, the HSC and MPP-like populations of MPAL1 showed mainly expression of the exon10/exon4 splice variant of *MLL-AF4*, thus resembling more the lymphoid than the myeloid phenotype (Figure 6F). In conclusion, altered isoform expression of *MLL-AF4* may contribute to lineage choice and the phenotypic switch.

The mutational landscape of lineage switch

Next, we examined the mutational landscape of lineage switched *MLL-AF4* leukaemias by performing exome sequencing on the entire cohort. In keeping with published data on newly presenting *MLL*r acute leukaemias (Dobbins *et al.*, 2013; Andersson *et al.*, 2015), exome sequencing of presentation ALL samples confirmed a relatively quiet mutational landscape in infant ALL cases, with median of 13 nonsynonymous somatic single nucleotide variants (SNVs) or insertions/deletions (indels) predicted to be deleterious to protein function (Table S7). Many of these were present in less than 30% of reads and considered sub-clonal. The most commonly mutated genes at presentation were *NRAS* (3 cases) and *KRAS* (5 cases) (Figure 7A) as described previously (Andersson *et al.*, 2015). In contrast, relapse AML samples contained a median of 46 deleterious somatic SNVs and indels (Table S7). This increase can be mainly attributed to three samples (LS03RAML, LS07RAML and LS08RAML) that carried deleterious mutations in DNA polymerase genes in the respective major clones linked to hypermutator phenotypes (Church *et al.*, 2013; Erson-Omay *et al.*, 2015; Mur *et al.*, 2020) (Figure 7A). However, we observed this phenotype only in three out of ten relapses, arguing against this phenomenon being a general requirement for the lineage switch in relapse. Similarly, many of the predominantly subclonal mutations identified in presentation ALL samples, including half of *RAS* mutations, were subsequently lost at relapse, indicating alternative subclones as the origin of relapse (Figure 7A). Finally, both MPALs harboured several mutations that were exclusively found in either the lymphoid or myeloid subpopulation indicating the presence of subclones with a lymphoid and myeloid bias (Figure 7A).

Next, we examined the global mutation patterns of patient LS01 in greater detail and focussed on SNVs with a VAF of ≥0.3. ALL and AML contained 104 and 3196 SNVs above this threshold, respectively, with only 22 shared SNVs between both samples. The most prevalent type of SNV was the C to T transition in the DNA of both ALL and AML samples (Figure 7B). However, the contribution of underlying single base substitution (SBS) signatures differed between diagnosis and relapse. Three different signatures (SBS16, 5 and 1) explained about 60% of the SNPs found in ALL, while SBS1 seemed to explain more than 50% of all SNPs in the AML (Figure 7B) suggesting a mutational clock as the main driver of the evolution of relapse (Alexandrov et al., 2015; Blokzijl et al., 2016). Further inspection of the pattern revealed a mutational signature mainly comprising C to T transitions and to a lesser degree C to G transversions in NCG triplets, raising the possibility that thiopurine maintenance treatment may have increased the mutational burden, resulting in lineage switched relapse in this patient (Li *et al.*, 2020).

Twelve deleterious SNVs were identified as unique to the relapse sample of case LS01. The availability of viable cellular material allowed us to investigate the order of acquisition of these secondary mutations within the structure of the normal haematopoietic hierarchy. We sorted this sample to isolate HSC-, MPP-, LMPP- and GMP-like and later populations. Using a targeted deep sequencing approach we then examined these populations for the presence of those 12 SNVs. This analysis showed an increasing number of mutations during the differentiation from MPPs through LMPPs to GMPs. Amongst them, only *PHF3* and *CHD4* mutations were present within the purified CD34+CD38-CD45RA-CD90- MPP-like fraction with VAF≥0.3 (Figure 7C and Table S3). In contrast, LMPP- and GMP-like populations contained all 12 SNVs at high VAF (Table S3). These findings identify the mutation of *CHD4* and *PHF3* as the earliest genetic events during relapse evolution and suggest them as potential drivers of an *MLL-AF4* positive, non-lineage committed, pre-leukaemic precursor population. Subsequent accumulation of additional mutations likely establishes the fully developed leukaemia with more mature haematopoietic/myeloid immunophenotypes.

Perturbation of *CHD4* and *PHF3* disrupts lymphoid development in *MLL-AF4* expressing cells

The two earliest mutations within LS01 relapse were identified in the Nucleosome Remodelling and Deacetylation complex (NuRD) member *CHD4* and the plant homeodomain finger containing *PHF3*. PHF3 is a member of a family of transcriptional regulators that have been suggested to link the deposition of histone marks to the regulation of transcription (Kinkelin *et al.*, 2013). PHF3 itself has been recently identified as an inhibitor of transcription elongation by competing with TFIIS for binding to the C-terminal domain of RNA polymerase II (Appel *et al.*, 2020). NuRD is a multiprotein transcriptional co-repressor complex with both histone deacetylase and ATP-dependent chromatin remodelling activity. It is a critical factor in the lymphoid lineage determination in part directed by the transcription factor IKZF1 (Kim *et al.*, 1999; Yoshida *et al.*, 2006; Ng *et al.*, 2009; Zhang *et al.*, 2011; Arends *et al.*, 2019). Both the CHD4 R1068H and PHF3 K1119I mutations affect highly conserved residues (Figures 8A, S8A) that are predicted by the Condel classifier (González-Pérez and López-Bigas, 2011) to disrupt protein function. Specifically, the CHD4 R1068H mutation has previously been linked to defects in cardiac development (Sifrim *et al.*, 2016).

Whilst mRNA expression levels of the mutant CHD4 R1068H were unaffected in case LS01, across the remaining cases analysed, expression of *CHD4* and additional NuRD complex members *MBD3*, *MTA1* and *RBBP4* was reduced following lineage switched myeloid relapse (Figures 8B). Furthermore, *CHD4* was affected in the MPAL myeloid populations by non-canonical alternative splicing leading to premature termination of translation, indicating that this particular pathway was severely disrupted across the cohort, irrespective of the putative cell of origin of relapse.

The relevance of *CHD4* and *PHF3* in regulating lymphoid versus myeloid lineage choice was further supported by ARACNE analysis using published ALL (Harvey *et al.*, 2010; Kang *et al.*, 2010) and AML (Gentles *et al.*, 2010) expression datasets (n=216) which we used to reverse-engineer a mutual information network (Margolin *et al.*, 2006). This network was trimmed to

represent only genes significantly associated with the difference between AML and ALL thus reflecting the likely influence of genes of interest upon genes associated with the difference between AML and ALL. This analysis found *CHD4* and *PHF3* to be the mutated genes with the highest number of edges within the network (*PHF3* – 21 edges, p=0.010; *CHD4* – 12 edges, p=0.0005, Figure S8B and Table S8), implicating them as causal to the lymphoid/myeloid distinction identified within primary ALL/AML.

To establish a direct functional link from these mutations to the loss of lymphoid lineage commitment in *MLL-AF4* ALL, we knocked down *CHD4* and *PHF3* in the *MLL-AF4* positive cell line, SEM. Depletion of *CHD4* and *PHF3* alone or in combination resulted in increased expression of the myeloid transcription factor *CEBPA* and reduced expression of lymphoid transcription factors including *LEF1*, *PAX5, TCF3* and *TCF12* (Figure 8C). These changes were accompanied by a more than twofold increase in CD33 expression in the two *MLL-AF4*-expressing ALL cell lines SEM and RS4;11, while CD33 levels in two *MLL-AF4*-negative ALL cell lines remained unaffected, supporting the importance of these epigenetic regulators specifically within the *MLL-AF4* context (Figures 8D and S8C). GSEA following *PHF3* and *CHD4* knockdown indicated loss of HSC and B-lymphocyte progenitor gene expression signatures (Figure 8E, left panel), similar to what was observed with the transcriptomes of the lineage switch leukaemia cases (Figure 8E, right panel). In line with these cell line experiments, knockdown of *CHD4* and *PHF3* in an ALL PDX generated from the first relapse of patient LS03 also resulted in a more than twofold increase in CD19/CD33 double and CD33 single positive cells, demonstrating that perturbation of these two genes is also able to change the immunophenotype of primary *MLL-AF4* ALL cells in a case susceptible to trans-differentiated relapse (Figure S8D).

In order to examine the role of additional mutations of chromatin modifiers found in our cohort and known to regulate lineage choice, we investigated the impact of the PRC1 members *PCGF6* and *AUTS2* on CD33 expression in SEM cells. *PCGF6* is mutated in LS07RAML and LS08RAML and has known roles in B lymphoid malignancy (Ferreira *et al.*, 2008; Wang *et al.*,

2015). *AUTS2* is both mutated in LS08RAML and highly expressed in all lymphoid populations examined (Figures 4A, 7A). STRING network (Szklarczyk *et al.*, 2019) analysis demonstrated close functional associations within PRC1 and NuRD complexes and their shared associations (Figure S8E). Mutual interactions include CBX2, a PRC1 complex member, which shows a similar expression pattern between lineage switch and MPAL cases (Figure S8F). While knockdown of *AUTS2* did not change CD33 levels, depletion of *PCGF6* also increased CD33 surface expression in SEM cells, further supporting the notion of epigenetic factors regulating lineage determination in ALL (Figures S8G). Furthermore, GSEA also indicated impaired function of PRC1 and PRC2 complexes in the AML relapse compared with the presentation ALL with down-regulation of their respective target genes (Figure S8H).

Given that the relapse-initiating cell can arise within an uncommitted, *MLL-AF4* translocated HSPC population, we went on to assess the impact of *CHD4* and *PHF3* function loss in a human cord blood model, which harbours a chimeric *MLL-Af4* fusion and can be differentiated both into myeloid and lymphoid lineages (Lin *et al.*, 2016). Knockdown of either *CHD4* or *PHF3* under lymphoid culture conditions significantly impaired lymphoid differentiation potential, whilst co-knockdown of *CHD4* and *PHF3* disrupted differentiation entirely (Figures 8F, G and Table S9). Transcriptomic analysis of the sorted populations revealed that CD33 positive cells exhibited metagene expression pattern similar to *MLL*r AML, while the pattern describing CD19+ cells was most similar to *MLLr* ALL, thus confirming that changes in surface marker expression were associated with the corresponding changes in the transcriptomic profiles (Figure S8I).

Taken together, our data show the important role of the NuRD (CHD4), PHF3 and other (PRC1*)* repressive complexes in the epigenetic control of lymphoid lineage choice. In particular, dysregulation of *CHD4*/NuRD was mediated by mutation, down-regulation of expression and differential splicing across the cohort, irrespective of the cellular/clonal origin of relapse. These data support a role for these factors in the strong lineage determining

capacity of *MLL-AF4* whilst their loss undermines both the execution and the maintenance of the lymphoid lineage fate.

Discussion

This study describes impaired epigenetic control as being central to the phenomenon of lymphoid-myeloid lineage switch in *MLL-AF4*-positive leukaemia and identifies the cell of origin of relapse into AML. We found that the origin of relapse was heterogeneous. Relapse can directly evolve from pro-B-like ALL blast populations, which agrees with the general self-renewal capacity of ALL cells (le Viseur *et al.*, 2008; Rehe *et al.*, 2013), but can also originate within the HSPC compartment. Indeed, analysis of both patient and xenotransplanted cell populations from diagnostic ALL identified *MLL-AF4* fusion transcripts in MPP- and HSC-like cells. This finding agrees with recently published data pointing at MPP cells as the origin of *MLL-AF4* leukaemia (Malouf and Ottersbach, 2018) and is in line with transcriptomic similarities between t(4;11) ALL and Lin-CD34+CD38-CD19- fetal liver cells, again suggesting an HSPC as the cell of origin (Agraz-Doblas *et al.*, 2019).

Irrespective of the cellular origin of the relapse, lineage switching was associated with a major rewiring of gene regulatory networks. At the level of transcriptional control, the decision for lymphoid development relies not only on the activation of a lymphoid transcriptional program, but also on the silencing of a default myeloid program (Nutt *et al.*, 1997, 1999). That decision is enacted by lymphoid master regulators including EBF1, PAX5 and IKAROS, which represent genes commonly mutated in precursor B-ALL (Mullighan *et al.*, 2007; Pongubala *et al.*, 2008; Boer *et al.*, 2016; Witkowski *et al.*, 2017). $Pax5^{-/-}$ pro-B cells which lack lymphoid potential, whilst capable of differentiating down erythro-myeloid lineages *in vitro*, still maintain expression of early B cell transcription factors *EBF1* and *E2A* (*TCF3*) (Nutt *et al.*, 1999). In contrast, we show that lineage switching *MLL-AF4* pro-B leukaemic relapse is associated with significant reduction in expression and binding of these earliest B lymphoid transcription factors. Their loss is linked with changes in the *MLL-AF4* transcriptional programme, notably

within *HOXA* cluster genes (Ayton and Cleary, 2003; Somervaille *et al.*, 2009; Gessner *et al.*, 2010)., which likely results in a wider reorganisation of malignant haematopoietic transcriptional networks, ultimately leading to a myeloid differentiation fate.

Similar to the *Pax5* knockout (Nutt *et al.*, 1999), loss of IKAROS DNA-binding activity prevents lymphoid differentiation (Zhang *et al.*, 2011). NuRD co-operates directly with IKAROS to repress HSC self-renewal and subsequent myeloid differentiation, permitting early lymphoid development (O'Neill *et al.*, 2000; Williams *et al.*, 2004; Yoshida *et al.*, 2008; Zhang *et al.*, 2011; Hosokawa *et al.*, 2013; Lu *et al.*, 2019). We found that the abrogation of this pathway through multiple mechanisms was central to the lineage switch from ALL to AML. Lineage switch was either associated with mutation, reduced expression or, in the case of two MPALs, alternative splicing of *CHD4* and other NuRD components. Long term knockdown of *CHD4* was not tolerated in our cord blood culture. This is in line with reports showing that complete loss of *CHD4* impairs leukaemic proliferation (Sperlazza *et al.*, 2015; Heshmati *et al.*, 2018), both myeloid and lymphoid differentiation of HSPCs and causes exhaustion of HSC pools (Yoshida *et al.*, 2008), indicating that basal CHD4 expression is required for maintaining AML. Moreover, our observation that a 60% *CHD4* knockdown is associated with the activation of pluripotency gene signatures is in line with the finding that a partial inhibition of *CHD4* supported induction of pluripotency in iPSCs, while a complete deletion eliminated cell proliferation (Mor *et al.*, 2018).

Whilst our study has investigated the rare clinical occurrence of lineage switching, recent studies have identified core NuRD and PRC1 complex members as being direct targets of *MLL-AF4* binding (Kerry *et al*., 2017; Harman *et al*., 2021). We therefore hypothesise that epigenetic regulator genes are co-opted during *MLL-AF4* leukemogenesis and mediate fundamental lineage specific decision-making processes, in this case the suppression of the myeloid lineage program. Multiple routes to their dysregulation may result in escape from this lineage restriction. Our finding that frontline chemotherapy itself may contribute to relapse highlights the urgent need to find alternative therapies for this high-risk leukaemia. Equally,

however, the associated loss of B cell surface markers (*e.g.*, CD19) provides an alternative mechanism for relapse following CAR-T cell or blinatumomab therapy (Gardner *et al*., 2016; Rayes, McMasters and O'Brien, 2016) in addition to mutations, alternative splicing (Sotillo *et al.*, 2015; Orlando *et al.*, 2018; Rabilloud *et al.*, 2021) and T cells trogocytosis (Hamieh *et al.*, 2019). Whilst these therapies target lineage specific surface markers, lineage-switched (pre-)leukaemic progenitor populations escape epitope recognition and provide a potential clonal source for the relapse. Given the increasing use of advanced immunological therapies, a detailed understanding of the molecular processes underlying lineage determination and switching will be critical for developing new strategies to avoid this route to clinical relapse.

Acknowledgements

We thank Jon Coxhead and Raf Hussain at the Newcastle University core genomics facility as well as Marc van Tuil at the Princess Maxima Center Diagnostic department for development of sequencing strategies. We acknowledge the Newcastle University Flow Cytometry Core Facility and Tomasz Poplonski at the Princess Maxima Center Flow Cytometry core facility for their assistance with the generation of flow cytometry data and cell sorting strategies as well as the Newcastle University Bioinformatics Support Unit for helping to develop the analysis approach for sequencing data. We thank Ruben van Boxtel and Eline Bertrums at the Princess Maxima Center for the generation of single cell derived clones, derived from early progenitors sorted from MPAL patient samples. We thank Monique den Boer, Frank van Leeuwen and Ronald Stam for critically reading the manuscript. This study was supported by a Cancer Research UK Centre Studentship (C27826/A17312) and Newcastle University Overseas Research Scholarship to RT, a CRUK program grant to JV and OH (C27943/A12788), grants from the North of England Children's Cancer Research Fund to OH, JV and SB, by Bloodwise grants 12055 and 15005 to OH and by a grant from the Kay Kendall Leukaemia Fund (KKL1142) to OH. SB was supported by an NIHR Academic Clinical Lectureship (CL-2012-01-002) and the Sir Bobby Robson Foundation Clinical Fellowship. Work in CB/PNC's lab was funded by a programme grant from Bloodwise (15001).

485 Work in JMA's lab was funded by a programme grant from Bloodwise (13044). EZ was
486 supported by an RFBR grant (№17-29-06052). Research in the VVG laboratory was
487 supported in part by the Ministry of Education of the Republic of Belarus, grant #3.04.3.
488 Research in the AK laboratory was supported by an RSF grant (20-75-10091).

489 This study makes use of data generated by the St. Jude Children's Research Hospital –
490 Washington University Pediatric Cancer Genome Project and the Therapeutically Applicable
491 Research to Generate Effective Treatments (TARGET) initiative, phs000218, managed by the
492 NCI (see supplementary methods).

493 Author contributions

494 Conceptualization, O.H., S.B., C.B.; Methodology, O.H., C.B., R.T., K.S., P.M., S.B., A.P.,
495 C.M., A.K., Z.K., J.B., V.B., R.M., J.V., J.M.A., S.L.; Software Programming, S.N., J.H.K.,
496 V.V.G., A.K., D.W., P.C.; Formal Analysis, S.N., J.H.K., V.V.G., A.K., D.W., P.C., C.B., O.H.;
497 Investigation, R.T., K.S., P.M., A.P., C.M., H.J.B., A.K., S.A., M.R.I., P.E., H.M., A.E., N.M.S.,
498 S.E.F., Y.S., D.P., P.C.; Resources, F.V., E.Z., A.S., J.C.M., L.J.R., C.E., O.A.H., S.Ba, R.S.,
499 N.M., M.C., V.B., R.M., M.W., C.J.H., C.A.C., D.S., Y.O., M.J.T., P.N.C., J.C.M., C.B., O.H.;
500 Data Curation, S.N., D.W., P.C.; Writing, S.B., O.H., C.B., R.T., K.S.; Supervision, O.H., S.B.,
501 J.M.A., J.V., C.B., ; Funding Acquisition, O.H., J.V., S.B., C.B., P.N.C., J.M.A., E.Z.

502 Declaration of interests

503 Z.K. and J.B. are employees of Illumina, a public company that develops and markets systems
504 for genetic analysis. The remaining authors declare no competing interests.

505 Figure titles and legends

506 **Figure 1. Characterisation of the *MLL-AF4* lineage switch cases.** (A) Morphological
507 change from lymphoblastic leukaemia (left panel) to acute monoblastic/monocytic leukaemia
508 (right panel). The scale bar represents 20 μm. (B) Sanger sequencing of *MLL-AF4* and

reciprocal *AF4*/*MLL* fusions in LS01 presentation ALL (upper panel) and relapse AML (lower panel) identifies a common breakpoint with identical filler sequence in ALL and AML samples. (C) Whole genome sequencing data of LS01 showing karyotype (outer circle), copy number changes (log2 depth ratio in 1Mb windows, loss <2 green dots, gain >2 red dots) and structural variants (translocations - red connecting lines, deletions – blue lines, inversions, purple lines). (D) Schematic representation of identified fusion variants, located within the major breakpoint MLL region, present in analysed t(4;11) cases (detailed breakpoint description presented at Fig.S1 and Table S1). Red connecting line indicates MLL-AF4 translocation positions of each gene partner.

Figure 2. Cellular origin of the *MLL-AF4* leukaemia. (A) Evaluation of B-cell receptor repertoires on ALL (presentation) and AML (relapse) lineage switch, and MPAL cases identified three different patterns. Pattern 1 - with clonotypes on the ALL only (LS01, LS02, LS04). Pattern 2 - B-cell receptor-containing clones on ALL and AML, but distinct to each other (LS03, LS06, LS07, LS08, LS10, MPAL2). Pattern 3 - B-cell receptor-containing clones shared between ALL and AML (LS05, LS09, MPAL1). Graphs represent frequency of clones identified with WES and/or RNAseq in LSAL and MPAL samples, respectively. (B) Presentation ALL sample only produces ALL in all mouse strains transplanted. In contrast, Relapse AML sample does not engraft in lymphoid supportive NSG mice and produces only AML in myeloid supportive MISTRG mice. (C) Post-transplantation flow cytometric analysis of LS01PALL produced in NSG mice (top panel) and LS01RAML produced in MISTRG mice (bottom panel).

Figure 3. Haematopoietic stem/progenitor populations carry *MLL-AF4*. (A) Flow cytometric sorting strategy for haematopoietic stem/progenitor cell (HSPC) populations. (B) PCR identification of the specific *MLL-AF4* fusion within sorted HSPC populations LS01 and MPAL1 cases. (C) Summary of *MLL-AF4* positivity within different HSPC populations analysed in patients LS01PALL and early progenitors and lymphoid fraction of MPAL1, presented as red and pink circles, respectively. (D) Flow cytometric analysis showing sequential transplantation of LS01 ALL presentation sample across four generations of NSG mouse

xenografts. The HSC population (CD34+CD38-CD45RA-CD90+) is maintained across four mice generations. (E) Proportion of bone marrow human CD34+ cells in CD34+CD38-CD45RA-CD90+ HSC gate in all analysed xenografts. (F) PCR identification of the specific *MLL-AF4* within the sorted HSPC xenograft sample.

Figure 4. Transcriptional reprogramming in lineage switch and MPAL cases. (A) Heatmap showing the top 100 differentially expressed genes between ALL and AML from six lineage switch (LS01, LS03, LS04, LS05, LS06, LS10) and two MPAL cases, ranked by stat value. (B) Enrichment of myeloid growth and differentiation signature in relapsed samples (left panel) identified by GSEA analyses, also pointing to downregulation of genes highly correlated with acute lymphoblastic leukemia (middle and right panel). Gene set enrichment analyses have been performed based on data derived from six lineage switch samples. FDR – false discovery rate, NES – normalised enrichment score. (C-E) Differential expression of (C) lineage specific, (D) immunoglobulin recombination machinery, and (E) genes encoding immunoglobulin heavy and light chains in lineage switch and MPAL cases. Error bars show standard error of the mean (SEM) for lineage switch cases and ranges for two MPAL cases.

Figure 5. Chromatin re-organisation and differential transcription factor binding underpins lineage switching. (A) DNase hypersensitivity site sequencing identifies 13619 sites with a log2 fold reduction and 12203 sites with a log2 fold increase following lineage switch to AML. (B) University of California, Santa Cruz (UCSC) genome browser screenshot displaying differential expression at lineage specific loci (lower red tracks) accompanied by altered DNase hypersensitivity (upper black tracks) proximal to the transcriptional start site (TSS) of *CD33*. (C) UCSC genome browser screenshot for *CD19* zoomed in on an ALL-associated DHS with EBF occupation as indicated by high resolution DHS-seq and Wellington analysis. FP - footprint. (D) Heat maps showing distal DHS regions specific for AML relapse on a genomic scale. Red and green indicate excess of positive and negative strand cuts, respectively, per nucleotide position. Sites are sorted from top to bottom in order of decreasing Footprint Occupancy Score. (E) De novo motif discovery in distal DHSs unique to AML relapse

as compared to ALL relapse as shown in (D). (F) EBF1 and C/EBP binding motifs demonstrate differential motif density in presentation ALL and relapse AML. (G) Enrichment of a myeloid C/EBPA network gene set in signatures associated with relapse AML and diagnostic ALL samples as identified by GSEA (H) Differential expression of TFs cognately binding to differentially accessible motifs shown in (F). TFs whose binding motifs show increased accessibility in ALL are in blue whilst those showing increased accessibility in AML are in red. RUNX1 in black reflects enriched accessibility of different RUNX1 binding sites in ALL and AML. Error bars show SEM or ranges in LSAL and MPAL cases, respectively.

Figure 6. Alternative splicing in lineage switch and MPAL cases. (A) Volcano plots demonstrating differential usage of exon-exon junctions in the transcriptome of AML/myeloid versus ALL/lymphoid cells of lineage switch (LS01, LS03 & LS04) or MPAL patients. The vertical dashed lines represent two-fold differences between the AML and ALL cells and the horizontal dashed line shows the FDR-adjusted q-value threshold of 0.05 (upper panel). Venn diagrams (lower panel) showing distribution of splice variants identified as significantly changed in AML (or myeloid fraction of MPAL patients), including exon-exon junctions (DEEj), differential exome usage (DEU) and retained introns (RI). (B) Pie charts showing the classification of non-differential (non-DEEj) and differential (DEEj) exon-exon junctions. Shown are the percentages of splicing events assigned to a particular mode of splicing. Complex splicing event corresponds to several (two or more) alternative splicing incidents occurring simultaneously in the same sample. (C) Enrichment analysis of affected signalling pathways by the exon-exon junctions (DEEj) and differential exome usage (DEU) in the LSAL AML relapse and myeloid compartment of MPAL patients. Pathway enrichment analysis has been performed with https://biit.cs.ut.ee/gprofiler/gost under the highest significance threshold, with multiple testing correction (g:SCS algorithm). (D) Fold change expression levels of total gene among genes identified to be affected by alternative splicing process (left panel) and differentially spliced variants in lineage switched and myeloid compartments of MPAL patients (right panel). (E) Schematic representation of the affected mRNA structure (and its

probable consequence depicted in red) within several selected genes (upper panel). Corresponding normalized expression level (vs reference gene *TBP*) in two tested lineage switch patients (LS03 and LS04) and one MPAL (lower panel). Shown is the ratio of analysed splice variant expression level between AML (or myeloid) and ALL (or lymphoid) populations. (F) Splice variants of *MLL-AF4* identified in MPAL patients and one lineage switch sample (LS10). Left panel represents junction read counts of the fusion oncogene, identified by the RNAseq analysis, with confirmation of the expression of both variants analysed by qRT-PCR (MPAL1, right panel). Both splice variants, further confirmed by Sanger sequencing, showed complete sequences of *MLL*ex9 or *MLL*e10 and either complete or truncated (for 3 nucleotides at the 5'end) *AF*4ex4, respectively.

Figure 7. Molecular characterisation of lineage switch *MLL-AF4* leukaemias. (A) Whole exome sequencing (WES) data showing genes recurrently mutated within the analysed cohort, involved in the cell growth, communication and metabolism and genes mutated in single cases belonging to the same function protein complexes (e.g. DNA polymerases, epigenetic complexes). Data are presented according to the disease timepoint/cell lineage and age of the patient. Depicted are major single nucleotide variants (SNVs) that were found in >33% of reads and minor SNVs in <33% reads. (B) Clonal SNV patterns identified by whole genome sequencing (WGS) in LS01 ALL and AML samples, constructed from counts of each mutation-type at each mutation context, corrected for the frequency of each trinucleotide in the reference genome. (C) Comparison of the whole exome sequencing and RNAseq data obtained for LS01 patient identifies 6, 98, and 10 SNVs expressed only in ALL, AML, and shared between ALL and AML, respectively. 12 SNVs exclusive for the AML relapse, predicted (by Condel scoring) to have deleterious effects, were subjected to multiplex PCR followed by next-generation sequencing analysed within each of the purified haematopoietic sub-populations. Circles with solid colour indicate VAF >30%, light colour and dashed line indicates VAF 5-30%. Remaining genes (yellow circle) represent the 10 other SNVs (out of 12 SNVs) which showed the same pattern in the frequency of mutation acquisition.

Figure 8. Epigenetic modulatory genes influence lineage specific expression profiles. (A) CHD4 scheme; the R1068H mutation is located in the critical helicase domain of CHD4 at a highly conserved residue. An * (asterisk) indicates positions which have a single, fully conserved residue, a : (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, a . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix. (B) Fold change in expression of NuRD complex members (*CHD4*, *MTA1*, *RBBP4*, *MBD3*) and *PHF3* following lineage switched relapse (left panel) and in MPAL cases (right panel). (C) Expression of lineage specific genes following knockdown of *PHF3*, *CHD4* or the combination, relative to non-targeting control construct in the *MLL-AF4* positive ALL cell line, SEM. (D) Flow cytometric analysis of the surface CD33 expression following knockdown of *PHF3*, *CHD4* or the combination in the SEM cell line. shNTC – non-targeting control. (E) Gene set enrichment analysis of RNA sequencing data derived from: knockdown of *PHF3* and *CHD4* in the SEM cell line (left panel) and lineage switch leukaemia cases (right panel). Shown is negative correlation of shCHD4/shPHF3 and relapse samples with Jaatinen haematopoietic stem cell signature (upper panel) and Haddad B-lymphocyte progenitor signature (lower panel). (F) Expression of lineage specific cell surface markers CD19 (lymphoid) and CD33 (myeloid) following culture of *MLL-Af4* transformed hCD34+ cord blood progenitor cells in lymphoid permissive conditions. Knockdown of *PHF3*, *CHD4* or the combination disrupts the dominant lymphoid differentiation pattern seen in non-targeting control (shNTC). (G) Assessment of *PHF3* knockdown influence on the surface marker expression after longer incubation period (33 days); *CHD4* knockdown impaired cellular survival upon longer *in vitro* culture.

Figure S1. Sequencing of *MLL-AF4* fusion breakpoints from DNA (LS02, LS03, LS04, LS07, LS08, LS09) or RNA (LS05, LS06, LS10, MPAL1, MPAL2).

Figure S2. Immunoglobulin rearrangement on ALL and AML LS01 according to the standard protocol BIOMED-2. Clonal peaks are coloured in dark blue, as indicated in the multiplex amplification FR1, FR2, and FR3 of the VH segments LS01PALL (left panel). The

multiplex results were confirmed by single amplification VH-JH primer sets, which further showed two clonal VH3, one clonal VH1, and clonal VH4 rearrangement. No clonal peaks are seen in LS01RAML (right panel).

Figure S3. Evaluation of *MLL-AF4* presence in HSPC populations. (A) Flow cytometry plots of single sorted HLA-DR+CD14+CD11c+ monocytes LS01PALL. The two *MLL-AF4* positive cells, cells 11 and 18, are highlighted in red throughout the sorting strategy (lower panel). (B) Amplification of the *MLL-AF4* in cells 11, 18 and bulk LS01 sample shows the expected 299 base pair band. (C) Flow cytometry plots showing sorting strategy for lymphoid and myeloid populations of MPAL1 and MPAL2. (D) Bulk sort plots for HSC and MPP populations (upper panel) and single cell plots for Lin-CD34+ populations (lower panel) sorted from MPAL1 sample. (E) Amplification of the *MLL-AF4* in single- cell derived clones Lin-CD34+, lymphoid, and myeloid fractions of MPAL1 patient.

Figure S4. Transcriptional profiles of ALL and AML lineage switch leukaemia. (A) Non-negative matrix factorisation analysis of paired presentation ALL cases and relapse AML cases, against normal haematopoietic precursors (Novershtern *et al*, 2011). Presentation ALL cases demonstrate a high expression of B cell metagene signature, whilst relapse AML cases continue to express this signature to a variable degree (upper panel). Relapse AML cases demonstrate high expression of myeloid metagene signature, at a comparable level to that of GMP (lower panel). (B) Non-negative matrix factorisation analysis of paired presentation ALL cases and relapse AML cases, against ALL and AML cases, with and without *MLL* rearrangement (Andersson *et al* 2015, Bolouri *et al* 2018). Heatmap shows clustering of presentation cases (dark blue) with B-ALL *MLL-AF4* and relapse cases (dark green) with AML with *MLL*r. (C) Differential promoter accessibility is associated with higher expression of *PAX5*, *LEF1*, and *CD79A* in presentation ALL cases, and of *CSF3R*, *KIT* and *CSF2RA* in relapse AML cases. TSS – transcriptional start site. (D) Gene set enrichment analysis indicates impaired DNA repair and cell cycle progression in AML relapse. NES – normalised enrichment score.

Figure S5. Impact of leukaemia lineage switch on *MLL-AF4*-regulated genes. (A) HOXA cluster analysis includes a general reduction on HOXA3-10 in relapse AML cases. (B) Differential chromatin accessibility and expression across the HOXA cluster at ALL presentation and AML relapse of case LS01, demonstrating differential DNase hypersensitivity of HOXA cluster. (C) Differential expression across the HOXA cluster in five cases of lineage switch from ALL (upper panel for each case) and AML (lower panel). (D) The list of direct *MLL-AF4* target genes obtained by overlaying SEM cells and t(4;11) patient cells ChIPseq data (Kerry *et al.*, 2017) and differentially expressed genes in our lineage switch and MPAL cases (Venn diagram, upper panel). Highly enriched genes are further listed on the lower panel. (E) GSEA showing loss of chromatin modifying enzymes signature in AML relapse.

Figure S6. High resolution DNase hypersensitivity sequencing. (A) UCSC genome browser screenshot for *RUNX1* focused on an AML-associated DHS with C/EBP occupation as indicated by high resolution DHS-seq and Wellington analysis. FP - footprint. (B) Heat maps showing distal DHS regions specific for ALL presentation on a genomic scale. Red and green indicate excess of positive and negative strand cuts, respectively, per nucleotide position. Sites are sorted from top to bottom in order of decreasing Footprint Occupancy Score. *De novo* motif discovery in distal DHSs unique to ALL as compared to AML relapse, as shown on the table, right panel.

Figure S7. Alternative splicing analysis. (A) RNAseq read counts across all analysed patient samples. (B) Qualitative analysis of exon-exon spanning reads from LS01, LS03 & LS04 samples. Upper panel represents distribution of the unfiltered and unnormalized reads, lower panel represents filtered and voom normalized reads. Expression (abundance) of the exon-exon junctions is presented as log2-transformed counts per million (CPM). (C) Abundance of non-differential versus differential exon-exon junctions in the transcriptome of LSALs (left panel) and MPAL series of samples (right panel). According to limma/diffSplice approach, junctions were divided into non-differential (upper panel) and differential (bottom panel) splicing events. In each boxplot, the horizontal line represents the median of

distribution, box shows the interquartile range, and whiskers are the minimum and maximum. p-values were calculated using the two-sided Mann-Whitney U test. (D) Venn diagrams showing distribution of identified retained intron (RI) across both LSAL and MPAL patients with corresponding GO terms describing cellular components affected in the AML relapse of lineage switched patients. (E) Venn diagrams showing distribution of identified differential exon-exon junctions (DEEj), significantly expressed in the AML relapse or myeloid compartment of MPAL patients with corresponding GO terms describing affected Reactome pathways and cellular components. GO terms analysis has been performed with https://biit.cs.ut.ee/gprofiler/gost under the highest significance threshold, with multiple testing correction (g:SCS algorithm).

Figure S8. The functional role of epigenetic modifiers on leukaemia lineage switching. (A) PHF3 scheme; the K1119I mutation is located at a highly conserved residue. An * (asterisk) indicates positions which have a single, fully conserved residue, a : (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, a . (period) indicates conservation between groups of weakly similar properties - scoring < 0.5 in the Gonnet PAM 250 matrix. (B) Mutual information sub-networks illustrate high centrality of mutated genes *PHF3* (left panel) and *CHD4* (right panel) within the AML/ALL transcriptional network. Size and colour of nodes reflect differential expression between AML and ALL. (C) Myeloid marker CD33 expression level upon knockdown of *CHD4*, *PHF3*, and non-targeting control (NTC) in *MLL*r (t(4;11)) cell line RS4;11 and non-*MLL*r cell lines, 697 and REH cells. (D) CD19 and CD33 surface expression level change upon knockdown of *CHD4* and *PHF3* in the PDX sample generated from the first relapse of patient LS03 (ALL relapse). (E) Protein network analysis of differentially expressed NuRD and PRC1 members, showing interactions between proteins (and clusters, depicted as blue and red colour); interaction strength is shown as increasing thickness of the joining line (https://string-db.org/). (F) Fold change expression level changes of PRC1 members following lineage switched relapse and in MPAL cases. (G) Expression of lineage specific surface markers following *PCGF6*

knockdown in SEM cells. Knockdown of *PCGF6* results in upregulation of *CD33*, *CSFR3* and *KIT* mRNA (upper panel). Surface CD33 expression increases substantially with shPCGF6_1 and shPCGF6_3, compared with shNTC (lower panel). (H) GSEA showing increased expression of PRC2 target genes and impaired function of PRC1 complex in the relapse AML samples. (I) Boxplot of NMF projections of CD19+ cord blood (CB_19pos) and CD33+ cord blood (CB_33pos) populations relative to the derived AML with *MLL*r metagene (red boxes) and B precursor ALL with *MLL*r metagene (turquoise boxes).

Table titles and legends

Supplementary Table 1. Clinical characteristics of patient cases

Supplementary Table 2. B cell receptor repertoires in lymphoid and myeloid populations. The frequency represents the stated specific sequence as a frequency of the total reads of that clonotype.

Supplementary Table 3. Read counts of *MLL-AF4*, wild type (wt), and mutant (mut) sequences from haematopoietic stem/progenitor cell populations sorted from presentation and relapse samples of LS01. The MPP populations first showing clonal mutations of candidates *CHD4* and *PHF3* are highlighted in yellow.

Supplementary Table 4. Differential gene expression across paired presentation/lymphoid and relapse/myeloid samples.

Supplementary Table 5. Alternative splice variants - mRNAs with retained introns (RI).

Supplementary Table 6. Alternative splice variants - genes with differential exon-exon junctions (DEEj).

Supplementary Table 7. Coding mutations identified in exome sequencing. Genes in red lettering are major clonal variants with variant allele frequencies >0.33. Genes in blue lettering are minor clones with variant allele frequencies 0.09-0.32. Genes in black lettering show a variant allele frequency of <0.09. Shared genes between presentation and relapse are indicated with red boxes.

Supplementary Table 8. ARACNE analysis of coding mutations projected onto a mutual information network constructed from published transcriptomic data from ALL and AML cases.

STAR Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD3-APC-H7	BD Biosciences	Cat#641397; RRID:AB_1645731
CD3-FITC	BD Biosciences	Cat#345763
CD3-BV650	Biolegend	Cat#300468
CD10-BV650	BD Biosciences	Cat#563734; RRID:AB_2738393
CD11b-APC/Fire750	Biolegend	Cat#301351
CD14-FITC	BD Biosciences	Cat#555397; RRID:AB_395798
CD14-BV605	Biolegend	Cat#367125
CD16-FITC	BD Biosciences	Cat#335035
CD19-PE-CF594	BD Biosciences	Cat#562294; RRID:AB_11154408
CD19-APCCy7	Biolegend	Cat#363010; RRID:AB_2564193
CD19-BV421	Biolegend	Cat#302234; RRID:AB_11142678
CD19-PECy7	Biolegend	Cat#363011
CD20-PE	BD Biosciences	Cat#345793
CD20-FITC	Biolegend	Cat#302303
CD33-APC	BD Biosciences	Cat#345800
CD33-APCCy7	Biolegend	Cat#366614; RRID:AB_2566416
CD33-BV421	Biolegend	Cat#303416; RRID:AB_2561690
CD34-PerCPCy5.5	BD Biosciences	Cat#347222
CD34-APCCy7	Biolegend	Cat#343514; RRID:AB_1877168
CD34-APC	Biolegend	Cat#343607
CD38-PeCy7	BD Biosciences	Cat#335825
CD38-PE	Biolegend	Cat#303506
CD45RA-BV510	Biolegend	Cat#304142; RRID:AB_2561947
CD56-FITC	BD Biosciences	Cat#345811
CD64-BV711	Biolegend	Cat#305041
CD90-A700	Biolegend	Cat#328120; RRID:AB_2203302
CD90-PerCPCy5.5	Biolegend	Cat#328118; RRID:AB_2303335
CD90-BV421	Biolegend	Cat#328122

CD117-PE	BD Biosciences	Cat#332785
CD117-BV605	BD Biosciences	Cat#562687; RRID:AB_2737721
CD123-BV421	BD Biosciences	Cat#306018; RRID:AB_10962571
HLA-DR-BV786	Biolegend	Cat#307642; RRID:AB_2563461
HLA-DR-A700	Biolegend	Cat#560743; RRID:AB_1727526
HLA-DR-PerCPCy5.5	Biolegend	Cat#307629
Bacterial and Virus Strains		
One Shot STBL3 chemically competent *E. coli*	Invitrogen	Cat#C7373-03
Chemicals, Peptides, and Recombinant Proteins		
Cytox Green-FITC	Life Tech	Cat#S34860
Deoxyribonuclease I (DNase I)	Worthington-Biochem	Cat# LS006328
AmpliTaq Gold	Applied Biosystems	Cat#4311818
Phusion® High-Fidelity PCR Master Mix with HF Buffer	NEB	Cat#M0531L
IL-6	MACS Miltenyi Biotec	Cat#130-095-352
IL-3	R&D Systems	Cat#203-IL-010
IL-7	R&D Systems	Cat#207-IL-005
SCF	StemCell Technologies	Cat#78062.1
TPO	R&D Systems	Cat#288-TP
FLT-3 Ligand	R&D Systems	Cat#308-FK
Critical Commercial Assays		
AMPure XP	Beckman Coulter	Cat#A63880
QIAamp DNA FFPE Tissue Kit	Qiagen	Cat#56404
AllPrep DNA/RNA Mini Kit	Qiagen	Cat#80204
QIAamp DNA Mini Kit	Qiagen	Cat#51306
innuPREP DNA/RNA Mini Kit	Analytik Jena	Cat#845-KS-2080050
RNeasy Mini Kit	Qiagen	Cat#74106
NEBNext Ultra Directional RNA Kit	NEB	Cat#E7420S
REPLI-g Single Cell Kit	Qiagen	Cat#150345
EndoFree Plasmid Maxi Kit	Qiagen	Cat#12362
Fluidigm Access Array Barcode Library for Illumina Sequencers	Fluidigm	Cat#100-4876
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Fisher Scientific	Cat#K1632

Platinum SYBR Green qPCR SuperMix-UDG with ROX, 2X	Invitrogen	Cat#11744500
Deposited Data		
Exome-seq data	This paper	PRJNA547947
Genome-seq data	This paper	PRJNA547815
Ig rearrangement sequencing data	This paper	PRJNA511413
RNA-seq data	This paper	GSE132396
DNase hypersensitivity sequencing data	This paper	GSE130142
Experimental Models: Cell Lines		
SEM	DSMZ	Cat#ACC546; RRID:CVCL_0095
REH	DSMZ	Cat#ACC22
697	DSMZ	Cat#ACC42
RS4;11	DSMZ	Cat#ACC508
MS-5	DSMZ	Cat#ACC441; RRID:CVCL_2128
Human cord blood $CD34^+$ MLL-Af4	(Lin *et al.*, 2016)	N/A
Experimental Models: Organisms/Strains		
NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice	Charles River Laboratories	N/A
M-CSF^h; IL-3/GM-CSF^h; hSIRPA^{tg};TPO^h;Rag2$^-$;γc^- (MISTRG) mice	Jackson Laboratory	N/A
Oligonucleotides		
shPHF3-1	GPP, Broad Institute	TRCN0000019118
shPHF3-2	GPP, Broad Institute	TRCN0000019114
shPHF3-3	GPP, Broad Institute	TRCN0000274376
shCHD4-1	GPP, Broad Institute	TRCN0000380981
shCHD4-2	GPP, Broad Institute	TRCN0000021363
shCHD4-3	GPP, Broad Institute	TRCN0000021360
shPCGF6-1	GPP, Broad Institute	TRCN0000229804
shPCGF6-2	GPP, Broad Institute	TRCN0000073109
shPCGF6-3	GPP, Broad Institute	TRCN0000073109
shAUTS2-1	GPP, Broad Institute	TRCN0000119058
shAUTS2-2	GPP, Broad Institute	TRCN0000304019
shAUTS2-3	GPP, Broad Institute	TRCN0000304081
shNTC (ATCTCGCTTGGGCGAGAGTAAG)	(Schwarzer *et al.*, 2017)	N/A
PHF3 Fw (ATGGACCTGGGCTTGAACTG)	This paper	N/A

PHF3 Rev (TGGTGGTGCACTTTCAGGAG)	This paper	N/A
CHD4 Fw (TGCTGACACAGTTATTATCTATGACTCTGA)	This paper	N/A
CHD4 Rev (ACGCACGGGTCACAAACC)	This paper	N/A
PCGF6 Fw (GGGAAATCCGACGTGCAAT)	This paper	N/A
PCGF6 Rev (GGAGAAACCACAAGACCATAATGA)	This paper	N/A
AUTS2 Fw (AAAAGGACCCGAGGTTGACA)	This paper	N/A
AUTS2 Rev (GCGATGTGAACATGCATAGCA)	This paper	N/A
GAPDH Fw (GAAGGTGAAGGTCGGAGTC)	This paper	N/A
GAPDH Rev (GAAGATGGTGATGGGATTTC)	This paper	N/A
CD19 Fw (TGACCCCACCAGGAGATTCTT)	This paper	N/A
CD19 Rev (CACGTTCCCGTACTGGTTCTG)	This paper	N/A
CD33 Fw (CTCGTGCCCTGCACTTTCTT)	This paper	N/A
CD33 Rev (CCCGGAACCAGTAACCATGA)	This paper	N/A
CSF3R Fw (CCCAGGCGATCTGCATACTT)	This paper	N/A
CSF3R Rev (AACAAGCACAAAAGGCCATTG)	This paper	N/A
KIT Fw (GGACCAGGAGGGCAAGTCA)	This paper	N/A
KIT Rev (GATAGCTTGCTTTGGACACAGACA)	This paper	N/A
Recombinant DNA		
pCMVD8.91 packaging vector	Life Science Market	Cat# PVT2323
pMD2.G envelope vector	Addgene	Cat#12259
pLKO5d.SFFV.miRNA30n	(Schwarzer *et al.*, 2017)	N/A
Software and Algorithms		
Bowtie2	(Langmead *et al.*, 2009)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
RUVSeq	(Risso *et al.*, 2014)	http://bioconductor.org/packages/RUVSeq
DESeq2	(Love, Huber and Anders, 2014)	http://bioconductor.org/packages/DESeq2
GSEA	(Subramanian *et al.*, 2005)	http://software.broadinstitute.org/gsea
ARACNe2	(Margolin *et al.*, 2006)	N/A
MiXCR	(Bolotin *et al.*, 2015)	http://mixcr.milaboratory.com/
MuTect	(Cibulskis *et al.*, 2013)	http://www.broadinstitute.org/cancer/cga/MuTect

MuTect2	(Cibulskis *et al.*, 2013)	http://www.broadinstitute.org/cancer/cga/MuTect
VEP, version 90	(Cibulskis *et al.*, 2013)	https://github.com/Ensembl/ensembl-vep
Tximport	(Soneson, Love and Robinson, 2016)	http://bioconductor.org/packages/tximport/
ISAAC	(Raczy *et al.*, 2013)	https://github.com/sequencing
Strelka	(Saunders *et al.*, 2012)	ftp://strelka@ftp.illumina.com/
Manta	(Chen *et al.*, 2016)	https://github.com/Illumina/manta
Wellington	(Piper *et al.*, 2015)	http://jpiper.github.com/pyDNase
Java Treeview	(Saldanha, 2004)	http://jtreeview.sourceforge.net
Cytoscape	(Shannon *et al.*, 2003)	http://www.cytoscape.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

As Lead Contact, Olaf Heidenreich (O.T.Heidenreich@prinsesmaximacentrum.nl) is responsible for all reagent and resource requests.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient samples and data

Patients were diagnosed by local haematology specialists according to contemporary clinical diagnostic criteria based on morphology and immunophenotypic analysis. All patient samples were collected at the point of diagnosis, remission following treatment or relapse and stored with written informed consent for research in one of six centres (Newcastle Haematology Biobank, Newcastle, UK; University Hospital Schleswig-Holstein, Kiel, Germany; Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia; Haematological Malignancy Diagnostic Service, Leeds, UK; Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands; Cincinnati Children's Hospital Medical Center, Cincinnati, USA). Mononuclear cells were isolated from

bone marrow or peripheral blood by density centrifugation followed by immediate extraction of DNA or RNA, or cryopreservation in the presence of 10% v/v DMSO.

Samples were requested and used in accordance with the ethical approvals granted to each of the local/institutional ethical review boards (NRES Committee North East - Newcastle & North Tyneside 1, UK, reference 07/H0906/109+5; Medical Faculty Christian-Albrechts University, Kiel, reference A 103/08; Dmitry Rogachev National Medical Research Center, Moscow, references MB2008: 22.01.2008, MB2015: 22.01.2015, ALL-REZ-2014: 28.01.2014; Haematological Malignancy Research Network, Yorkshire, UK, reference 04/Q1205/69; Haematological Malignancy Diagnostic Service, Leeds, UK, reference 14/WS/0098; Erasmus MC METC, Netherlands, reference MEC-2016-739; IRB of Cincinnati Children's Hospital, USA, reference 2010-0658) and in accordance with the Declaration of Helsinki. Each patient/sample was allocated an anonymised reference and no identifiable information was shared.

Cell Lines

SEM, 697, REH and RS4;11 cells were grown in RPMI-1640 with 10% fetal bovine serum (FBS, Sigma)] at 37°C in a humidified 5% CO2 incubator. All cell lines contain no PHF3, CHD4 (with exception of the REH cell line - missense mutation at p.R61Q), PCGF6, AUTS2 mutation according to the CCLE (https://portals.broadinstitute.org/ccle) database.

MLL-Af4 transduced cord blood cells (Lin *et al.*, 2016) were cultured in IMDM with 10% FBS, 2 mM Glutamine supplemented with recombinant human SCF, IL-3, IL-6, FLT-3L, and TPO (10 ng/ml each) at 37°C in a humidified 5% CO_2 incubator. To prime towards lymphoid differentiation, the cells were co-cultured with MS-5 (RRID:CVCL_2128, murine) stroma cells

in α-MEM with 10% FBS, 2 mM glutamine supplemented with SCF, FLT-3L, and IL-7 (10 ng/ml each). The cells were demi-populated weekly.

***In Vivo* Mouse Studies**

In vivo studies were conducted in accordance with the UK Animals (Scientific Procedures) Act 1986 under project licences PPL60/4552 and PPL60/4222 following institutional ethical review.

NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG, Charles River Labs and bred in-house) mice (both sexes) aged 8 -10 weeks old were transplanted intra-femorally under isoflurane anaesthesia with 10^6 cells from sample LS01PALL. Sample LS01RAML was transplanted into 5 NSG mice without evidence of engraftment.

MISTRG mice, expressing four human cytokines (M-CSF^h; IL-3/GM-CSF^h; hSIRPA^{tg};TPO^h;Rag2$^-$;γc$^-$, Jax Lab (Rongvaux *et al.*, 2014), bred in house) were transplanted intra-hepatically. Three day old mice (n = 2) were sub-lethally irradiated (1.5 Gy) 24 h before the transplantation. Each mouse received 10^6 mononuclear cells from sample LS01RAML.

Mice were humanely killed before or when they displayed end points specified by the licenses (tumours reached 1.5 cm in diameter, lost >10% weight in 3 consecutive days or 20% at any time, or displayed signs of ill). PDX cells were harvested from engrafted spleen (NSG) or abdominal masses (MISTRG).

METHOD DETAILS

Long-distance inverse PCR experiments

MLL rearrangement sequences were identified using long-distance inverse PCR (LDI-PCR) (Meyer *et al.*, 2005, 2013). DNA (1 µg) from each patient was digested with restriction enzyme *Bam*HI and purified by phenol extraction and ethanol precipitations to remove the residual

enzyme. The digested product was re-ligated by incubation at 16°C overnight in the presence of 5 units T4 DNA ligase, 50 µl total volume, to form DNA circles. Five microliters (100 ng) re-ligated genomic DNA was used for the LDI-PCR.

Five *MLL*-specific primers (primers A-E (Meyer *et al.*, 2005)) were designed and used in four different combinations (A–B, A–C, A–D, and A–E). The LDI-PCR reactions were performed using TripleMaster PCR System (Eppendorf) according to the manufacturer's instructions. The amplicons were separated on 0.8% agarose gels, extracted, and sequenced. Sequencing results were annotated by blasting the human genome database www.ncbi.nlm.nih.gov/genome/seq/Blast.

PCR-based clonality immunoglobulin gene targets

Clonality analysis based on immunoglobulin gene rearrangements was performed according to the BIOMED-2 Concerted Action BMH-CT98-3936 using standardised protocols and primer sets (van Dongen *et al.*, 2003). DNA was amplified by multiplex PCR assays IGH (VH–JH and DH–JH) and IGK (Vκ–Jκ and Kde primers) using AmpliTaq Gold (Applied Biosystems) in duplicate reactions. The PCR reaction consists of 100 ng DNA, 1 X ABI Gold Buffer, 10 pmol of each primer, 200 µM final concentration dNTP, 1.5 mM final concentration $MgCl_2$, and 1 U Taq enzyme in 50 µl volume. The PCR amplification parameters were: one cycle at 95°C for 10 min, thirty five cycles of 95°C for 45 s, 60°C for 45 s, and 72°C for 1.5 min, followed by one cycle at 72°C for 10 min. The duplicate samples were analysed by GeneScan analysis on a Life Technologies 3100 platform.

Exome sequencing

Germline DNA from cases LS08 and LS09 were extracted from formalin fixed paraffin embedded remission bone marrow using QIAamp DNA FFPE Tissue Kit (Qiagen, Cat#56404). Other DNA samples were extracted from either bone marrow or peripheral blood using AllPrep DNA/RNA Mini Kit (Qiagen, Cat#80204), QIAamp DNA Mini Kit (Qiagen,

Cat#51306), or innuPREP DNA/RNA Mini Kit (Analytik Jena, Cat#845-KS-2080050), according to manufacturers' instructions. The exons were captured using SureSelect XT2 Human All Exon V6 (Agilent), and sequenced by paired-end 75 bp sequencing on HiSeq4000 (Illumina), resulting in roughly 45 million reads per sample. DNA from the myeloid and lymphoid cellular compartments derived from MPAL patients samples, were pre-processed with KAPA HyperPlus Kit (Roche) followed by exons enrichment with KAPA HyperCapture Kit (Roche), and sequenced by paired-end 300 bp sequencing on NovaSeq6000 (Illumina), resulting in roughly 25 million reads per sample.

Raw reads were aligned to human reference genome (hg19 or hg38 for lineage switch or MPAL patients, respectively) using Burrows-Wheeler Aligner (BWA) 0.7.12 (Li and Durbin, 2009) and were processed using the Genome Analysis Toolkit (GATK, v3.8 or 4.1) best practices recommended workflow for variant discovery analysis (Mckenna *et al.*, 2010; DePristo *et al.*, 2011; van der Auwera *et al.*, 2013). MuTect (v1.1.7) and MuTect2 (4.1) were used to identify somatic variants for each matched sample pair (Cibulskis *et al.*, 2013). Variants were annotated using Ensembl Variant Effect Predictor (VEP, version 90) (Cibulskis *et al.*, 2013).

RNA sequencing

Total RNA was extracted with AllPrep DNA/RNA Mini Kit (Qiagen, Cat#80204), innuPREP DNA/RNA Mini Kit (Analytik Jena, Cat#845-KS-2080050), or TRIzol (Thermo Fisher Scientific, Cat# 15596026) followed by RNeasy Mini Kit (Qiagen, Cat#74106) from either bone marrow or peripheral blood, according to manufacturers' instructions. Messenger RNA was captured using NEBNext Ultra Directional RNA Kit in combination with NEBNext poly(A) mRNA Magnetic Isolation Module or KAPA RNA HyperPrep Kit with RiboErase (HMR) in case of lineage switch or MPAL patients respectively, and submitted for paired-end 150 bp sequencing on HiSeq4000 (Illumina) or paired-end 300 bp sequencing on NovaSeq6000 (Illumina) depending on the analysed patients group. For each sample, transcript abundance was

quantified from raw reads with Salmon (version 0.8.2) (Patro *et al.*, 2017) using the reference human transcriptome (hg38) defined by GENCODE release 27. An R package Tximport (version 1.4.0) (Soneson, Love and Robinson, 2016) was used to estimate gene-level abundance from Salmon's transcript-level counts. Gene-level differential expression analysis was performed using DESeq2 (version 1.16.1) (Love, Huber and Anders, 2014).

Alternative splicing analysis

The differential splicing events were detected in RNAseq data obtained from all analysed patient samples. Detailed analyses were restricted to samples with more than 15 million reads per sample, which included four lineage switched (LS01, LS03, LS04 and LS10) and both MPAL patients. The differential splicing events, including exon-exon junctions (DEEj), differential exon usage (DEU) and retained introns (RI) were identified in both presentation/relapse pairs or lymphoid/myeloid cellular fractions, using pipeline described previously (Grinev *et al.*, 2021). Validation of identified different splice variants was performed by real-time qPCR experiments.

Whole genome sequencing

Presentation, remission and relapse DNA samples from case LS01 were sequenced by Illumina UK and analysed using the remission sample as the matching normal. Sequencing reads were aligned to the human GRCh37.1 reference genome using ISAAC (Raczy *et al.*, 2013). Identification of somatic SNVs and small somatic indels (<50 bp) was performed by Strelka (Saunders *et al.*, 2012). Large structural variants (including deletions, inversions, duplications and insertions all >50bp and translocations) were called by Manta (Chen *et al.*, 2016).

Non-negative matrix factorization analysis

Non-negative matrix factorisation (NMF) was performed using data sets obtained with permission. Data from Bolouri *et al.* (Bolouri *et al.*, 2018) were generated by the

Therapeutically Applicable Research to Generate Effective Treatments (TARGET) initiative, phs000218, managed by the NCI. The data used for this analysis are available [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000218.v21.p7]. Information about TARGET can be found at http://ocg.cancer.gov/programs/target. Data from Andersson *et al.* (Andersson *et al.*, 2015) was obtained with permission of the Pediatric Cancer Genome Project from the European Genome-Phenome Archive (EGA study accession EGAS00001000246). Data from Novoshtern *et al.* (Novershtern *et al.*, 2011) was obtained openly from Gene Expression Omnibus (GEO dataset GSE24759). NMF was used to extract metagenes from three model gene expression datasets including only leukaemic samples for which information on recurrent genetic abnormalities was known (Novershtern *et al.*, 2011; Andersson *et al.*, 2015; Bolouri *et al.*, 2018). We classified samples into the following 7 subgroups of interest: ALL without *MLL*r, ALL with *MLL-AF4* fusion, ALL with other types of *MLL*r, AML with *MLL*r, AML with inv(16), AML with *RUNX1-ETO* fusion, AML with normal karyotype. A matrix of gene expression data combining per gene read counts of samples from each dataset was created, normalized and variance-stabilized transformed (vst) using DESeq2 (Love, Huber and Anders, 2014). Batch effect between leukaemia derived datasets (Andersson *et al.*, 2015; Bolouri *et al.*, 2018) was minimized using the remove unwanted variation method from RUVSeq (Risso *et al.*, 2014) on upper-quartile normalized counts whilst retaining variation associated with biological covariates of interest. To remove genes that did not vary sufficiently across the model dataset, for RNAseq data, we capped expression values at 20 and 10,000 read counts and only included genes with > 10-fold and >1000 read counts between minimum and maximum. For the Affymetrix U133A microarray, we capped expression values at 0.01 and 10,000 units and only probes with > 3-fold and >5 units difference between minimum and maximum were considered in the NMF analysis. To identify the most robust combination of metagenes and subgroups/clusters in the model datasets, consensus bootstrapped NMF clustering was performed on filtered gene count matrix as previously described (Schwalbe *et al.*, 2013). Briefly, we performed NMF and K-means clustering, testing all combinations of 2-15 metagenes and clusters with bootstrapped

resampling method (n=100) to test for reproducibility. Cluster stability measures (Cohen's kappa, average silhouette scores) were assessed to determine optimal combinations of metagenes and clusters. Samples assigned to the same cluster fewer than 90% of replicates were removed from the dataset. All genes in the refined model dataset were then column rank normalised and metagenes scores were recalculated.

Metagenes were projected onto our RNAseq datasets (both lineage switch cases and *MLL-Af4* cord blood samples) by the pseudoinverse method as outlined in (Brunet *et al.*, 2004; Tamayo *et al.*, 2007). All analysis and figures were generated using R/Bioconductor and by using a modified version of the NMF scripts provided by (Brunet *et al.*, 2004).

DHS library generation, sequencing, and mapping

DHS analysis was performed as described previously (Ptasinska *et al.*, 2014; Cauchy *et al.*, 2015). DNase I (Worthington, Cat# LS006328) digestion was performed using ~5 million patient sample cells using 8 units (presentation) or 14 units (relapse) for 3 min at 22°C in a 1 mM $CaCl_2$ supplemented buffer. Nuclear proteins were digested with 1 mg/ml Proteinase K overnight at 37°C. DNase I digestion products were size-selected on an agarose gel, cutting below 150 bp. High-throughput sequencing libraries were prepared from 10 ng of size-selected material, using the Kapa Hyperprep kit as per manufacturer's instruction. Libraries were sequenced with 50 bp single-end reads on an Illumina HiSeq 2500 sequencer according to manufacturer's instructions.

Fastq files were generated using bcl2fastq (1.8.4) and subsequently aligned to the hg19 assembly (NCBI Build 37) with the use of bowtie2 (2.1.0) (Langmead *et al.*, 2009), with –very-sensitive-local as a parameter. Read coverage generation and peak detection were carried out using MACS 1.4.1 (Zhang *et al.*, 2008) using --keep-dup=all –g hs -w -S. Pairwise comparisons were performed as previously described (Cauchy *et al.*, 2015). Digital footprinting was carried out using the Wellington package using default parameters (Piper *et al.*, 2013). Differential footprinting analysis was carried out on footprints using the Wellington-

bootstrap (Piper *et al.*, 2015) package with default parameters. Average profiles and heatmaps were obtained using the functions dnase_average_profile and dnase_to_javatreeview from the Wellington package. Heatmaps were plotted using Java TreeView (Saldanha, 2004).

Flow cytometry and cell sorting

One million cells (of all tested cell lines, cord blood or LS03ALL PDX cells) were used for flow cytometry analysis of CD19 (BV421) and CD33 (APCCy7). The antibody details are listed in the Key Resources Table. Cells were collected and washed in a 3 ml wash buffer containing 0.2% BSA in PBS, re-suspended in 100 µl wash buffer and 5 µl each antibody was added, followed by incubation at RT for 20 min in the dark. Cells were washed and re-suspended in a 0.5 ml wash buffer. Flow cytometry was performed on FACSCanto II from Becton Dickinson (BD) and data analysed with FlowJo (Treestar).

Haematopoietic hierarchy analysis was performed from 10^7 cells of primary and PDX samples. The cells were collected and incubated for 30 min in a wash buffer containing 0.5% FBS, 2 mM EDTA in PBS. Following the incubation, the cells were washed and resuspended in a wash buffer. Fluorescence activated cell sorting (FACS) was performed using a FACS Aria Fusion (BD) or ASTRIOS EQ (Beckman Coulter). In order to separate early progenitors or lymphoid and myeloid fractions, MPAL patient samples were sorted from 10^7 or $3*10^6$ cells for MPAL1 and MPAL2 respectively. Sorted cells were collected into 1.5 ml microfuge tubes containing 500 µl RPMI with 10% fetal calf serum or SFEM II media in case of cell lines or primary samples, respectively. In order to analyse single cell derived clones, sorted early progenitor cells derived from MPAL samples were deposited onto 384-well plates containing 75ul of SFEM II media (supplemented with SCF, FLT3L, IL6, IL3 and TPO). Single cells were grown into 60-90% confluency colonies followed by RNA/DNA extraction and MLL-AF4 detection by PCR.

977 **Single cell sorting and whole genome amplification (WGA)**

978 Single cells were labelled (see flow cytometry and cell sorting section above) and sorted into
979 96 well plate (Eppendorf twin.tec PCR Plate 96 full skirt) using FACSAria Fusion (BD). The
980 plates were filled with 3.5 µl PBS/well prior to the sorting. Each plate included forty-five single
981 cell samples, two 0 cells as negative control, and one bulk (300 – 1,000 cells) as positive
982 control sample. The cells were prepared from as many as 10^5 cells in 500 µl buffer (0.5% FBS,
983 2 mM EDTA in PBS).

984 Cell sorting was performed using the "single cell" setting on FACSAria Fusion. Following the
985 sorting, the plate was centrifuged at 900 g for 1 min, snap-frozen on dry ice, and then kept at
986 -20°C until the whole genome amplification (WGA) procedure.

987 WGA was performed using REPLI-g Single Cell Kit (Qiagen, Cat#150345). The cells were
988 lysed by adding 1.5 µl Buffer D2 and incubation at 65°C for 10 min in a thermal cycler (HYBAID
989 PCRExpress). Subsequently, the reaction was terminated by adding 1.5 µl Stop Solution and
990 stored on the ice. The mixture of 20 µl REPLI-g sc Reaction Buffer and REPLI-g sc DNA
991 Polymerase was added and incubated at 30°C for 8 h. The reaction was terminated by heat
992 inactivation at 65°C for 3 min. The products were diluted 100 fold with TE Buffer ahead of PCR
993 amplification.

994 **Nested multiplex PCR and targeted sequencing**

995 The 12 mutation candidate driver genes and *MLL-AF4* LS01RAML were amplified using gDNA
996 as the template by nested multiplex PCR method. The primers were designed using Primer
997 Express (Applied Biosystems) software.

998 PCR amplifications were carried out using Phusion® High-Fidelity PCR Master Mix with HF
999 Buffer (NEB, Cat#M0531L) according to the manufacturer's instructions (25 µl reaction), plus
1000 80 – 200 nM of each primer set. The first multiplex reaction parameters were: one cycle at
1001 98°C for 2 min, thirty cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 30 s, followed by one

cycle at 72°C for 10 min. The products were diluted 500-fold, and 1 µl used as the template for the second PCR reactions (25 µl reaction). The second/nested multiplex reaction parameters were: one cycle at 98°C for 2 min, twenty cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 30 s, followed by one cycle at 72°C for 10 min. The amplicons that were taken forward for next-generation targeted sequencing had additional CS1 (ACACTGACGACATGGTTCTACA) and CS2 (TACGGTAGCAGAGACTTGGTCT) Fluidigm tag sequences on the nested PCR primer forward and reverse, respectively. The amplicons were barcoded using Fluidigm Access Array Barcode Library for Illumina Sequencers (Cat#100-4876) by taking 0.8 µl multiplex PCR products (multiplex group A-C), 4 µl Fluidigm barcode primer (400 nM final concentration), 10 µl of 2X Phusion Master Mix (NEB, Cat#M0531L), and 3.6 µl H_2O (20 µl reaction). The barcoding PCR reaction parameters were: one cycle at 98°C for 2 min, six cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 1 min, followed by one cycle at 72°C for 10 min. The products were run on 2% agarose gel and extracted using the QIAquick Gel Extraction Kit (Qiagen, Cat#28706). The purified products were submitted for paired-end 300 bp sequencing on MiSeq (Illumina), resulting in >1,000 coverage per gene.

***MLL-Af4* stem cell expression analysis**

Following myeloid or lymphoid culture of *MLL/Af4* transduced CD34+ cord blood cells, CD19+CD33-, CD19-CD33+ and CD19-CD33- populations were flow sorted, lysed and RNA extracted using RNeasy Micro Kit (Qiagen), according to the manufacturer's instructions. Input RNA was equilibrated to a starting input cell number of 300 cells per population before cDNA and sequencing library production were performed using SMARTSeqv4 (Clontech) and NexteraXT (Illumina) kits, according to manufacturer's instructions. The resultant libraries were submitted for paired-end 150 bp sequencing on a NEXTSeq500 (Illumina). For each sample, transcript abundance was quantified from raw reads with Salmon (version 0.8.2) (Patro *et al.*, 2017) using the reference human transcriptome (hg38) defined by GENCODE release 27. An R package Tximport (version 1.4.0) (Soneson, Love and Robinson, 2015) was

used to estimate gene-level abundance from Salmon's transcript-level counts. Gene-level differential expression analysis was performed using DESeq2 (version 1.16.1) (Love, Huber and Anders, 2014) prior NMF analysis as described above.

***PHF3*, *CHD4*, *PCGF6*, and *AUTS2* shRNAs**

shRNAs against PHF3 (TRCN0000019118, TRCN0000019114, TRCN0000274376), CHD4 (TRCN0000380981, TRCN0000021363, TRCN0000021360), PCGF6 (TRCN0000229804, TRCN0000073109, TRCN0000073109), AUTS2 (TRCN0000119058, TRCN0000304019, TRCN0000304081), or non-targeting control (ATCTCGCTTGGGCGAGAGTAAG) were cloned into pLKO5d.SFFV.miR30n (Schwarzer *et al.*, 2017). Each target gene was linked with different fluorescent protein, including dTomato (PHF3 and AUTS2), eGFP (CHD4 and PCGF6), and RFP657 (non-targeting control).

The vector contained BsmBI at the cloning sites. The oligonucleotides were designed with the appropriate complementary overhang sequences of BsmBI-cleaved vector. They were ligated and transformed into STBL3 chemically competent cells (Invitrogen) according to manufacturer's instructions and plated on agar plates containing 100 µg/ml ampicillin incubated for 16 h at 37°C. Single colonies were inoculated for DNA preparation, and extracted using EndoFree Plasmid Maxi Kit (Qiagen, Cat#12362). All clones were verified by sequencing prior to lentivirus production.

Lentivirus Production and Transduction

Viral particles were produced using calcium phosphate precipitation method on 293T cells by co-transfecting an envelope plasmid pMD2.G, packaging plasmid pCMVΔR8.91, and the shRNA vector pLKO5d.SFFV.miR30n. The cells were grown in 100 mm tissue culture dishes at a concentration of 1-2 x 10^6 cells in 10 ml medium the day prior to co-transfection. On the following day when the cells had reached 30-50% confluence, 20 µg of shRNA vector, 15 µg of pCMVΔR8.91, and 5 µg of pMD2.G were mixed. The volume of the mixture was adjusted

with HEPES buffer solution (2.5 mM HEPES containing deionized water at pH7.3) to 250 µl. A volume of 250 µl of 0.5 M $CaCl_2$ was added to the mixture. This solution was added to 500 µl of 2X HeBS (0.28 M NaCl, 0.05 M HEPES and 1.5 mM Na_2HPO_4 in deionized water at pH 7.00) and mixed by vortexing. It was incubated at RT for 30-40 min to allow the formation of the calcium phosphate precipitate, before adding dropwise on the 293T cells. After 16 h, the cells were washed with 10 ml PBS and added with 10 ml culture media. They were incubated at 37°C in a humidified atmosphere with 5% CO_2 for the next two days. Lentivirus particles were collected by centrifuging the supernatant at 400 g for 10 min at 4°C and filtered through Acrodisc Syringe 0.45 filters. They were stored in aliquots at -80°C. Cell lines and PDX samples were transduced with lentivirus as previously described (Bomken *et al.*, 2013).

Lymphoid differentiation of transduced *MLL-Af4* cord blood cells

MLL-Af4 cord blood cells (5 x 10^5 cells per transduction, Lin *et al.*, 2016) were transduced with short hairpin constructs as described and co-cultured with MS-5 stroma cells in lymphoid culture conditions described above. Single and triple transduced populations were identified using the construct specific fluorophores and lineage specific surface markers assessed as a proportion of the total transduced leukocyte population.

qRT-PCR

One million cells were collected and RNA was extracted using RNeasy Mini Kit (Qiagen, Cat#74106). cDNA was synthesised from 1 µg RNA using RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Cat#K1632). The product was diluted by adding 80 µl H_2O (Livak and Schmittgen, 2001). Sequences of primers used in this study were designed using Primer Express (Applied Biosystems) software.

QUANTIFICATION AND STATISTICAL ANALYSIS

Reverse Engineering of Transcriptional Networks

A mutual information network was developed such that the connections between genes within the network represented an inferred likelihood of a causal relationship within ALL/AML. The network was further refined by retaining only those nodes (genes) and edges (causal connections) which are significantly associated, in expression terms, to the distinction between normal myeloid and lymphoid cells. Thus, the centrality of the mutations within the network reflects the estimated extent of their causal influence upon a lymphoid/myeloid distinction within the framework of primary ALL/AML.

Mutual Information networks were reverse engineered from 216 published AML/ALL Affymetrix HGU133p2 expression profiles (GSE11877 & GSE24006) processed using Rma (affy package R/Bioconductor) using the ARACNe2 algorithm (Margolin *et al.*, 2006) set to adaptive partitioning with a DPI tolerance of 0.1 and a p-value threshold of 0.01 and 1000 bootstraps. The probes/genes were filtered to remove those that were insufficiently variable as per the Aracne method. Probes had to be greater than 3 fold expression and 300 delta between the min and max excluding the 5% most extreme values to be included. Probes without >20 intensity in greater than 20% of samples were also removed. All probes which fulfilled these requirements were included in the construction of this network (n=9780).

Nodes within the network were annotated with a metagene value calculated using Non-Negative Matrix Factorisation (k=2) and reflecting the differences between myeloid and lymphoid cells (GSE24759) and the other. A metagene cutoff of 0.13 (based on a calculated alpha of 0.1) was applied to trim the network. Cytoscape (Shannon *et al.*, 2003) was used to merge nodes where probe sets represented common genes and to calculate centrality statistics for each of the mutated genes of interest. Nodes of interest (i.e. mutated nodes) were ranked according to their centrality (e.g. degree). Nodes were sized and coloured according

to their differential expression between AML and ALL types; size = significance, colour = log2 fold change.

DATA AND SOFTWARE AVAILABILITY

Exome sequencing data and genome sequencing data presented in this manuscript have been deposited in the NCBI Sequence Read Archive (SRA) under project numbers PRJNA547947 and PRJNA547815 respectively. Immunoglobulin and TCR sequencing data have been deposited in NCBI SRA under project number PRJNA511413. RNA sequencing data and DNase hypersensitivity sequencing data were deposited in Gene Expression Omnibus under accession numbers GSE132396 and GSE130142 respectively. All deposited data will be publically available following publication of the manuscript.

References

Agraz-Doblas, A., Bueno, C., Bashford-Rogers, R., Roy, A., Schneider, P., Bardini, M., Ballerini, P., Cazzaniga, G., Moreno, T., Revilla, C., Agraz-Doblas, A., *et al*. (2019). Unravelling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. Haematologica, 104, 1176-1188. 10.3324/haematol.2018.206375.

Alexander, T.B., Gu, Z., Iacobucci, I., Dickerson, K., Choi, J.K., Xu, B., Payne-Turner, D., Yoshihara, H., Loh, M.L., Horan, J., *et al*. (2018) The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature 562, 373-379. 10.1038/s41586-018-0436-0.

Alexandrov, L. B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., Stratton, M.R. (2015) Clock-like mutational processes in human somatic cells. Nature Genetics 47, 1402–1407. 10.1038/ng.3441.

Andersson, A.K., Ma, J., Wang, J., Chen, X., Gedman, A.L., Dang, J., Nakitandwe, J., Holmfeldt, L., Parker, M., Easton, J., *et al*. (2015). The landscape of somatic mutations in infant *MLL*-rearranged acute lymphoblastic leukemias. Nature Genetics 47, 330-337. 10.1038/ng.3230.

Appel, L-M., Franke, V., Bruno, M., Grishkovskaya, I., Kasiliauskaite, A., Schoeberl, U.E., Puchinger, M.G., Kostrhon, S., Beltzung, E., Mechtler, K., *et al*. (2020) PHF3 regulates neuronal gene expression through the new Pol II CTD reader domain SPOC. bioRxiv, p. 2020.02.11.943159. 10.1101/2020.02.11.943159.

Arends, T., Dege, C., Bortnick, A., Danhorn, T., Knapp, J.R., Jia, H., Harmacek, L., Fleenor, C.J., Straign, D., Walton, K., *et al*. (2019). CHD4 is essential for transcriptional repression and lineage progression in B lymphopoiesis. Proc Natl Acad Sci U S A 116, 10927-10936. 10.1073/pnas.1821301116.

Ayton, P. M., and Cleary, M. L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9. Genes Dev 17, 2298–2307. 10.1101/gad.1111603.

Ferreira, B.I., García, J.F., Suela, J., Mollejo, M., Camacho, F.I., Carro, A., Montes, S., Piris, M.A., Cigudosa, J.C. (2008). Comparative genome profiling across subtypes of low-grade B-cell lymphoma identifies type-specific and common aberrations that target genes with a role in B-cell neoplasia. Haematologica 93, 670–679. 10.3324/haematol.12221.

Blokzijl, F. de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. Nature 538, 260–264. doi: 10.1038/nature19768.

Boer, J.M., van der Veer, A., Rizopoulos, D., Fiocco, M., Sonneveld, E., de Groot-Kruseman, H.A., Kuiper, R.P., Hoogerbrugge, P., Horstmann, M., Zaliova, M., *et al*. (2016). Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. Leukemia, 30, 32–38. 10.1038/leu.2015.199.

Bolotin, D.A., Poslavsky, S., Davydov, A.N., Frenkel, F.E., Fanchi, L., Zolotareva, O.I., Hemmers, S., Putintseva, E.V., Obraztsova, A.S., Shugay, M., *et al*. (2017). Antigen receptor repertoire profiling from RNA-seq data. Nat Biotechnol, 35, 908–911. 10.1038/nbt.3979.

Bolouri, H., Farrar, J.E., Triche, T.Jr., Ries, R.E., Lim, E.L., Alonzo, T.A., Ma, Y., Moore, R., Mungall, A.J., Marra, M.A., *et al*. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med, 24, 103–112.

10.1038/nm.4439 PMID - 29227476.

Bomken, S., Buechler, L., Rehe, K., Ponthan, F., Elder, A., Blair, H., Bacon, C.M., Vormoor, J., Heidenreich, O. (2013). Lentiviral marking of patient-derived acute lymphoblastic leukaemic cells allows in vivo tracking of disease progression. Leukemia, 27, 718-721. 10.1038/leu.2012.206

Brunet, J.P., Tamayo, P., Golub, T.R., Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 101, 4164–4169. 10.1073/pnas.0308531101.

Carulli, G., Marini, A., Ferreri, M.I., Azzarà, A., Ottaviano, V., Lari, T., Rocco, M., Giuntini, S., Petrini, M. (2012). B-cell acute lymphoblastic leukemia with t(4;11)(q21;q23) in a young woman: evolution into mixed phenotype acute leukemia with additional chromosomal aberrations in the course of therapy. Hematol Rep 4, e15. 10.4081/hr.2012.e15.

Cauchy, P., James, S.R., Zacarias-Cabeza, J., Ptasinska, A., Imperato, M.R., Assi, S.A., Piper, J., Canestraro, M., Hoogenkamp, M., Raghavan, M., *et al*. (2015). Chronic FLT3-ITD Signaling in Acute Myeloid Leukemia Is Connected to a Specific Chromatin Signature. Cell Rep 12, 821–836. 10.1016/j.celrep.2015.06.069.

Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., *et al*. (2014). Transcriptional diversity during lineage commitment of human blood progenitors. Science 345, 1251033. 10.1126/science.1251033.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222. 10.1093/bioinformatics/btv710.

Church, D.N., Briggs, S.E., Palles, C., Domingo, E., Kearsey, S.J., Grimes, J.M., Gorman, M., Martin, L., Howarth, K.M., Hodgson, S.V., *et al*. (2013). DNA polymerase ε and δ exonuclease domain mutations in endometrial cancer. Hum Mol Genet 22, 2820–2828. 10.1093/hmg/ddt131.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213–219. 10.1038/nbt.2514.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., *et al*. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491–498. 10.1038/ng.806.

Dobbins, S.E., Sherborne, A.L., Ma, Y.P., Bardini, M., Biondi, A., Cazzaniga, G., Lloyd, A., Chubb, D., Greaves, M.F., Houlston, R.S. (2013). The silent mutational landscape of infant MLL-AF4 pro-B acute lymphoblastic leukemia. Genes Chromosomes Cancer 52, 954–960. 10.1002/gcc.22090.

Erson-Omay, E.Z., Çağlayan, A.O., Schultz, N., Weinhold, N., Omay, S.B., Özduman, K., Köksal, Y., Li, J., Serin Harmancı, A., Clark, V., *et al*. (2015). Somatic POLE mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis. Neuro Oncol 17, 1356–1364. 10.1093/neuonc/nov027.

Gardner, R., Wu, D., Cherian, S., Fang, M., Hanafi, L.A., Finney, O., Smithers, H., Jensen, M.C., Riddell, S.R., Maloney, D.G., *et al*. (2016). Acquisition of a CD19-negative myeloid phenotype allows immune escape of *MLL*-rearranged B-ALL from CD19 CAR-T-cell therapy. Blood 127, 2406-2410. 10.1182/blood-2015-08-665547.

1192 Germano, G., Pigazzi, M., del Giudice, L., Campo Dell'Orto, M., Spinelli, M., Zangrando, A., Paolucci, P.,
1193 Ladogana, S., Basso, G. (2006). Two consecutive immunophenotypic switches in a child with MLL-
1194 rearranged acute lymphoblastic leukemia. Haematologica 91, ECR09.

1195 Gentles, A.J., Plevritis, S.K., Majeti, R., Alizadeh, A.A. (2010). Association of a leukemic stem cell gene
1196 expression signature with clinical outcomes in acute myeloid leukemia. JAMA 304, 2706–2715.
1197 10.1001/jama.2010.1862.

1198 Gessner, A., Thomas, M., Castro, P.G., Büchler, L., Scholz, A., Brümmendorf, T.H., Soria, N.M., Vormoor,
1199 J., Greil, J., Heidenreich, O. (2010). Leukemic fusion genes MLL/AF4 and AML1/MTG8 support leukemic
1200 self-renewal by controlling expression of the telomerase subunit TERT. Leukemia 24, 1751–1759.
1201 10.1038/leu.2010.155.

1202 González-Pérez, A., and López-Bigas, N. (2011). Improving the assessment of the outcome of
1203 nonsynonymous SNVs with a consensus deleteriousness score, Condel', Am J Hum Genet 88, 440–449.
1204 10.1016/j.ajhg.2011.03.004.

1205 Grinev, V.V., Barneh, F., Ilyushonak, I.M., Nakjang, S., Smink, J., van Oort, A., Clough, R., Seyani, M.,
1206 McNeill, H., Reza, M., *et al*. (2021). RUNX1/RUNX1T1 mediates alternative splicing and reorganises the
1207 transcriptional landscape in leukemia. Nat Commun 12, 520. 10.1038/s41467-020-20848-z.

1208 Hamieh, M., Dobrin, A., Cabriolu, A., van der Stegen, S.J.C., Giavridis, T., Mansilla-Soto, J., Eyquem, J.,
1209 Zhao, Z., Whitlock, B.M., Miele, M.M., *et al*. (2019). CAR T cell trogocytosis and cooperative killing
1210 regulate tumour antigen escape. Nature 568, 112–116. 10.1038/s41586-019-1054-1.

1211 Harman, J.R., Thorne, R., Jamilly, M., Tapia, M., Crump, N.T., Rice, S., Beveridge, R., Morrissey, E., de
1212 Bruijn, M.F.T.R., Roberts, I., *et al*. (2021) A KMT2A-AFF1 gene regulatory network highlights the role
1213 of core transcription factors and reveals the regulatory logic of key downstream target genes. Genome
1214 Res, Jun 4 online ahead of print. doi: 10.1101/gr.268490.120.

1215 Harvey, R.C., Mullighan, C.G., Wang, X., Dobbin, K.K., Davidson, G.S., Bedrick, E.J., Chen, I.M., Atlas,
1216 S.R., Kang, H., Ar, K. *et al*. (2010). Identification of novel cluster groups in pediatric high-risk B-
1217 precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide
1218 DNA copy number alterations, clinical characteristics, and outcome. Blood. 116, 4874–4884.
1219 10.1182/blood-2009-08-239681.

1220 Heshmati, Y., Türköz, G., Harisankar, A., Kharazi, S., Boström, J., Dolatabadi, E.K., Krstic, A., Chang, D.,
1221 Månsson, R., Altun, M., *et al*. (2018). The chromatin-remodeling factor CHD4 is required for
1222 maintenance of childhood acute myeloid leukemia. Haematologica 103, 1169-1181.
1223 10.3324/haematol.2017.183970

1224 Hohaus, S., Petrovick, M.S., Voso, M.T., Sun, Z., Zhang, D.E., Tenen, D.G. (1995). PU.1 (Spi-1) and C/EBP
1225 alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha
1226 gene. Mol Cell Biol 15, 5830–5845. 10.1128/mcb.15.10.5830.

1227 Hosokawa, H., Tanaka, T., Suzuki, Y., Iwamura, C., Ohkubo, S., Endoh, K., Kato, M., Endo, Y., Onodera,
1228 A., Tumes, D.J., *et al*. (2013). Functionally distinct Gata3/Chd4 complexes coordinately establish T
1229 helper 2 (Th2) cell identity. Proc Natl Acad Sci U S A 110, 4691–4696. 10.1073/pnas.1220865110.

1230 Ivanov, I.C., Jitam, D., Grigore, G.E., Zlei, M., Ivanoc, A.V., Dumitras, S., Carasevici, E., Miron, I.C. (2013)
1231 Infant acute leukemia with lineage switch at relapse expressing a novel t(4;11)(q21;q23) MLL-AF4
1232 fusion transcript. Romanian Review of Laboratory Medicine 21, 47–58. 10.2478/rrlm-2013-0017.

1233 Jiang, J.G., Roman, E., Nandula, S.V., Murty, V.V., Bhagat, G., Alobeid, B. (2005). Congenital MLL-
1234 positive B-cell acute lymphoblastic leukemia (B-ALL) switched lineage at relapse to acute myelocytic
1235 leukemia (AML) with persistent t(4;11) and t(1;6) translocations and JH gene rearrangement. Leuk
1236 Lymphoma 46, 1223-1227. 10.1080/10428190500086055.

1237 Jung, R., Jacobs, U., Krumbholz, M., Langer, T., Keller, T., De Lorenzo, P., Valsecchi, M.G., van der
1238 Velden, V.H., Moericke, A., Stanulla, M., *et al*. (2010). Bimodal distribution of genomic MLL
1239 breakpoints in infant acute lymphoblastic leukemia treatment. Leukemia 24, 903–907.
1240 10.1038/leu.2010.14.

1241 Kang, H., Chen, I.M., Wilson, C.S., Bedrick, E.J., Harvey, R.C., Atlas, S.R., Devidas, M., Mullighan, C.G.,
1242 Wang, X., Murphy, M., *et al*. (2010). Gene expression classifiers for relapse-free survival and minimal
1243 residual disease improve risk classification and outcome prediction in pediatric B-precursor acute
1244 lymphoblastic leukemia. Blood 115, 1394–1405. 10.1182/blood-2009-05-218560.

1245 Kerry, J., Godfrey, L., Repapi, E., Tapia, M., Blackledge, N.P., Ma, H., Ballabio, E., O'Byrne, S., Ponthan,
1246 F., Heidenreich, O., *et al*. (2017). MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from
1247 Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. Cell Rep 18, 482–495.
1248 10.1016/j.celrep.2016.12.054.

1249 Kim, J., Sif, S., Jones, B., Jackson, A., Koipally, J., Heller, E., Winandy, S., Viel, A., Sawyer, A., Ikeda, T.,
1250 *et al*. (1999). Ikaros DNA-Binding Proteins Direct Formation of Chromatin Remodeling Complexes in
1251 Lymphocytes. Immunity 10, 345-355. 10.1016/s1074-7613(00)80034-5.

1252 Kinkelin, K., Wozniak, G.G., Rothbart, S.B., Lidschreiber, M., Strahl, B.D., Cramer, P. (2013). Structures
1253 of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. Proc Natl
1254 Acad Sci U S A 110, 15277–15282. 10.1073/pnas.1311010110.

1255 Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of
1256 short DNA sequences to the human genome. Genome Biol 10, R25. 10.1186/gb-2009-10-3-r25.

1257 Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G.,
1258 Schönheit, J., Vockentanz, L., *et al*. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus
1259 in B cells and myeloid cells. Blood 117, 2827–2838. 10.1182/blood-2010-08-302976.

1260 le Viseur, C., Hotfilder, M., Bomken, S., Wilson, K., Röttgers, S., Schrauder, A., Rosemann, A., Irving, J.,
1261 Stam, R.W., Shultz, L.D., *et al*. (2008) In childhood acute lymphoblastic leukemia, blasts at different
1262 stages of immunophenotypic maturation have stem cell properties. Cancer Cell, 14, 47–58.
1263 10.1016/j.ccr.2008.05.015

1264 Li, B., Brady, S.W., Ma, X., Shen, S., Zhang, Y., Li, Y., Szlachta, K., Dong, L., Liu, Y., Yang, F., *et al*. (2020).
1265 Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia.
1266 Blood 135, 41–55. 10.1182/blood.2019002220.

1267 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
1268 Bioinformatics 25, 1754–1760. 10.1093/bioinformatics/btp324.

1269 Lin, S., Luo, R.T., Ptasinska, A., Kerry, J., Assi, S.A., Wunderlich, M., Imamura, T., Kaberlein, J.J., Rayes,
1270 A., Althoff, M.J., *et al*. (2016). Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11)
1271 Pro-B Acute Lymphoblastic Leukemia. Cancer Cell 30, 737–749. 10.1016/j.ccell.2016.10.008.

1272 Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time
1273 Quantitative PCR and the 2−ΔΔCT Method. Methods 25, 402–408. 10.1006/meth.2001.1262.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. 10.1186/s13059-014-0550-8.

Lu, X., Chu, C.S., Fang, T., Rayon-Estrada, V., Fang, F., Patke, A., Qian, Y., Clarke, S.H., Melnick, A.M., Zhang, Y., *et al*. (2019). MTA2/NuRD Regulates B Cell Development and Cooperates with OCA-B in Controlling the Pre-B to Immature B Cell Transition. Cell Rep 28, 472-485. 10.1016/j.celrep.2019.06.029.

Malouf, C., and Ottersbach, K. (2018). The fetal liver lymphoid-primed multipotent progenitor provides the prerequisites for the initiation of t(4;11) MLL-AF4 infant leukemia. Haematologica, 103, e571–e574. 10.3324/haematol.2018.191718.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. BMC Bioinformatics 7, S7. doi: 10.1186/1471-2105-7-S1-S7.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303. 10.1101/gr.107524.110.

Meyer, C., Schneider B., Dingermann, T., Klingebiel, T., Marschalek, R. (2005). A LDI-PCR based method allows the identification of any MLL rearrangement. Blood 106, 2848. 10.1182/blood.V106.11.2848.2848.

Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M., *et al*. (2009). New insights to the MLL recombinome of acute leukemias. Leukemia 23, 1490–1499. 10.1038/leu.2009.33.

Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., *et al*. (2013). The MLL recombinome of acute leukemias in 2013. Leukemia 27, 2165–2176. 10.1038/leu.2013.135.

Meyer, C., Burmeister, T., Gröger, D., Tsaur, G., Fechina, L., Renneville, A., Sutton, R., Venn, N.C., Emerenciano, M., Pombo-de-Oliveira, M.S. (2018). The MLL recombinome of acute leukemias in 2017. Leukemia 32, 273–284. 10.1038/leu.2017.213.

Moorman, A.V., Ensor, H.M., Richards, S.M., Chilton, L., Schwab, C., Kinsey, S.E., Vora, A., Mitchell, C.D., Harrison, C.J. (2010). Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. Lancet Oncol 11, 429–438. 10.1016/S1470-2045(10)70066-8.

Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula, S., Krupalnik, V., *et al*. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. Cell Stem Cell 23, 412-425. 10.1016/j.stem.2018.07.004.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., *et al*. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 446, 758–764. 10.1038/nature05690.

Mur, P., García-Mulero, S., Del Valle, J., Magraner-Pardo, L., Vidal, A., Pineda, M., Cinnirella, G., Martín-Ramos, E., Pons, T., López-Doriga, A., *et al*. (2020). Role of POLE and POLD1 in familial cancer. Genet

Med 22, 2089–2100. 10.1038/s41436-020-0922-2.

Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. Genes Dev 12, 2403–2412. 10.1101/gad.12.15.2403.

Ng, S.Y., Yoshida, T., Zhang, J., Georgopoulos, K. (2009). Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity 30, 493–507. 10.1016/j.immuni.2009.01.014.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., *et al*. (2011). Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. Cell 144, 296–309. 10.1016/j.cell.2011.01.004.

Nutt, S.L., Urbánek, P., Rolink, A., Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. Genes Dev 11, 476–491. 10.1101/gad.11.4.476.

Nutt, S.L., Heavey, B., Rolink, A.G., Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature 402, 14–20. 10.1038/35005514.

O'Neill, D. W., Schoetz, S.S., Lopez, R.A., Castle, M., Rabinowitz, L., Shor, E., Krawchuk, D., Goll, M.G., Renz, M., Seelig, H.P., Han, S., *et al*. (2000). An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. Mol Cell Biol 20, 7572–7582. 10.1128/mcb.20.20.7572-7582.2000.

Orlando, E.J., Han, X., Tribouley, C., Wood, P.A., Leary, R.J., Riester, M., Levine, J.E., Qayed, M., Grupp, S.A., Boyer, M., *et al*. (2018). Genetic mechanisms of target antigen loss in CAR19 therapy of acute lymphoblastic leukemia. Nat Med 24, 1504–1506. 10.1038/s41591-018-0146-z.

Park, M., Koh, K.N., Kim, B.E., Im, H.J., Jang, S., Park, C.J., Chi, H.S., Seo, J.J. (2011). Lineage Switch at Relapse of Childhood Acute Leukemia: A Report of Four Cases. J Korean Med Sci 26, 829–831. 10.3346/jkms.2011.26.6.829.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419. 10.1038/nmeth.4197.

Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., Hovi, L., LeBlanc, T., Szczepanski, T., Ferster, A., *et al*. (2007). A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. Lancet 370, 240–250. 10.1016/S0140-6736(07)61126-X.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Research 41, e201–e201. 10.1093/nar/gkt850.

Piper, J., Assi, S.A., Cauchy, P., Ladroue, C., Cockerill, P.N., Bonifer, C., Ott, S. (2015). Wellington-bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors. BMC Genomics 16, 1000. 10.1186/s12864-015-2081-4.

Pongubala, J.M., Northrup, D.L., Lancki, D.W., Medina, K.L., Treiber, T., Bertolino, E., Thomas, M., Grosschedl, R., Allman, D., Singh, H. (2008). Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. Nat Immunol 9, 203–215. 10.1038/ni1555.

Ptasinska, A., Assi, S.A., Martinez-Soria, N., Imperato, M.R., Piper, J., Cauchy, P., Pickin, A., James, S.R., Hoogenkamp, M., Willimason, D. *et al*. (2014). Identification of a Dynamic Core Transcriptional Network in t(8;21) AML that Regulates Differentiation Block and Self-Renewal. Cell Rep 8, 1974–1988. 10.1016/j.celrep.2014.08.024 .

Rabilloud, T.,Potier, D., Pankaew, S., Nozais, M., Loosveld, M., Payet-Bornet, D. (2021). Single-cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse after CAR-T therapy. Nature Commun 12, 865. 10.1038/s41467-021-21168-6.

Raczy, C., Petrowski, R., Saunders, C.T., Chorny, I., Kruglyak, S., Margulies, E.H., Chuang, H.-Y., Källberg, M., Kumar, S.A., Liao, A . (2013). Isaac: Ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 29, 2041–2043. 10.1093/bioinformatics/btt314.

Rayes, A., McMasters, R.L., O'Brien, M.M. (2016). Lineage switch in MLL-rearranged infant leukemia following CD19-directed therapy. Pediatr Blood Cancer 63, 1113-510. 10.1002/pbc.25953

Rehe, K. Wilson, K., Bomken, S., Williamson, D., Irving, J., den Boer, M.L., Stanulla, M., Schrappe, M., Hall, A.G., Heidenreich, O., Vormoor, J.(2013). Acute B lymphoblastic leukaemia-propagating cells are present at high frequency in diverse lymphoblast populations. EMBO Mol Med 5, 38–51. 10.1002/emmm.201201703.

Risso, D. Ngai, J., Speed, T.P., Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples.Nature Biotechnol 32, 896–902. 10.1038/nbt.2931.

Rongvaux, A. Willinger, T., Martinek, J., Strowig, T., Gearty, S.V., Teichmann, L.L., Saito, Y., Marches, F., Halene, S., Palucka, A.K. *et al*. (2014). Development and function of human innate immune cells in a humanized mouse model.Nature Biotechnol 32, 364–372. 10.1038/nbt.2858.

Rossi, J.G., Andrea R Bernasconi, Cristina N Alonso, Patricia L Rubio, Marta S Gallego, Carolina A Carrara, Myriam R Guitter, Silvia Eandi Eberle, Mariela Cocce, Pedro A Zubizarreta, *et al*. (2012). Lineage switch in childhood acute leukemia: An unusual event with poor outcome. Am J Hematol 87, 890–897. 10.1002/ajh.23266.

Saldanha, A. J. (2004). Java Treeview - Extensible visualization of microarray data. Bioinformatics 20, 3246–3248.10.1093/bioinformatics/bth349.

Saunders, C.T. Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., Cheetham, R.K. (2012). Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28, 1811–1817. 10.1093/bioinformatics/bts271.

Schwalbe, E.C., Williamson, D., Lindsey, J.C., Hamilton, D., Ryan, S.L., Megahed, H., Garami, M., Hauser, P., Dembowska-Baginska, B., Perek, D. *et al*. (2013). DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. Acta Neuropathol 125, 359–371. 10.1007/s00401-012-1077-2.

Schwarzer, A. Emmrich, S., Schmidt, F., Beck, D., Ng, M., Reimer, C., Adams, F.F., Grasedieck, S., Witte, D., Käbler, S. *et al*. (2017). The non-coding RNA landscape of human hematopoiesis and leukemia. Nature Commun 8, 218.10.1038/s41467-017-00212-4.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498–2504. 10.1101/gr.1239303.

Sifrim, A. Hitz, M.-P., Wilsdon, A., Breckpot, J., Al Turki, S.H., Thienpont, B., McRae, J., Fitzgerald, T.W.F., Singh, T., Swaminathan, G.J. *et al*. (2016). Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nature Gen 48, 1060–1065. 10.1038/ng.3627.

Somervaille, T.C.P., Christina J Matheny, C.J., Gary J Spencer, G.J., Masayuki Iwasaki, M., John L Rinn, J.L., Daniela M Witten, D.M., Howard Y Chang, H.Y., Sheila A Shurtleff, S.A., James R Downing, J.R., Michael L Cleary, M.L. (2009). Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. Cell Stem Cell 4, 129–140. 10.1016/j.stem.2008.11.015.

Soneson, C., Love, M.I., Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000 Research 4, 1521. 10.12688/f1000research.7563.

Sotillo, E., Barrett, D.M., Black, K.L., Bagashev, A., Oldridge, D., Wu, G., Sussman, R., Lanauze, C., Ruella, M., Gazzara, M.R., *et al*. (2015). Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. Cancer Discovery 5, 1282–1295. 10.1158/2159-8290.CD-15-1020.

Sperlazza, J., Rahmani, M., Beckta, J., Aust, M., Hawkins, E., Wang, S., Zhu, S., Podder, S., Dumur, C., Archer, K., *et al*. (2015). Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation. Blood 126, 1462–1472. 10.1182/blood-2015-03-631606.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550. 10.1073/pnas.0506580102

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., *et al*. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research 47, D607–D613. 10.1093/nar/gky1131.

Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W.M., Mesirov, J.P. (2007). Metagene projection for cross-platform, cross-species characterization of global transcriptional states. . Proc Natl Acad Sci U S A 104, 5959 LP – 5964. 10.1073/pnas.0701068104.

van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., *et al*. (2013). From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11.10.1-11.10.33. 10.1002/0471250953.bi1110s43.

van Dongen, J.J.M., Langerak, A.W., Brüggemann, M., Evans, P.A.S., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuuring, E., García-Sanz, R. *et al*. (2003). Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17, 2257–2317. 10.1038/sj.leu.2403202.

Wang, J., Mi, J.-Q., Debernardi, A., Vitte, A.-L., Emadali, A., Meyer, J.A., Charmpi, K., Ycart, B., Callanan, M.B., Carroll, W.L., *et al*. (2015). A six gene expression signature defines aggressive subtypes and predicts outcome in childhood and adult acute lymphoblastic leukemia. Oncotarget 6, 16527–16542. 10.18632/oncotarget.4113.

1437 Wilkinson, A.C., Ballabio, E., Geng, H., North, P., Tapia, M., Kerry, J., Biswas, D., Roeder, R.G., Allis, C.D.,
1438 Melnick, A., *et al*. (2013). RUNX1is a key target in t(4;11) leukemias that contributes to gene activation
1439 through an AF4-MLL complex interaction. Cell Reports 3, 116–127. 10.1016/j.celrep.2012.12.016.

1440 Williams, C.J., Naito, T., Arco, P.G., Seavitt, J.R., Cashman, S.M., De Souza, B., Qi, X., Keables, P., Von
1441 Andrian, U.H., Georgopoulos, K. (2004). The chromatin remodeler Mi-2β is required for CD4
1442 expression and T cell development. Immunity 20, 719–733. 10.1016/j.immuni.2004.05.005.

1443 Witkowski, M.T., Hu, Y., Roberts, K.G., Boer, J.M., McKenzie, M.D., Liu, G.J., Le Grice, O.D., Tremblay,
1444 C.S., Ghisi, M., Willson, T.A., *et al*. (2017). Conserved IKAROS-regulated genes associated with B-
1445 progenitor acute lymphoblastic leukemia outcome. J Exp Med 214, 773–791. 10.1084/jem.20160048.

1446 Yoshida, T., Ng, S.Y.-M., Zuniga-Pflucker, J.C., and Georgopoulos, K. (2006). Early hematopoietic
1447 lineage restrictions directed by Ikaros. Nature Immunology 7, 382–391. 10.1038/ni1314.

1448 Yoshida, T., Hazan, I., Zhang, J., Ng, S.Y., Naito, T., Snippert, H.J., Heller, E.J., Qi, X., Lawton, L.N.,
1449 Williams, C.J., *et al*. (2008).The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell
1450 self-renewal and multilineage differentiation. Genes Dev 22, 1174–1189. 10.1101/gad.1642808.

1451 Zangrando, A., Dell'orto, M.C., Te Kronnie, G., Basso, G. (2009). MLL rearrangements in pediatric acute
1452 lymphoblastic and myeloblastic leukemias: MLL specific and lineage specific signatures. BMC Med
1453 Genomics 236, 10.1186/1755-8794-2-36.

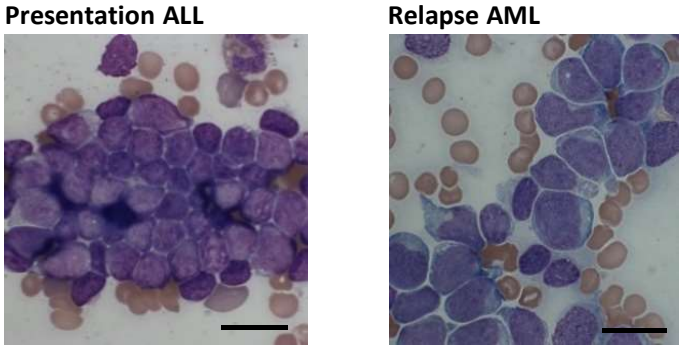
1454 Zhang, J., Jackson, A.F., Naito, T., Dose, M., Seavitt, J., Liu, F., Heller, E.J., Kashiwagi, M., Yoshida, T.,
1455 Gounari, F., et al. (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls
1456 lymphocyte development and prevents leukemogenesis. Nature Immunology 13, 86–94.
1457 10.1038/ni.2150.

1458 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M.,
1459 Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137.
1460 10.1186/gb-2008-9-9-r137.

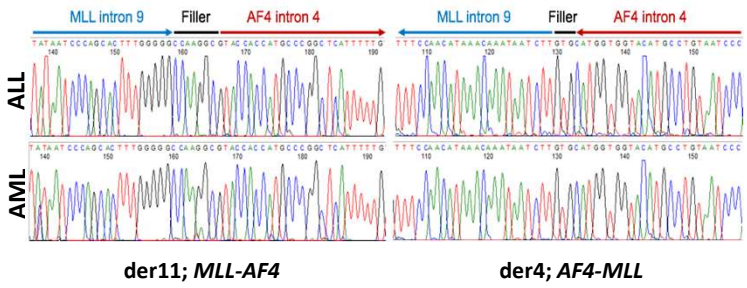
1461
1462

Figure 1

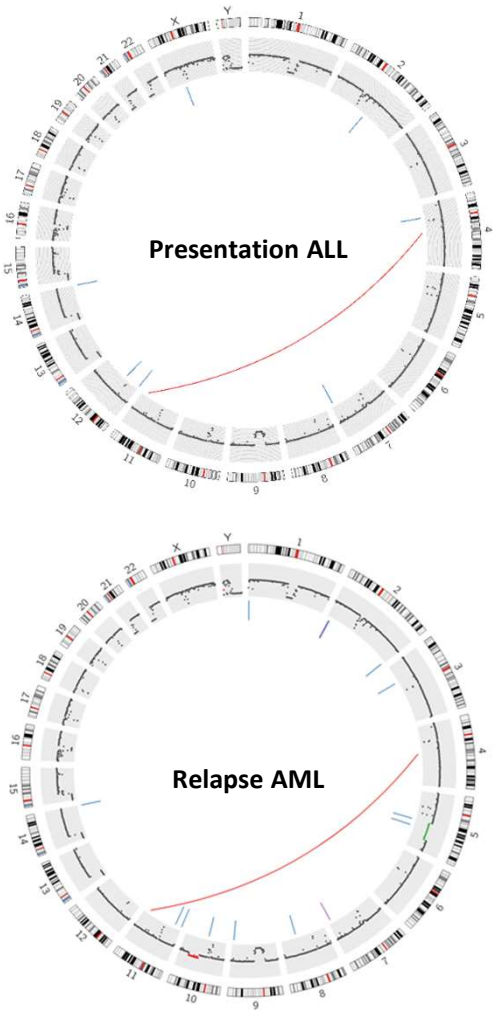
A



B



C



D

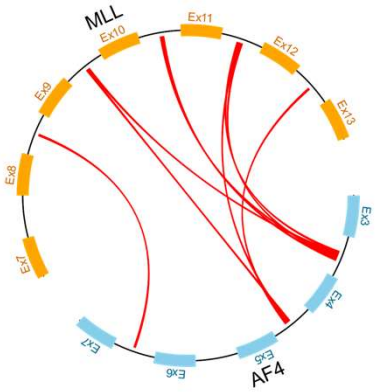
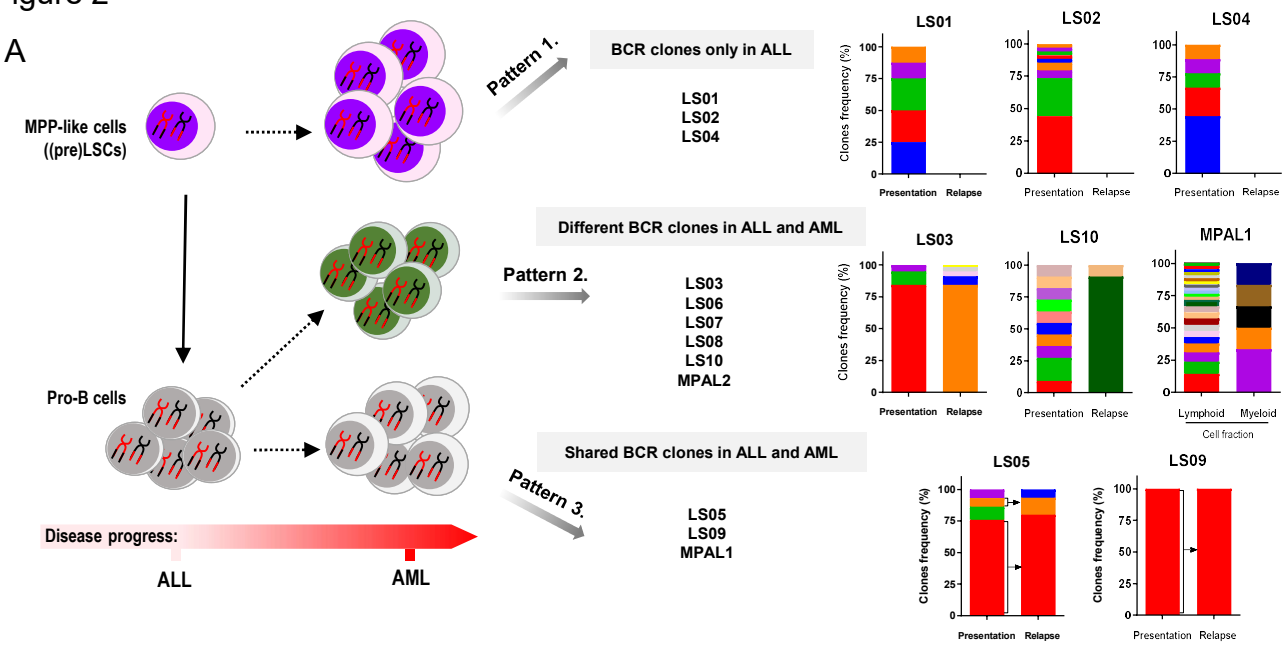
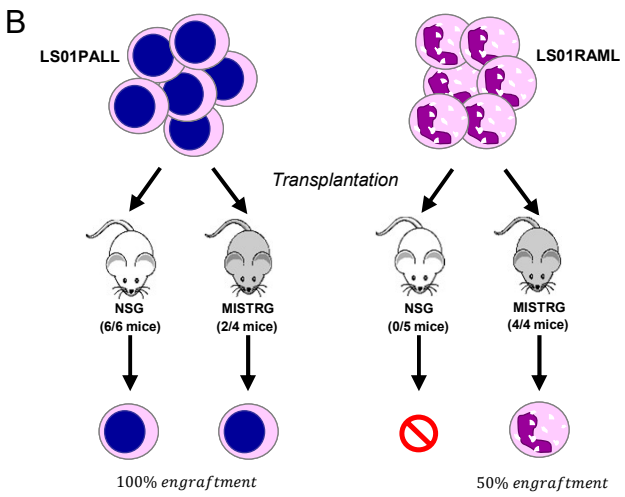


Figure 2





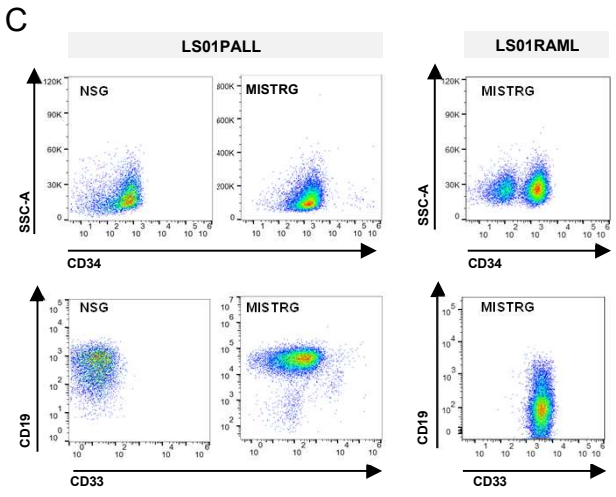


Figure 3

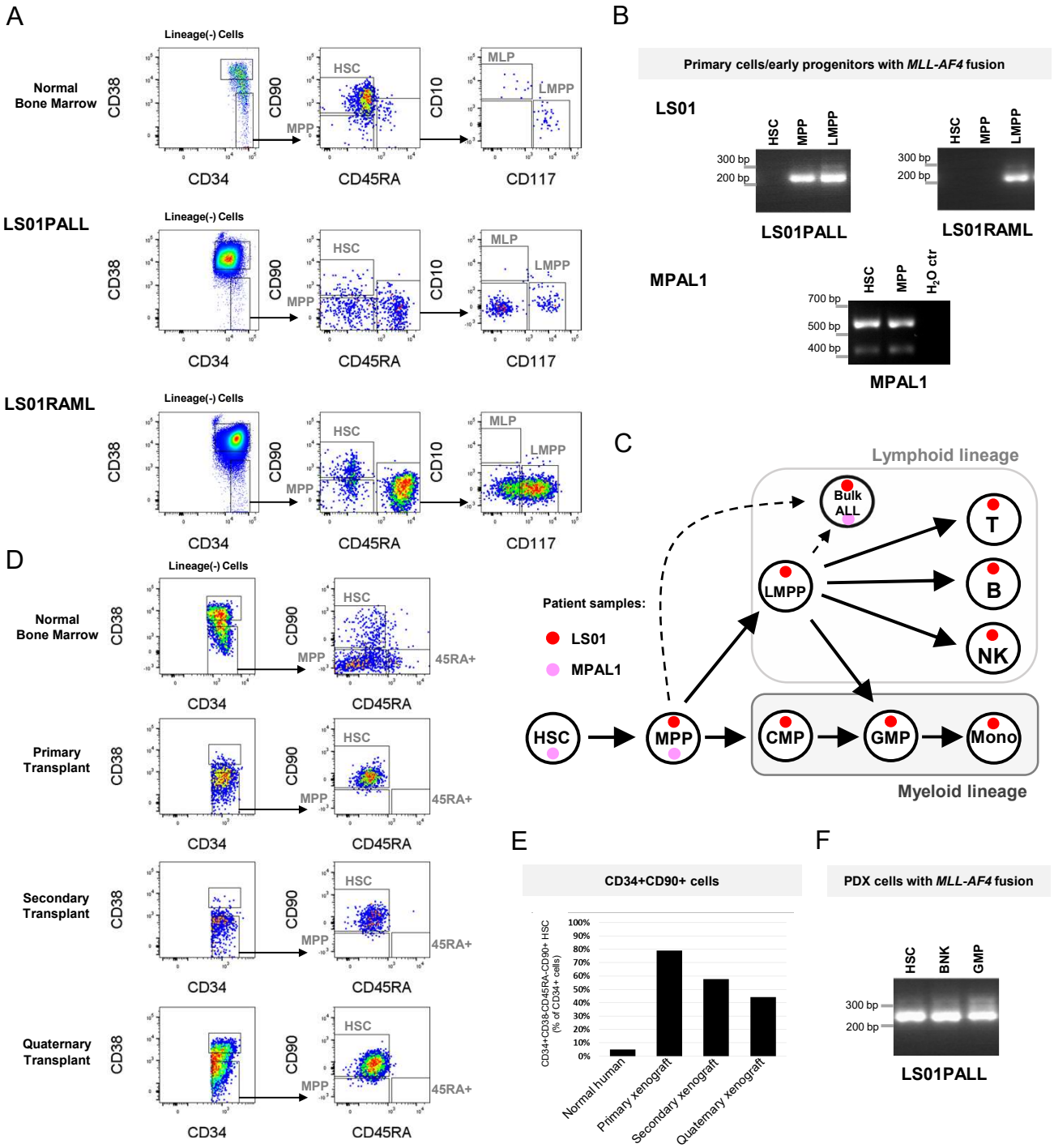


Figure 4

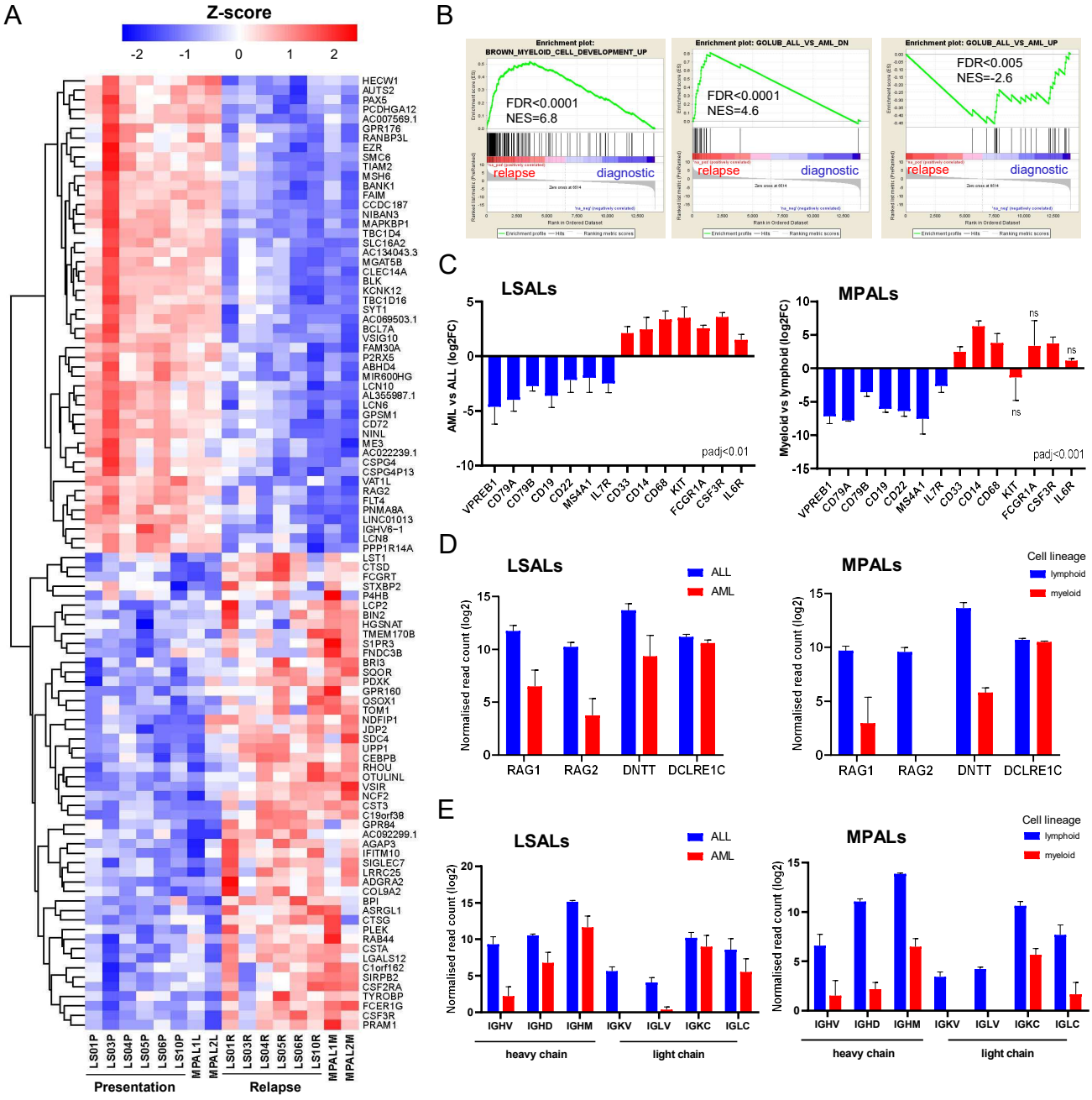


Figure 5

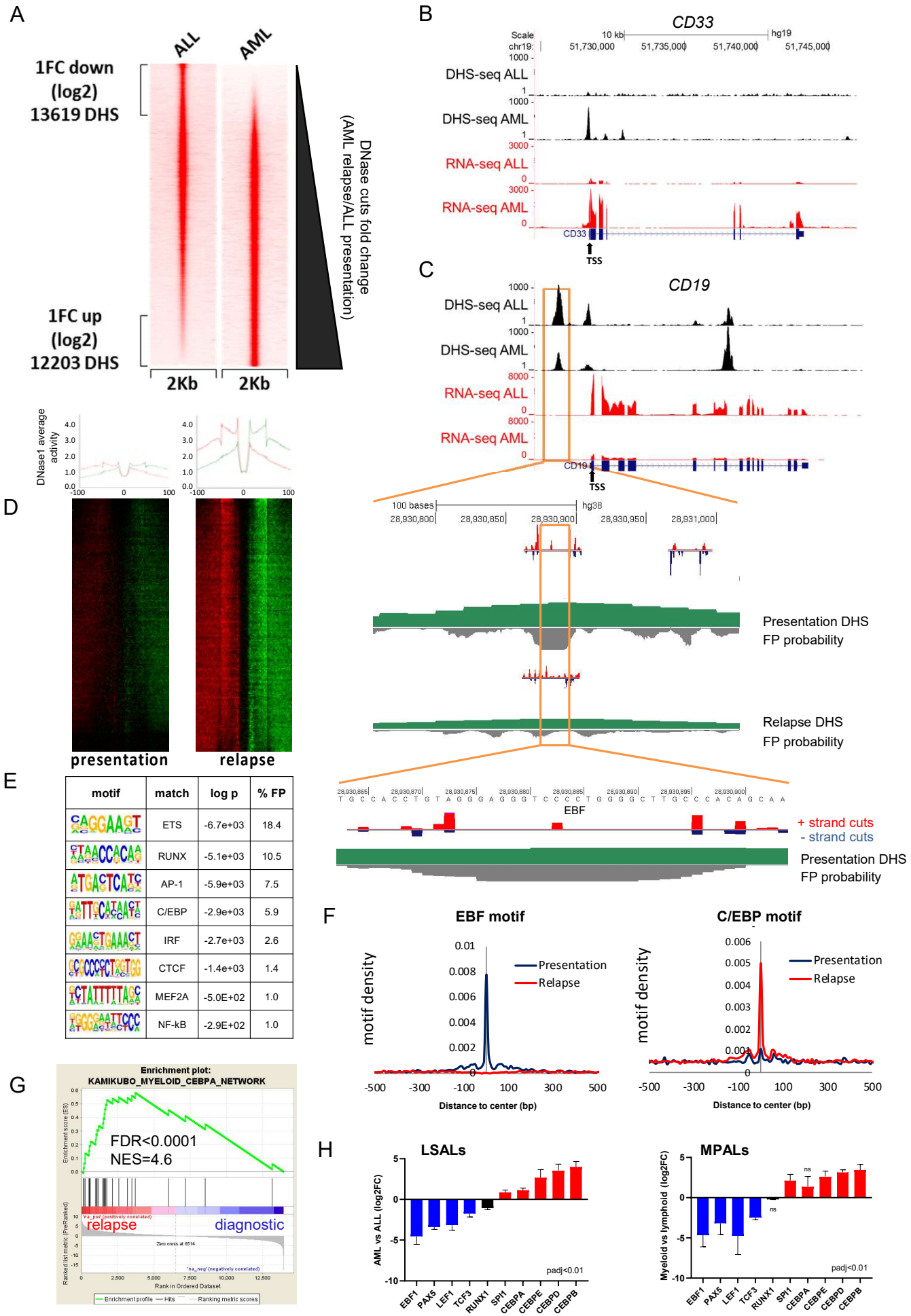


Figure 6

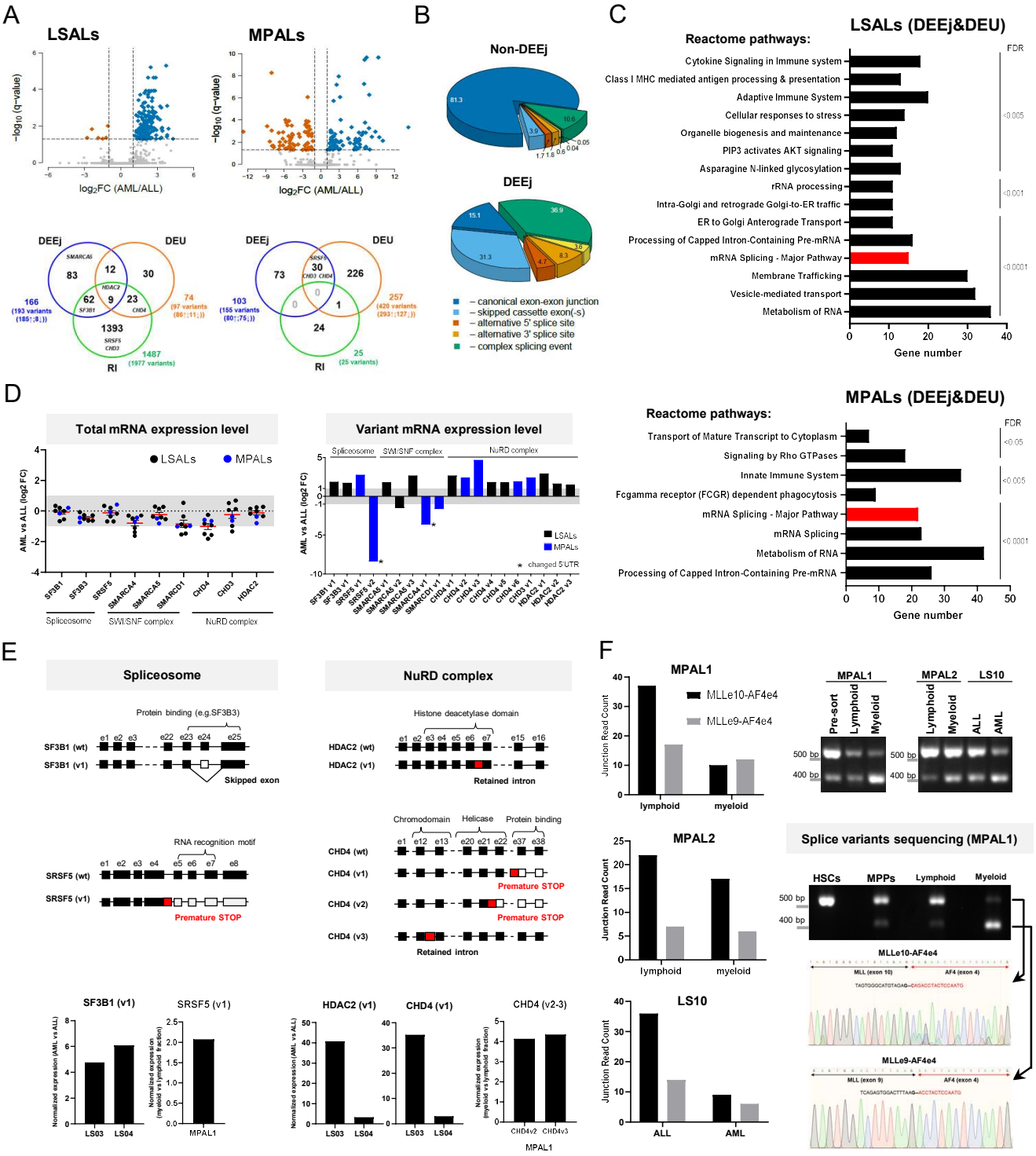


Figure 7

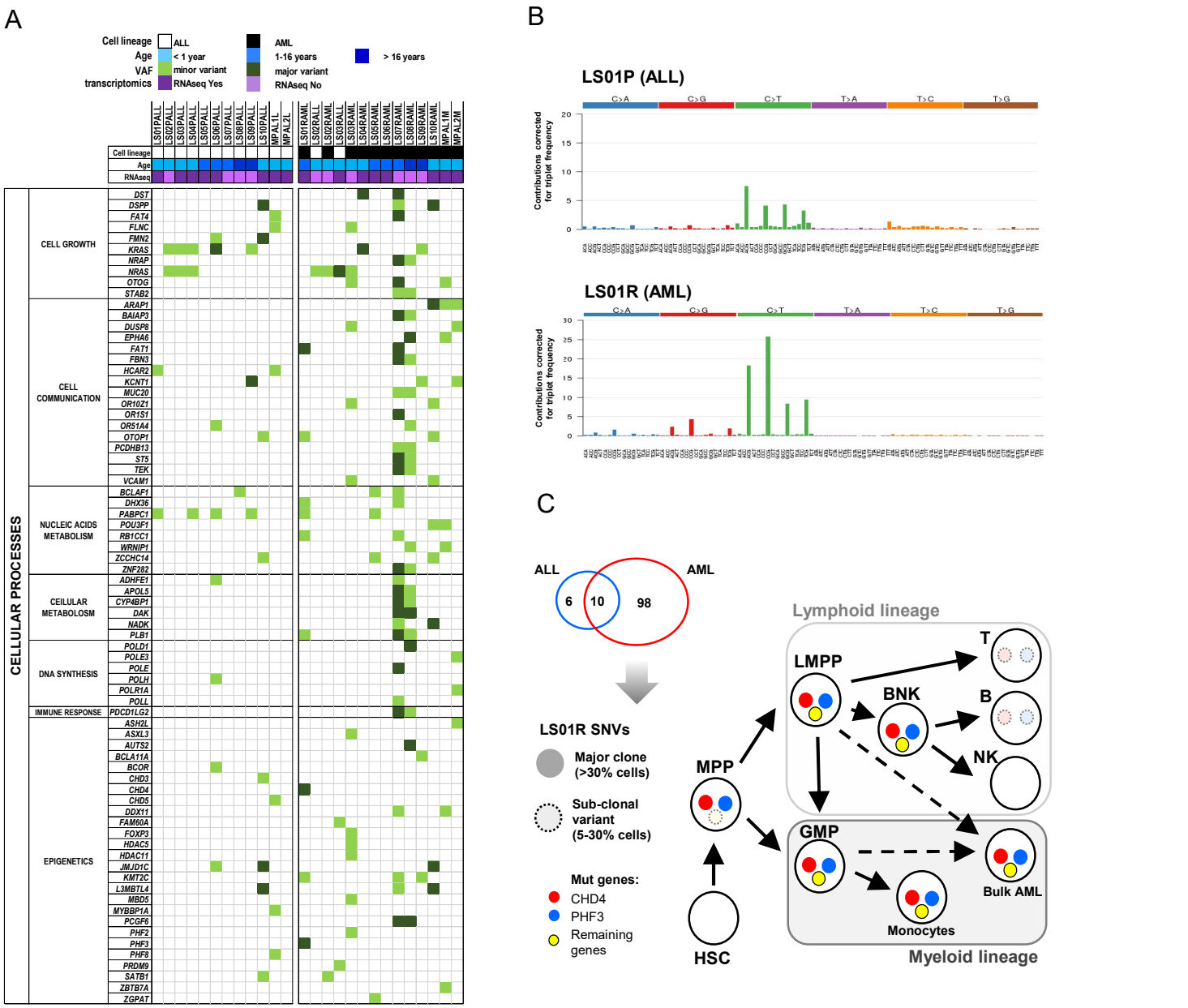


Figure 8

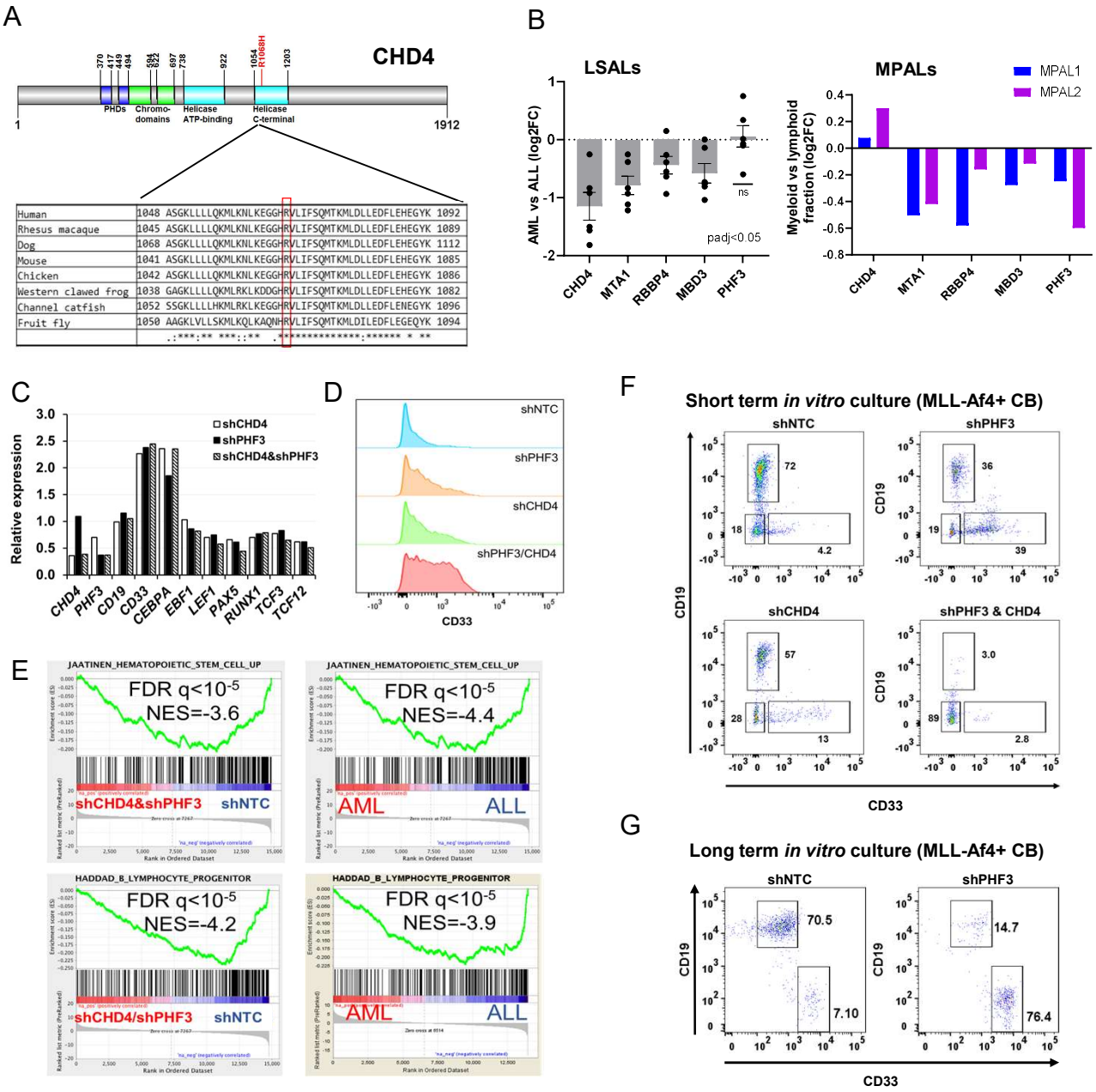
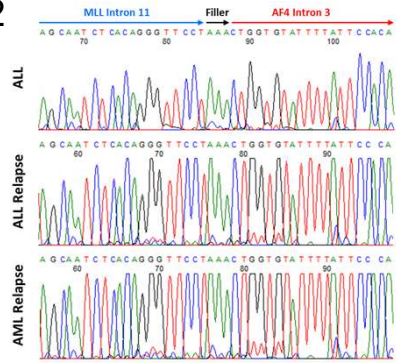
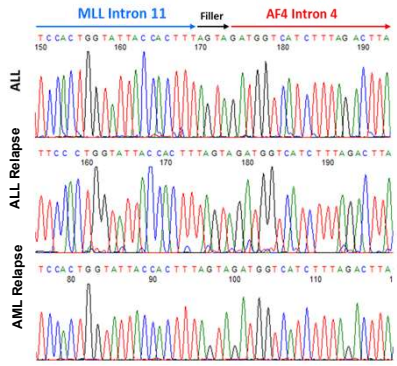


Figure S1

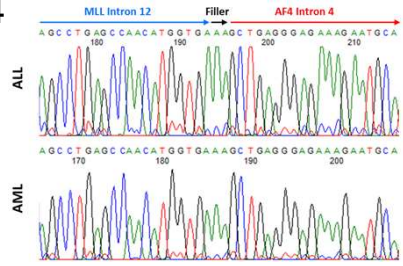
LS02



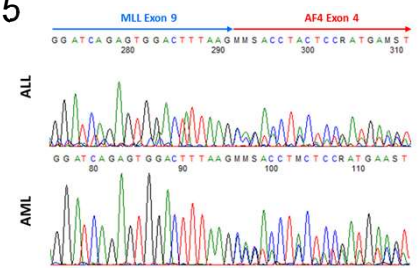
LS03



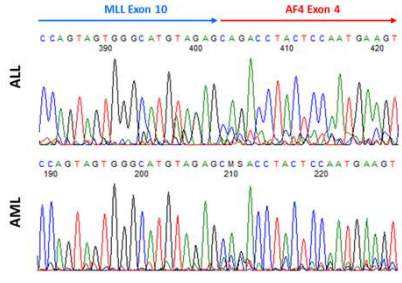
LS04



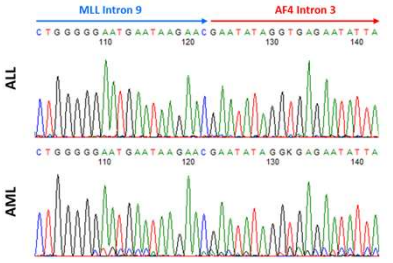
LS05



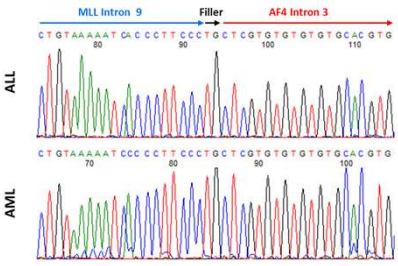
LS06



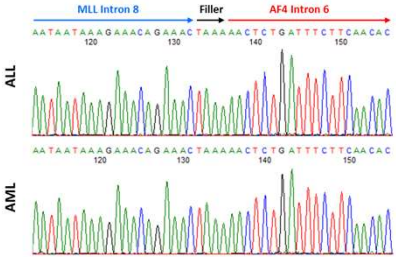
LS07



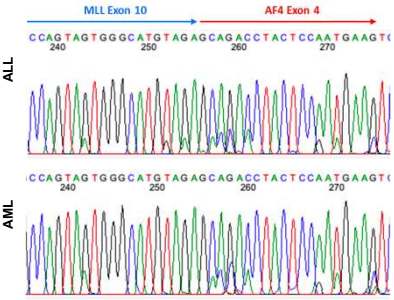
LS08



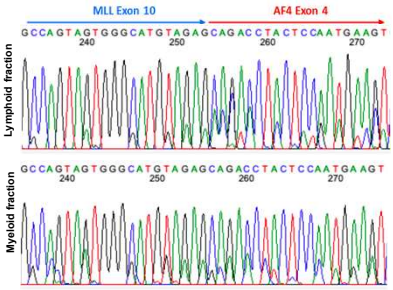
LS09



LS10



MPAL1



MPAL2

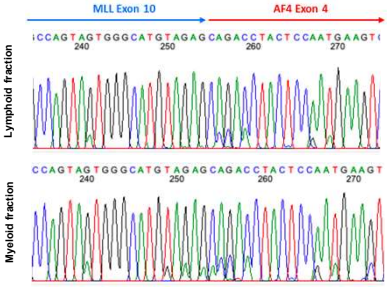


Figure S2

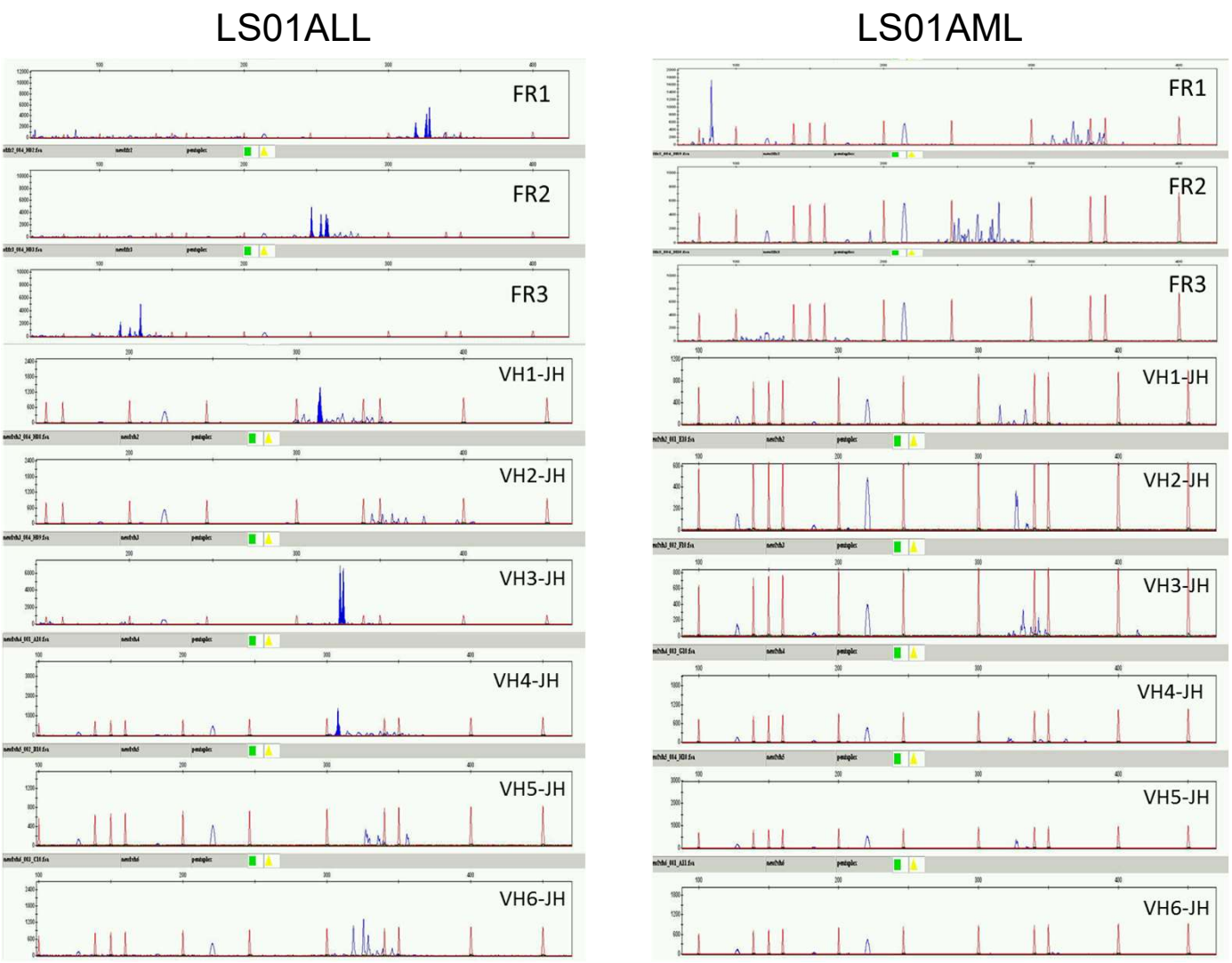


Figure S3

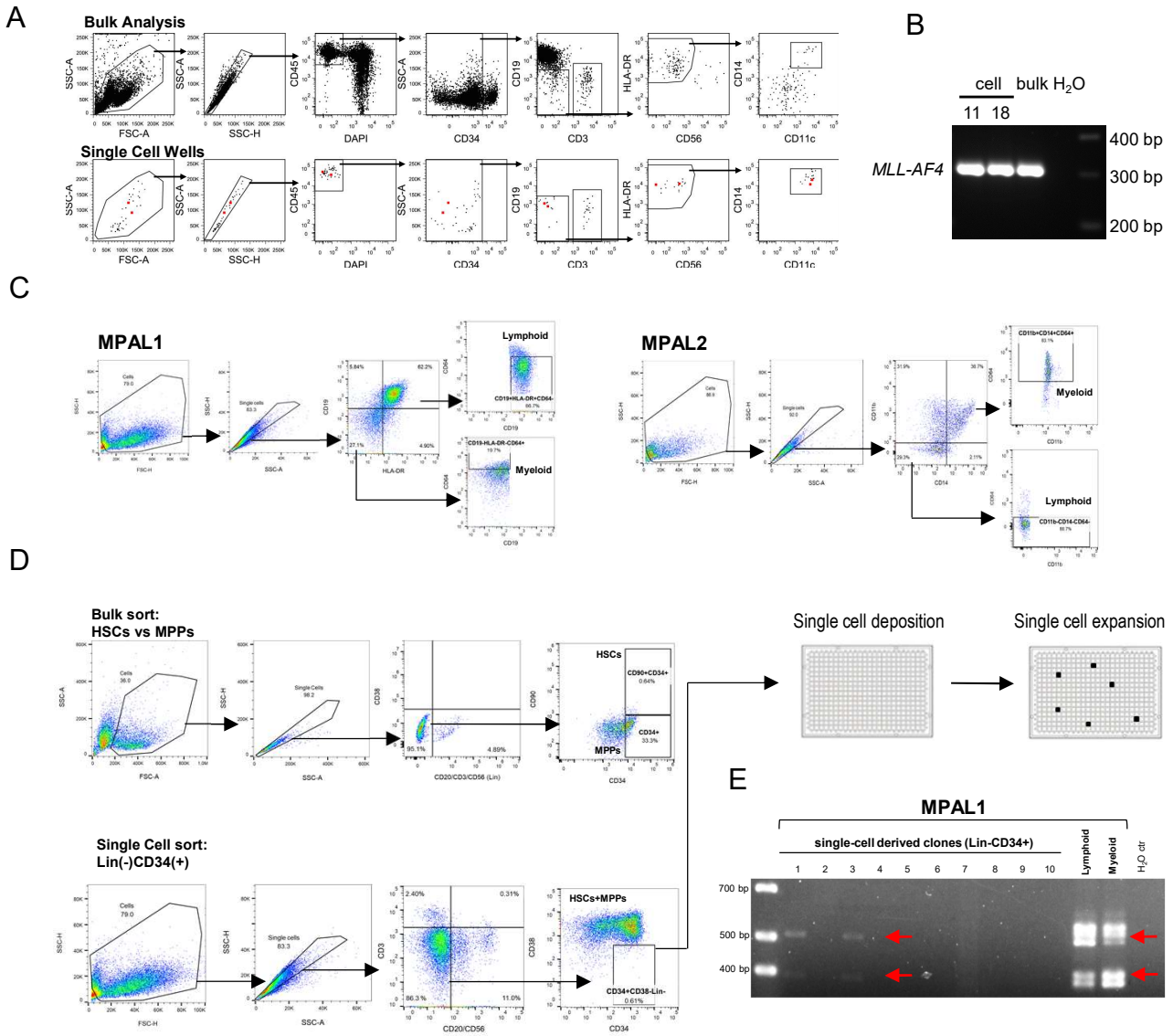


Figure S4

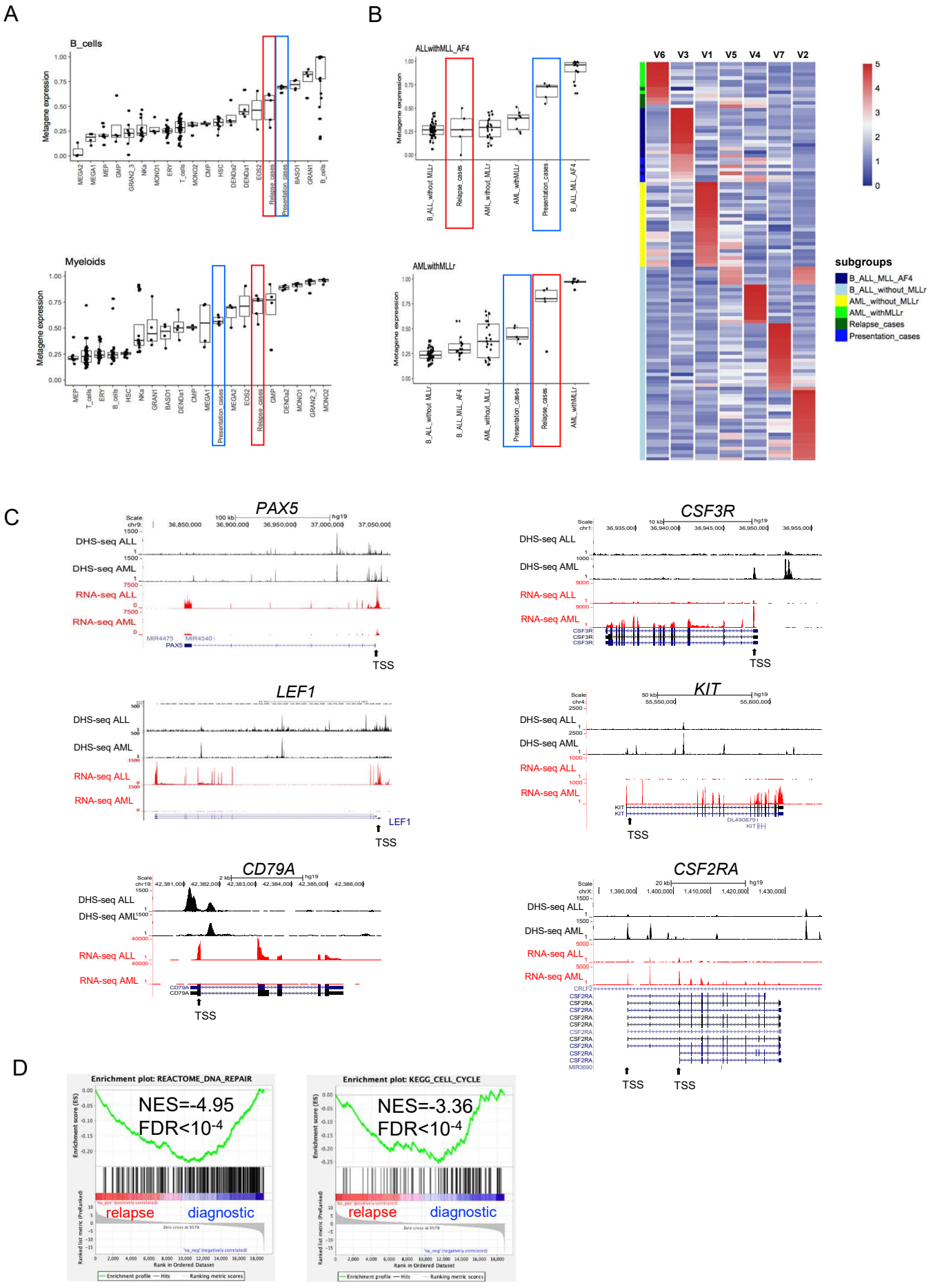


Figure S5

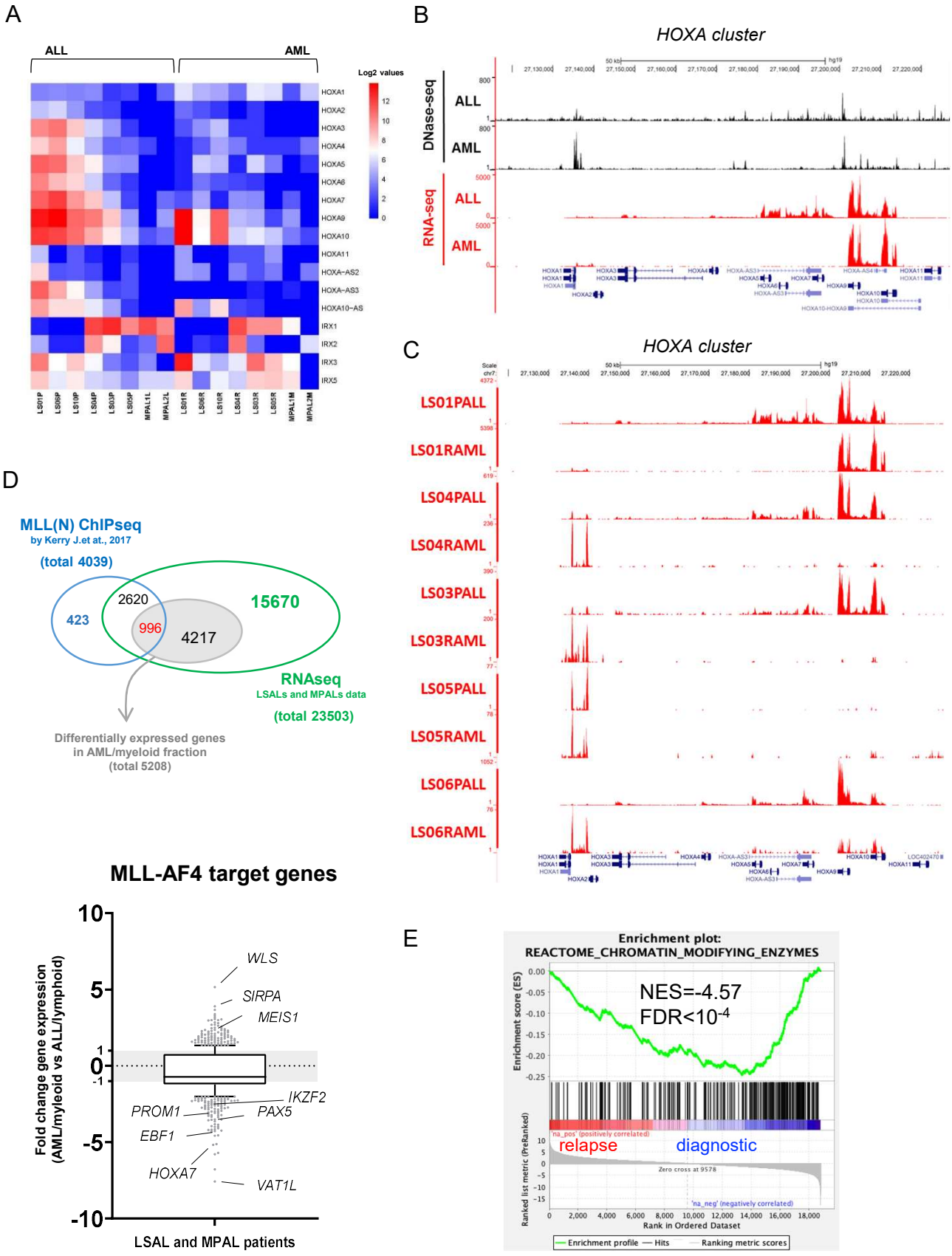
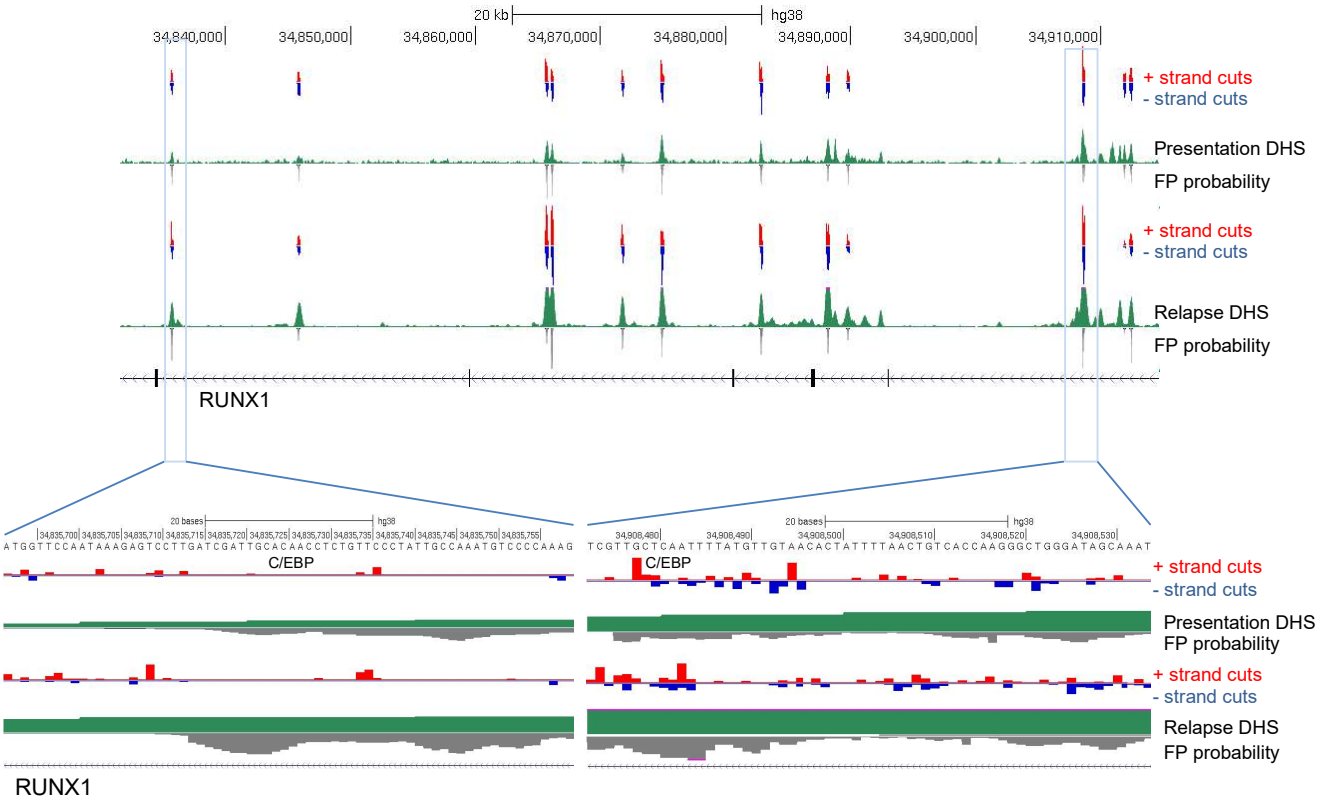
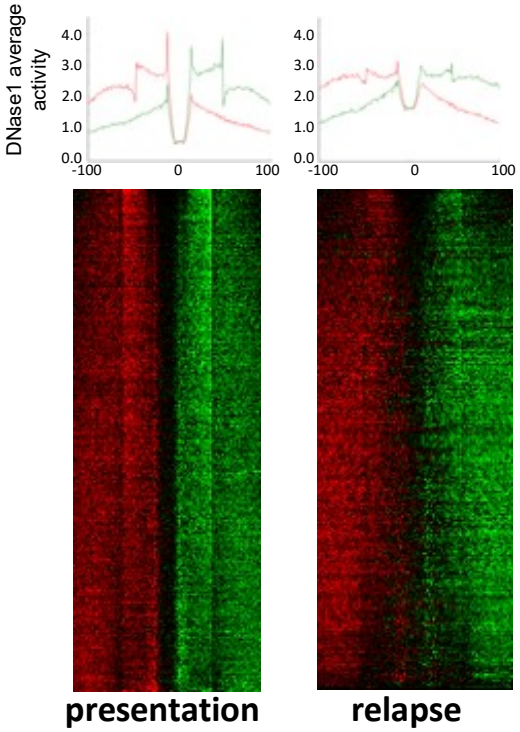


Figure S6

A



B



motif	match	p-value	% FP
	ETS	1e-516	13.6
	RUNX	1e-378	10.8
	EBF	1e-984	10.30
	Ebox	1e-113	9.9
	AP1	1e-640	6.7
	Ebox	1e-71	4.9
	PAX	1e-146	1.6

Figure S7

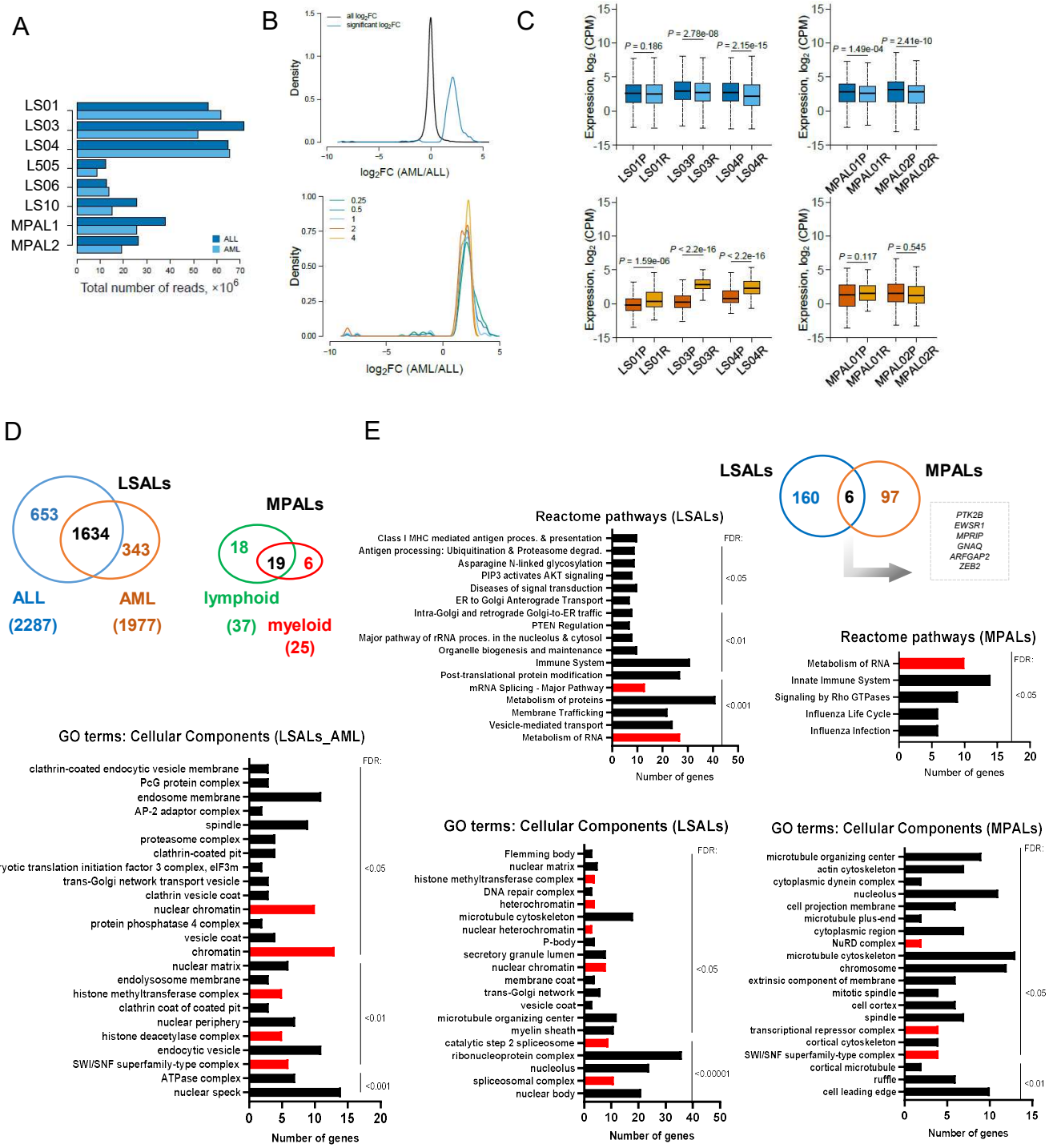


Figure S8