Northumbria University
NEWCASTLE

University**Library**

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

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

Tracking no: BLD-2021-015036R2

Ricky Tirtakusuma (Newcastle University, United Kingdom) Katarzyna Szoltysek (Newcastle University, United Kingdom) Paul Milne (Newcastle University, United Kingdom) Vasily Grinev (Belarusian State University, Belarus) Anetta Ptasinska (University of Birmingham, United Kingdom) Paulynn Chin (University of Birmingham, United Kingdom) Claus Meyer (DCAL, Institute of Pharmaceutical Biology, Goethe-University Frankfurt, Germany) Sirintra Nakjang (Newcastle University, United Kingdom) Jayne Hehir-Kwa (Princess Maxima Centrum for Pediatric Oncology, Netherlands) Daniel Williamson (Newcastle University, United Kingdom) Pierre Cauchy (University of Birmingham, United Kingdom) Peter Keane (University of Birmingham, United Kingdom) salam assi (University of Birmingham, United Kingdom) Minoo Ashtiani (Princess Maxima Centrum for Pediatric Oncology, Netherlands) Sophie Kellaway (University of Birmingham, United Kingdom) Maria Imperato (University of Birmingham, United Kingdom) Fotini Vogiatzi (University Hospital Schleswig-Holstein, Campus Kiel, Germany) Elizabeth Schweighart-James (Princess Maxima Centrum for Pediatric Oncology, Netherlands) Shan Lin (Cincinnati Children's Hospital Medical Center, United States) Mark Wunderlich (Cincinnati Children's Hospital Medical Center, United States) Janine Stutterheim (Princess Maxima Center, Netherlands) Alexander Komkov (Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Federation) Elena Zerkalenkova (Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Russian Federation) Paul Evans (Leeds Teaching Hospitals National Health Services Trust,) hesta mcneill (Newcastle University, United Kingdom) Alex Elder (Newcastle University, United Kingdom) Natalia Martínez-Soria (Newcastle University, United Kingdom) Sarah Fordham (Newcastle University, United Kingdom) Yuzhe Shi (Northern Institute for Cancer Research, United Kingdom) Lisa Russell (Newcastle University, United Kingdom) Deepali Pal (Northumbria university, United Kingdom) Alexandra Smith (University of York, United Kingdom) zoya kingsbury (Illumina Cambridge Ltd., United Kingdom) Jennifer Becq (Illumina UK, United Kingdom) Cornelia Eckert (Charite Medical Center, Campus Virchow-Klinikum, Germany) Oskar Haas (CCRI & St. Anna Children's Hospital, Austria) peter carey (Department of Haematology, Royal Victoria Infirmary,) Simon Bailey (Newcastle University, United Kingdom) Roderick Skinner (Newcastle University, United Kingdom) Natalia Miakova (Federal Research Centre of Pediatric Hematology, Oncology and Immunology, Russian Federation) Matthew Collin (Newcastle University, United Kingdom) Venetia Bigley (Newcastle University, United Kingdom) Muzlifah Haniffa (Newcastle University, United Kingdom) Rolf Marschalek (Goethe-University of Frankfurt/Main, Germany) Christine Harrison (Newcastle University, United Kingdom) Catherine Cargo (Haematological Malignancy Diagnostic Service, United Kingdom) Denis Schewe (Otto-von-Guericke University Magdeburg, Germany) Yulia Olshanskaya (Dmitriy Rogachev National Medical Center of pediatric hematology, oncology and immunology, Russian Federation) Michael Thirman (University of Chicago, United States) Peter Cockerill (Institute of Cancer and Genomic Sciences, United Kingdom) James Mulloy (Cincinnati Children's Hospital Medical Center, United States) Helen Blair (Newcastle University,) H. Vormoor (Newcastle University, United Kingdom) James Allan (Newcastle University, United Kingdom) Constanze Bonifer (University of Birmingham, United Kingdom) Olaf Heidenreich (Newcastle University, United Kingdom) Simon Bomken (Newcastle University, United Kingdom)

Abstract:
The fusion gene *MLL/AF4* defines a high-risk subtype of pro-B acute lymphoblastic leukaemia. Relapse can be associated with a lineage switch from acute lymphoblastic to acute myeloid leukaemia resulting in poor clinical outcomes due to resistance towards chemo- and immuno-therapies. Here we show that the myeloid relapses share oncogene fusion breakpoints with their matched lymphoid presentations and can originate from varying differentiation stages from immature progenitors through to committed B-cell precursors. Lineage switching is linked to substantial changes in chromatin accessibility and rewiring of transcriptional programmes, including alternative splicing. These findings indicate that the execution and maintenance of lymphoid lineage differentiation is impaired. The relapsed myeloid phenotype is recurrently associated with the altered expression, splicing or mutation of chromatin modifiers, including *CHD4* coding for the ATPase/helicase of the nucleosome remodelling and deacetylation complex, NuRD. Perturbation of *CHD4* alone or in combination with other mutated epigenetic modifiers induces myeloid gene expression in *MLL/AF4*-positive cell models indicating that lineage switching in *MLL/AF4* leukaemia is driven and maintained by disrupted epigenetic regulation.

Conflict of interest: COI declared - see note

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

Preprint server: Yes; biorxiv https://doi.org/10.1101/2021.07.16.452676

Author contributions and disclosures: Conceptualization, O.H., S.B., C.B.; Methodology, O.H., C.B., R.T., K.S., P.M., S.B., A.P., 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., V.V.G., A.K., D.W., P.C.; Formal Analysis, S.N., M.A., J.H.K., V.V.G., A.K., D.W., P.C., P.K., C.B., O.H.; Investigation, R.T., K.S., P.M., A.P., C.M., P.S.C., H.J.B., S.G.K.., A.K., S.A., M.R.I., E.K.S., P.E., H.M., A.E., N.M.S., 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., 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.; Data Curation, S.N., D.W., P.C.; Writing, S.B., O.H., C.B., R.T., K.S.; Supervision, O.H., S.B., 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.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: 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. Requests for additional specific data/materials should be made to Olaf Heidenreich (O.T.Heidenreich@prinsesmaximacentrum.nl).

Clinical trial registration information (if any):

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], Paulynn S Chin[6], Claus Meyer[7], Sirintra Nakjang[1], Jayne Y Hehir-Kwa[2], Daniel Williamson[1], Pierre Cauchy[6], Peter Keane[6], Salam A Assi[6], Minoo Ashtiani[2], Sophie G Kellaway[6], Maria R Imperato[6], Fotini Vogiatzi[9], Elizabeth K Schweighart[2], 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], Muzlifah 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, 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, Newcastle upon Tyne, UK

[5]Department of Genetics, the Faculty of Biology, Belarusian State University, 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]St. Anna Children´s Cancer Research Institute (CCRI), Vienna, Austria

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

[17]Biosciences Institute, Newcastle University, 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

***Authors contributed equally**
****Co-senior authors**
†Co-corresponding Authors.

Running title: Lineage switching in *MLL/AF4* leukaemias

Keywords: *MLL/AF4*; *KMT2A*/*AFF1*; acute lymphoblastic leukaemia; nucleosome remodelling and deactylation complex (NuRD); chromatin remodelling

Corresponding authors

Dr Simon Bomken
Wolfson Childhood Cancer Research Centre
Translational and Clinical Research Institute
Level 6 Herschel Building
Brewery Lane
Newcastle University
Newcastle upon Tyne
NE1 7RU, UK

Tel: +44 (0)191 2082231
E mail: s.n.bomken@ncl.ac.uk

Professor Olaf Heidenreich
Princess Maxima Center for Pediatric Oncology
Heidelberglaan 25
3584 CS Utrecht
The Netherlands

Tel: +31 (0)88 972 7272
E mail: o.t.heidenreich@prinsesmaximacentrum.nl

Abstract word count - 169
Manuscript word count - 4054

Figures - 7
Tables - 0
References - 62

Keypoints

- Myeloid relapse can originate from varying differentiation stages of *MLL/AF4*-positive ALL.
- Dysregulation of epigenetic regulators underpins fundamental lineage reprogramming.

Abstract

The fusion gene *MLL/AF4* defines a high-risk subtype of pro-B acute lymphoblastic leukaemia. Relapse can be associated with a lineage switch from acute lymphoblastic to acute myeloid leukaemia resulting in poor clinical outcomes due to resistance towards chemo- and immuno-therapies. Here we show that the myeloid relapses share oncogene fusion breakpoints with their matched lymphoid presentations and can originate from varying differentiation stages from immature progenitors through to committed B-cell precursors. Lineage switching is linked to substantial changes in chromatin accessibility and rewiring of transcriptional programmes, including alternative splicing. These findings indicate that the execution and maintenance of lymphoid lineage differentiation is impaired. The relapsed myeloid phenotype is recurrently associated with the altered expression, splicing or mutation of chromatin modifiers, including *CHD4* coding for the ATPase/helicase of the nucleosome remodelling and deacetylation complex, NuRD. Perturbation of *CHD4* alone or in combination with other mutated epigenetic modifiers induces myeloid gene expression in *MLL*/*AF4*-positive cell models indicating that lineage switching in *MLL/AF4* leukaemia is driven and maintained by disrupted 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 (AML) and lymphoblastic leukaemias (ALL) and is generally associated with poor prognosis[1,2]. 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[1]. 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[3-5]. 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[6].

Lineage plasticity has been associated with the loss of original therapeutic targets [7,8]. 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. These findings provide 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.

Methods

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.

Additional methods are described in Supplemental Methods.

158

Results

Characterisation of MLL/AF4 acute leukaemias with lineage switch

We focussed on lineage switches which originally presented as ALL and relapsed as AML, and mixed phenotype acute leukaemias (MPALs) presenting with distinct lymphoid and myeloid populations. Lymphoid and myeloid phenotypes were defined by morphology and by expression of either B lymphoid (CD19, CD22, CD79A) or myeloid antigens (CD33, CD117/KIT, CD64/FCGR1A) (Figure 1A, Table S1). To exclude *de novo* and therapy-associated AMLs, which are unrelated to the original ALL and do not share the initiating event, the lymphoid and myeloid presentations and relapses had to display identical *MLL*/*AF4* breakpoints as genetic proof of relationship (Figures 1B,S1, Table S1). Using these definitions, we collected a cohort of 12 cases of *MLL/AF4* ALL comprising 6 infant, 2 paediatric and 2 adult patients who relapsed with acute myeloid leukaemia (AML), including one infant patient (LS10) who relapsed following B-lineage directed blinatumomab treatment and two infant *MLL/AF4* mixed phenotype acute leukaemias (MPALs)(Table S1).

Lineage switch leukaemia is associated with transcriptional reprogramming

We hypothesized that lineage switch would be linked with changes in gene expression. Since the changes in transcriptome composition may include altered regulation of both transcription and mRNA maturation[9], we compared gene expression and splicing between lymphoid and myeloid populations from lineage switch and MPAL patients. Cluster analysis of differential gene expression robustly separated both population types (Figure 2A). We identified 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 linked to reduced lymphoid and increased myeloid gene expression (Figure 2B, Table S2). Changed gene expression included the loss of lymphoid genes such as *PAX5, EBF1, CD19, CD20* (*MS4A1*) *and CD22,* diminished gene expression of immunoglobulin genes and genes involved in the VDJ recombination (*RAG1, RAG2, DNTT*), and a gain of myeloid gene including *CLEC12A, PRAM1, CSF3R* and members of the *CEBP* transcription factor family (Figures

2C,D,S2A,B)[10-12]. Moreover, almost 30% of direct *bona fide* target genes of *MLL/AF4* including *PROM1,* encoding the stem cell marker CD133, *IKZF2* and *HOXA7* showed lower expression in myeloid cells despite sharing the same *MLL/AF4* isotype (Figures S3A-D, Table S2)[13-15]. These data show that lineage switch also involves differential *MLL/AF4*-driven gene expression.

The analysis of RNA isoform compositions showed that lineage switch is associated with altered splicing, comprising changes in intron retention and differential usage of exons and exon-exon linkages (Figure 3A, Tables S3,S4). Interestingly, 85% of all differentially used exon-exon linkages were non-canonical and mainly consisted of exon skipping and complex splicing events (Figures 3A,B, Table S4). Pathway analysis revealed an enrichment of alternatively spliced genes in immune pathways, including antigen processing and membrane trafficking, suggesting that alternative splicing is linked to the change from a lymphoid to a myeloid differentiation state (Figure 3C).

Interestingly, lineage switch also affected total expression and the composition of alternatively spliced fusion transcript isoforms for both *MLL/AF4* and *AF4/MLL*. For instance, we detected in relapse material from patient LS01 a fusion variant skipping *MLL* exon 9 (Figure S3E, Table S5). In addition, we also observed changes in transcription and splicing for genes regulating the chromatin landscape. Several epigenetic regulators, including the polycomb PRC1 like complex component *AUTS2 and* the SWI/SNF complex component *BCL7A* were down-regulated in myeloid compared to lymphoid cells (Figure 2A). Several other spliceosome and SWI/SNF members were either differentially expressed or spliced. Amongst all NuRD complex members, only *CHD4* demonstrated differential expression whilst *CHD4*, *CHD3* and *HDAC2* were differentially spliced in AML relapse cells or myeloid subpopulations of MPALs (Figures 3D,E, Table S4). For instance, *CHD4* encoding the ATPase/helicase subunit of the histone-modifying NuRD complex showed a significantly lower expression in AML relapses of patients with lineage switch, but was differentially

spliced in MPAL patients resulting in premature stops or intron retention most likely leading loss of function isoforms.

Reorganisation of chromatin accessibility and transcription factor binding site occupancy upon lineage switch

The substantial gene expression changes, including those affecting epigenetic regulators and lineage-determining transcription factors, prompted us to link transcriptional changes to altered genome-wide chromatin accessibility. High resolution DNaseI hypersensitive site (DHS) mapping combined with digital footprinting analysis using the Wellington algorithm[16] uncovered multiple differentially accessible genes including the hematopoietic surface marker genes CD33 and CD19 and transcription factors (Figures 4A-C,S4A,B). These alterations occurred both at locations distal and proximal to transcriptional start sites (TSS) indicating the involvement of enhancers and promoters (Figures 4D,S4C). Digital footprinting is now generally accepted to highlight factors important for regulating specific cell fates[17-19]. These analyses showed that changes in chromatin accessibility after lineage switch were linked to an altered pattern of transcription factor binding site occupancy (Figures 4E,S4D) with a loss of occupancy of consensus binding sites for lymphoid transcription factors including EBF or PAX5 and a corresponding increased occupancy of binding motifs for myeloid factors including C/EBP family members (Figures 4E,F). We also observed a redistribution of footprinted sites for transcription factors controlling both lymphoid and myeloid maturation such as RUNX, AP-1 and ETS family members to alternative cognate motifs (Figures 4E,S4D)[20,21]. This finding is exemplified by decreased accessibility of a region located 1 kb upstream of the *CD19* TSS with concomitant loss of EBF binding site occupancy at this element (Figure 4C). In conclusion, the transition from lymphoid to myeloid immunophenotype is associated with genome-wide alterations in chromatin accessibility and transcription factor binding site occupancy.

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 agreement with previously reported mutation rates in *MLL*-rearranged leukaemias, presentation ALLs displayed a relatively quiet mutational landscape with a median of 25 nonsynonymous somatic single nucleotide variants (SNVs) or insertions/deletions (indels) (Figures S5A,B, Table S6)[10,22]. Most of them were sub-clonal with less than 30% of the reads. The group of AML relapses showed on average 92 SNVs and indels. However, this increase was due to the more heterogeneous composition of the relapse group: two cases (LS07AML and LS08AML) carried mutated DNA polymerase genes resulting in increased mutational burden. We observed this phenotype in only two out of ten relapses, arguing against this phenomenon being a general requirement for the lineage switch.

In general, we found only a limited overlap between mutations in ALL presentation and AML relapse (Figures 5A,B, Table S6). While ALL mutations were not associated with genes belonging to specific functional pathways, AML-specific mutations were associated with the regulation of transcription and chromatin binding and modification, further emphasising the notion of transcriptional reprogramming during lineage switch. Most of the subclonal mutations identified in presentation samples were subsequently lost at relapse, indicating alternative subclones as the origin of relapse. This included *KRAS* and *NRAS* mutations, which have previously been shown to confer a worse clinical outcome to infants with an *MLL*-rearranged ALL (Figure 5C)[23]. Also the MPALs harboured many mutations that were exclusively found in either the lymphoid or myeloid subpopulation indicating the presence of subclones with a lymphoid and myeloid bias (Figures 5A, B). These combined data show that lymphoid and myeloid leukaemic phenotypes are associated with distinctive mutation signatures both in lineage switches and in MPALs.

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

To identify factors contributing to the lineage plasticity in *MLL*/*AF4*-positive leukaemic cells, we compared all genes demonstrating differential expression, alternative splicing or mutation in the AML relapse (Figure 6A). This comparison highlighted eight genes common to all lineage-switched patients. One common gene was *CHD4*, which codes for the ATPase/helicase subunit of the Nucleosome Remodelling and Deacetylation complex (NuRD), a multiprotein transcriptional co-repressor complex with both histone deacetylase and ATP-dependent chromatin remodelling activity. NuRD is critical for lymphoid lineage determination by interacting with the transcription factor IKZF1[24-26]. *CHD4* shows significantly lower expression in lineage switched AML when compared to ALL presentation and is differentially spliced in the MPAL cases (Figure 3E, 6B). Finally, whilst *CHD4* mutations have been reported in <1.5% *MLL*-germline childhood ALL cases[27], as with the R1068H mutation found in the relapse of patient LS01, these variants commonly affect highly conserved residues in the helicase/ATPase domains and are predicted to disrupt its activity (Figure 6C,S5C)[28-30]. In contrast, recurrent mutations in other NuRD complex members have not been described in ALL and no other NuRD complex member was clonally mutated in our cohort (Table S6).

We therefore hypothesised that CHD4 was important in maintaining lineage fidelity in *MLL*/*AF4*-positive ALL. To test this idea, we performed knockdown experiments in the *MLL*/*AF4*-expressing and CD33-negative ALL cell line SEM, where we also included *ACAP1, DHX36, NCOA2, PHF3* and *PPP1R7* as five additional genes with potentially deleterious mutations in patient LS01 (Figures S6A). Reverse engineering of a mutual gene network from 216 ALL and AML gene expression data sets identified *CHD4* and *PHF3*, a co-factor in RNA Pol II-mediated transcription[31], as the most relevant network components of the mutated genes investigated (*PHF3* – 21 edges, p=0.010; *CHD4* – 12 edges, p=0.0005) (Figure S6B, Table S7)[32,33].

Only knockdown of *CHD4* and of *PHF3* robustly induced expression of the myeloid surface marker CD33 with a combined knockdown resulting in an even stronger CD33 expression (Figures S6A,C). Moreover, knockdown of either *CHD4* or *PHF3* also increased CD33 levels in RS4;11, another *MLL*/*AF4* ALL cell line, but not in the two *MLL*-germline ALL cell lines 697 and REH (Figure S6D), indicating that loss of CHD4 or PHF3 may only affect CD33 in the context of *MLL*/*AF4*. Finally, the combined knockdown of *CHD4* and *PHF3* in PDX from diagnostic ALL cells significantly increased the fraction of CD33+ cells from 8% to more than 25% (Figure S6E). These combined data suggest that *CHD4* and *PHF3* restrict *MLL/AF4*-positive leukaemic cells to a lymphoid phenotype.

In order to examine the role of additional mutations of chromatin modifiers found in our cohort, we investigated the impact of the PRC1 members *PCGF6* and *AUTS2,* genes with known roles in B lymphoid malignancy[34] and mutated in LS07RAML and LS08RAML (Figure 5A). While knockdown of *AUTS2* did not change CD33 levels, depletion of *PCGF6* increased CD33 surface expression in SEM cells, further supporting the notion of epigenetic factors in regulating lineage determination in ALL (Figure S6F).

In order to establish a direct link between CHD4 / PHF3 binding to the upregulation of myeloid genes, we investigated the impact of *CHD4* or *PHF3* perturbation on gene expression and chromatin organisation by performing RNA-seq, ATAC-seq and ChIP-seq for CHD4 in SEM cells and the *MLL* germline cell line 697 (Figures 6D,S7A,B, Table S8). In this analysis we ranked the ATAC-Seq and ChIP-Seq signals according to their fold-changes alongside the control patterns, which demonstrated that ATAC-seq analysis of control-treated SEM cells show a very similar pattern to CHD4 binding (Figure 6D) confirming that this factor is a global regulator of chromatin accessibility. Knockdown of both factors caused a shift in the overall chromatin accessibility pattern as shown by clustering analysis (Figure S7A,B bottom panels) suggesting that the after knockdown cells shifted their cistrome and thus their identity, whereby CHD4 knockdown resulted in a gain of open chromatin sites (Figure 6D, top panel). The knockdown of PHF3 caused both a loss and a gain of open

chromatin sites (Figure 6D, bottom panel). GSEA demonstrated a strong correlation of these gene expression changes in SEM cells after knockdown of CHD4 and PHF3 and lineage switch cases (Figure S7C,D). However, these changes were particular to *MLL*/*AF4* cells since in *MLL* germline 697 cells, CHD4 knockdown-induced changes in chromatin accessibility were not linked to altered gene expression, and knockdown of PHF3 did not affect chromatin accessibility (Figure 6D, right panels).

Knockdown of CHD4 or PHF3 in SEM cells changed chromatin structure and reduced expression of *CD79B, RAG2, VPREB1* and *CD22,* while concomitantly increasing transcription of *CEBPA, LYZ, SIRPA* and *CD33* (Figures 6E,S8A,B). However, 697 cells neither showed a change in immunophenotype nor altered expression of these genes suggesting that CHD4- and PHF3-mediated changes in gene expression correlate with the presence of an *MLL* fusion gene.

Given that the relapse-initiating cell may arise within an uncommitted, *MLL/AF4* translocated HSPC population, we assessed the impact of *CHD4* and *PHF3* function loss in a human cord blood model, which harbours a chimeric *MLL/Af4* fusion[35]. 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 6F,G, Table S9). Transcriptomic analysis of the sorted populations revealed that CD33 positive cells exhibited a metagene expression pattern similar to *MLL*r AML, while the pattern describing CD19+ cells was most similar to *MLLr* ALL, confirming that changes in surface marker expression were associated with the corresponding changes in the transcriptomic profiles (Figure S6G).

Taken together, our data show the important role of CHD4 and PHF3 in the epigenetic control of lymphoid lineage maintenance in *MLL/AF4*-positive leukaemia. In particular, dysregulation of *CHD4*/NuRD is mediated by mutation, down-regulation of expression and differential splicing across the entire cohort. These data support a role for these factors in

342 the lineage determining capacity of *MLL/AF4,* whilst their loss undermines execution and
343 maintenance of the lymphoid lineage fate.

344 ***Clonal evolution of AML relapse***

345 The observed cooperation of CHD4 and PHF3 in the control of lineage determination
346 predicted that both mutations co-occur in the same cell. Furthermore, since both mutations
347 might be required for the lineage switch in patient LS01, we hypothesised that they should
348 be detectable in the most immature populations of this AML sample, for which we had viable
349 cellular material. We therefore investigated the order of acquisition of these secondary
350 mutations within the structure of the normal haematopoietic hierarchy. Dissecting the relapse
351 AML sample using cell sorting, we isolated HSC-, MPP-, LMPP- and GMP-like, as well as
352 more differentiated populations, followed by targeted deep sequencing examining *MLL/AF4*
353 and 12 SNVs including mutated *CHD4* and *PHF3* that were unique to the relapse sample.
354 The fusion oncogene was found in the multipotential progenitor population (MPP,
355 CD34+CD38-CD45RA-CD90-) and in the lymphoid-primed multipotent progenitor-like
356 population (LMPP, CD34+CD38-CD45RA+; with lymphoid, myeloid, but not megakaryocyte-
357 erythroid potential) (Figures S9A,B; Table S10). When examining the presence of the 12
358 SNVs across the different populations, only *PHF3* and *CHD4* mutations were present within
359 the purified MPP-like fraction with VAF≥0.3 (Figure 7A, Table S10). In contrast, LMPP- and
360 GMP-like populations contained all 12 SNVs at high VAF. These findings place the *CHD4*
361 and *PHF3* mutations amongst the earliest genetic events in this patient during the evolution
362 of lineage-switched relapse. Moreover, they suggest, at least for this patient, an MPP-like or
363 even more immature cell population as the origin of relapse.

364 ***Cellular origin of lineage switched relapse***

365 In order to examine whether lineage-switched relapse regularly arises from lymphoid primed
366 or even earlier leukaemic populations, we examined whether relapsed AML cells contained
367 and even shared B-cell receptor (BCR) rearrangements with the preceding ALL. To

interrogate the developmental stage at which the myeloid relapse arose we analysed (BCR) rearrangements with RNA-seq and whole exome-seq (WES) derived data[36]. All ALL cases showed classical oligoclonal rearrangements of BCR loci, supporting the lymphoid lineage decision (Figure S9C, Table S11). We observed three distinct patterns for AML relapses (Figure 7B). Pattern 1 comprises AML cells with no BCR rearrangements implying the presence of a relapse-initiating cell residing in a primitive precursor population prior to early DJ recombination. This pattern was seen with patient LS01 and, together with the presence of *CHD4* and *PHF3* mutations, strongly supports an MPP-like population as a putative origin of relapse (Figure 7A). As a second pattern, we found unrelated BCR rearrangements, which may indicate either aberrant rearrangement in a myeloid cell or relapse initiating from B-lymphoid cell committed to undergo rearrangement, or a transdifferentiated minor ALL clone with an alternative rearrangement (Figure 7C, cases LS03, LS06, LS07, LS08, MPAL1, MPAL2). Interestingly, this pattern is found in a relapse after blinatumomab treatment (LS10) suggesting that immune escape may occur by direct transdifferentiation (Figure 7C). Pattern 3 shows shared BCR rearrangements between diagnostic and relapse material, which suggests a transdifferentiated myeloid relapse from the major ALL clone (cases LS05 and LS09). These data demonstrate that AML relapses can originate from different stages of lymphoid leukaemogenesis.

Discussion

This study describes impaired epigenetic control as being central to the phenomenon of lymphoid-myeloid lineage switch in *MLL/AF4* leukaemia, and demonstrates a heterogeneous cellular origin of relapse. The comparison of BCR rearrangements between matched ALL presentation and AML relapse cases demonstrates that whilst relapse can evolve directly from pro-B-like ALL blast populations, in keeping with the general self-renewal capacity of ALL cells[37], it can alternatively originate within the HSPC compartment. Indeed, the identification of *MLL*/*AF4*-expressing MPP-like cells shows that lineage switched relapse can

originate from very immature haemopoietic progenitor populations. This finding agrees with recently published data pointing at MPP cells as the origin of *MLL/AF4* leukaemia[38] 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[23]. Furthermore, the identification of *MLL*/*AF4* within HSPC populations is consistent with the recent identification of an early lymphoid progenitor, ELP-like signature specifically in *MLL*-rearranged ALL[39]. Nevertheless, and in agreement with previously published findings for MPALs[6], the data derived from the present cohort strongly support a non-lineage committed progenitor compartment as one source for lineage switched relapse. However, we can not exclude additional cells-of-origin of *MLL/AF4* ALL.

Irrespective of the cellular origin of the relapse, lineage switching is 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[40]. That decision is enacted by lymphoid master regulators including EBF1, PAX5 and IKAROS, which represent genes commonly mutated in precursor B-ALL and do not just upregulate B-cell specific genes, but also repress the myeloid program[40-44]. Pax5$^{-/-}$ pro-B cells which lack lymphoid potential, whilst capable of erythro-myeloid differentiation *in vitro*, still maintain expression of early B cell transcription factors *EBF1* and *E2A* (*TCF3*)[40]. In contrast, we show that lineage switching *MLL/AF4* pro-B leukaemic relapse is associated with a significant reduction in expression of these earliest B lymphoid transcription factors, which links to changes in the *MLL/AF4* transcriptional programme, ultimately establishing a myeloid differentiation fate. Unfortunately, we were not able to directly prove changes in transcription factor binding and associated changes in histone modifications due to the lack of available primary patient material. However, high resolution DHS-seq clearly demonstrated changes in chromatin accessibility and loss of occupation of the corresponding transcription factor binding sites.

The opposite scenario is observed when myeloid transcription factors are expressed in B-lymphoid cells[45]. Here, overexpression of C/EBPα efficiently reprograms such cells into macrophages by suppressing lymphoid genes. *CEBPA* is strongly upregulated after *CHD4* knockdown (Figure 6E) and is likely to be a driving force behind the lineage switch. Taken together, these published and newly presented data confirm that (i) the balance between lymphoid and myeloid transcription factors is instructive for lineage choice, and (ii) the down-regulation of the myeloid program is essential for the maintenance of the lymphoid fate.

How can the mutation of global chromatin regulators cause a switch in cell fate? Similar to the *Pax5* knockout, loss of IKAROS DNA-binding activity prevents lymphoid differentiation[26]. NuRD co-operates directly with IKAROS to repress HSC self-renewal and myeloid differentiation, permitting early lymphoid development[26,46,47]. Lineage switch was either associated with heterozygous mutation, reduced expression or, in the case of two MPALs, alternative splicing of *CHD4* and other NuRD components. These gene dosage effects are consistent with reports showing that complete loss of *CHD4* impairs normal and leukaemic proliferation[48,49], myeloid and lymphoid differentiation of HSPCs and causes exhaustion of HSC pools[46], indicating that basal CHD4 expression is required for maintaining AML. Moreover, a partial inhibition of *CHD4* supported induction of pluripotency in iPSCs, while a complete deletion eliminated cell proliferation, demonstrating that lowering *CHD4* expression may facilitate lineage promiscuitiy[50].

Recent studies have identified core NuRD and PRC1 complex members as being direct targets of *MLL/AF4* binding[51,52]. Moreover, NuRD components including CHD4 were shown to be part of an MLL supercomplex[53]. We therefore hypothesise that epigenetic regulator genes are recruited by lineage specific factors during *MLL/AF4* leukemogenesis and mediate fundamental lineage specific decision-making processes, in this case the repression of the myeloid lineage program. Multiple routes to their dysregulation may result in escape from this lineage restriction and may be enacted at different stages of haematopoiesis. However, importantly and in keeping with a previous murine study of lineage conversion following

447 CAR-T cell therapy, we did not identify evidence of relapse from a pre-existent myeloid
448 clone[54].

449 Of substantial clinical importance, lineage switch results in the loss of B cell surface markers
450 (*e.g.*, CD19), providing an alternative mechanism for relapse following CAR-T cell or
451 blinatumomab therapy[55,56] in addition to mutations, alternative splicing and T cell
452 trogocytosis[57-59]. Whilst these therapies target lineage specific surface markers, lineage-
453 switched (pre-)leukaemic progenitor populations escape epitope recognition and provide a
454 potential clonal source for the relapse[60]. As recognition of lineage switching following eg
455 CD19 CAR-T cell therapy grows, two recent studies have highlighted the particular
456 vulnerability of patients with *MLL*r ALL[54,61,62]. Given the increasing use of advanced
457 immunological therapies, a detailed understanding of the molecular processes underlying
458 lineage determination and switching will be critical for developing new strategies to avoid this
459 route to clinical relapse. Here we highlight an important role of epigenetic regulatory
460 complexes in the context of *MLL/AF4* leukaemia.

461

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 Monique den Boer, Frank van Leeuwen and Ronald Stam for critically reading the manuscript.

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

Author contributions

Conceptualization, O.H., S.B., C.B.; Methodology, O.H., C.B., R.T., K.S., P.M., S.B., A.P., 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., V.V.G., A.K., D.W., P.C.; Formal Analysis, S.N., M.A., J.H.K., V.V.G., A.K., D.W., P.C., P.K., C.B., O.H.; Investigation, R.T., K.S., P.M., A.P., C.M., P.S.C., H.J.B., S.G.K, A.K., S.A., M.R.I., E.K.S., P.E., H.M., A.E., N.M.S., 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., 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.; Data Curation, S.N., D.W., P.C.; Writing, S.B., O.H., C.B., R.T., K.S.; Supervision, O.H., S.B., 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.

Data 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. Requests for additional specific data/materials should be made to Olaf Heidenreich (O.T.Heidenreich@prinsesmaximacentrum.nl).

Conflict of interest statement

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

Financial support

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), a Kika programme grant to OH and JV (329), 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), the Sir Bobby Robson Foundation Clinical Fellowship and a Medical Research Council Clinician Scientist Fellowship (MR/S021590/1). Work in CB/PNC's lab was funded by a programme grant from Bloodwise (15001). Work in JMA's lab was funded by a programme grant from Bloodwise (13044). EZ was supported by an RFBR grant (№17-29-06052). Research in the VVG laboratory was supported in part by the Ministry of Education of the Republic of

510 Belarus, grant #3.04.3. Research in the AK laboratory was supported by an RSF grant (20-
511 75-10091).

512

References

1. Meyer C, Burmeister T, Groger D, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia*. 2018;32(2):273-284.
2. Moorman AV, Ensor HM, Richards SM, et al. 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*. 2010;11(5):429-438.
3. Germano G, Pigazzi M, del Giudice L, et al. Two consecutive immunophenotypic switches in a child with MLL-rearranged acute lymphoblastic leukemia. *Haematologica*. 2006;91(5 Suppl):ECR09.
4. Jiang JG, Roman E, Nandula SV, Murty VV, Bhagat G, Alobeid B. Congenital MLL-positive B-cell acute lymphoblastic leukemia (B-ALL) switched lineage at relapse to acute myelocytic leukemia (AML) with persistent t(4;11) and t(1;6) translocations and JH gene rearrangement. *Leuk Lymphoma*. 2005;46(8):1223-1227.
5. Rossi JG, Bernasconi AR, Alonso CN, et al. Lineage switch in childhood acute leukemia: an unusual event with poor outcome. *Am J Hematol*. 2012;87(9):890-897.
6. Alexander TB, Gu Z, Iacobucci I, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature*. 2018;562(7727):373-379.
7. Le Magnen C, Shen MM, Abate-Shen C. Lineage Plasticity in Cancer Progression and Treatment. *Annu Rev Cancer Biol*. 2018;2:271-289.
8. Quintanal-Villalonga A, Chan JM, Yu HA, et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat Rev Clin Oncol*. 2020;17(6):360-371.
9. Chen L, Kostadima M, Martens JHA, et al. Transcriptional diversity during lineage commitment of human blood progenitors. *Science*. 2014;345(6204):1251033.
10. Andersson AK, Ma J, Wang J, et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat Genet*. 2015;47(4):330-337.
11. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.
12. Zangrando A, Dell'orto MC, Te Kronnie G, Basso G. MLL rearrangements in pediatric acute lymphoblastic and myeloblastic leukemias: MLL specific and lineage specific signatures. *BMC Med Genomics*. 2009;2:36.
13. Gessner A, Thomas M, Castro PG, et al. Leukemic fusion genes MLL/AF4 and AML1/MTG8 support leukemic self-renewal by controlling expression of the telomerase subunit TERT. *Leukemia*. 2010;24(10):1751-1759.
14. Somervaille TC, Matheny CJ, Spencer GJ, et al. Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell*. 2009;4(2):129-140.
15. Wilkinson AC, Ballabio E, Geng H, et al. RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interaction. *Cell Rep*. 2013;3(1):116-127.
16. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res*. 2013;41(21):e201.
17. Bonifer C, Cockerill PN. Chromatin Structure Profiling Identifies Crucial Regulators of Tumor Maintenance. *Trends Cancer*. 2015;1(3):157-160.
18. Kreher S, Bouhlel MA, Cauchy P, et al. Mapping of transcription factor motifs in active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma. *Proc Natl Acad Sci U S A*. 2014;111(42):E4513-4522.
19. Howell ED, Yzaguirre AD, Gao P, et al. Efficient hemogenic endothelial cell specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-regulated TGFbeta target genes. *Genes Dev*. 2021;35(21-22):1475-1489.

20. Hohaus S, Petrovick MS, Voso MT, Sun Z, Zhang DE, Tenen DG. PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene. *Mol Cell Biol*. 1995;15(10):5830-5845.
21. Leddin M, Perrod C, Hoogenkamp M, et al. Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. *Blood*. 2011;117(10):2827-2838.
22. Dobbins SE, Sherborne AL, Ma YP, et al. The silent mutational landscape of infant MLL-AF4 pro-B acute lymphoblastic leukemia. *Genes Chromosomes Cancer*. 2013;52(10):954-960.
23. Agraz-Doblas A, Bueno C, Bashford-Rogers R, et al. Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. *Haematologica*. 2019;104(6):1176-1188.
24. Arends T, Dege C, Bortnick A, et al. CHD4 is essential for transcriptional repression and lineage progression in B lymphopoiesis. *Proc Natl Acad Sci U S A*. 2019;116(22):10927-10936.
25. Ng SY, Yoshida T, Zhang J, Georgopoulos K. Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity*. 2009;30(4):493-507.
26. Zhang J, Jackson AF, Naito T, et al. Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat Immunol*. 2011;13(1):86-94.
27. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555(7696):371-376.
28. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*. 2011;88(4):440-449.
29. Sifrim A, Hitz MP, Wilsdon A, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*. 2016;48(9):1060-1065.
30. Novillo A, Fernandez-Santander A, Gaibar M, et al. Role of Chromodomain-Helicase-DNA-Binding Protein 4 (CHD4) in Breast Cancer. *Front Oncol*. 2021;11:633233.
31. Appel LM, Franke V, Bruno M, et al. PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat Commun*. 2021;12(1):6078.
32. Kang H, Chen IM, Wilson CS, et al. Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood*. 2010;115(7):1394-1405.
33. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA*. 2010;304(24):2706-2715.
34. Ferreira BI, Garcia JF, Suela J, et al. 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*. 2008;93(5):670-679.
35. Lin S, Luo RT, Ptasinska A, et al. Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell*. 2016;30(5):737-749.
36. Bolotin DA, Poslavsky S, Davydov AN, et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol*. 2017;35(10):908-911.
37. le Viseur C, Hotfilder M, Bomken S, et al. In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties. *Cancer Cell*. 2008;14(1):47-58.
38. Malouf C, Ottersbach K. The fetal liver lymphoid-primed multipotent progenitor provides the prerequisites for the initiation of t(4;11) MLL-AF4 infant leukemia. *Haematologica*. 2018;103(12):e571-e574.
39. Khabirova E, Jardine L, Coorens THH, et al. Single-cell transcriptomics reveals a distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic leukemia. *Nat Med*. 2022;28(4):743-751.

40. Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*. 1999;401(6753):556-562.
41. Mullighan CG, Kennedy A, Zhou X, et al. Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. *Leukemia*. 2007;21(9):2000-2009.
42. Boer JM, van der Veer A, Rizopoulos D, et al. Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. *Leukemia*. 2016;30(1):32-38.
43. Witkowski MT, Hu Y, Roberts KG, et al. Conserved IKAROS-regulated genes associated with B-progenitor acute lymphoblastic leukemia outcome. *J Exp Med*. 2017;214(3):773-791.
44. Pongubala JM, Northrup DL, Lancki DW, et al. Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat Immunol*. 2008;9(2):203-215.
45. van Oevelen C, Collombet S, Vicent G, et al. C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports*. 2015;5(2):232-247.
46. Yoshida T, Hazan I, Zhang J, et al. The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev*. 2008;22(9):1174-1189.
47. Lu X, Chu CS, Fang T, et al. MTA2/NuRD Regulates B Cell Development and Cooperates with OCA-B in Controlling the Pre-B to Immature B Cell Transition. *Cell Rep*. 2019;28(2):472-485 e475.
48. Sperlazza J, Rahmani M, Beckta J, et al. Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation. *Blood*. 2015;126(12):1462-1472.
49. Heshmati Y, Turkoz G, Harisankar A, et al. The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia. *Haematologica*. 2018;103(7):1169-1181.
50. Mor N, Rais Y, Sheban D, et al. Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naive Pluripotency. *Cell Stem Cell*. 2018;23(3):412-425 e410.
51. Kerry J, Godfrey L, Repapi E, et al. MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia. *Cell Rep*. 2017;18(2):482-495.
52. Harman JR, Thorne R, Jamilly M, et al. A KMT2A-AFF1 gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes. *Genome Res*. 2021.
53. Nakamura T, Mori T, Tada S, et al. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell*. 2002;10(5):1119-1128.
54. Jacoby E, Nguyen SM, Fountaine TJ, et al. CD19 CAR immune pressure induces B-precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic plasticity. *Nat Commun*. 2016;7:12320.
55. Gardner R, Wu D, Cherian S, et al. Acquisition of a CD19-negative myeloid phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy. *Blood*. 2016;127(20):2406-2410.
56. Rayes A, McMasters RL, O'Brien MM. Lineage Switch in MLL-Rearranged Infant Leukemia Following CD19-Directed Therapy. *Pediatr Blood Cancer*. 2016;63(6):1113-1115.
57. Sotillo E, Barrett DM, Black KL, et al. Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. *Cancer Discov*. 2015;5(12):1282-1295.
58. Rabilloud T, Potier D, Pankaew S, Nozais M, Loosveld M, Payet-Bornet D. Single-cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse after CAR-T therapy. *Nat Commun*. 2021;12(1):865.

59. Hamieh M, Dobrin A, Cabriolu A, et al. CAR T cell trogocytosis and cooperative killing regulate tumour antigen escape. *Nature*. 2019;568(7750):112-116.
60. Liao W, Kohler ME, Fry T, Ernst P. Does lineage plasticity enable escape from CAR-T cell therapy? Lessons from MLL-r leukemia. *Exp Hematol*. 2021;100:1-11.
61. Lamble AJ, Myers RM, Taraseviciute A, et al. KMT2A Rearrangements Are Associated with Lineage Switch Following CD19 Targeting CAR T-Cell Therapy. *Blood*. 2021;138(Supplement 1):256-256.
62. Leahy AB, Devine KJ, Li Y, et al. Impact of high-risk cytogenetics on outcomes for children and young adults receiving CD19-directed CAR T-cell therapy. *Blood*. 2022;139(14):2173-2185.

Figure Legends

Figure 1. Characterisation of *MLL/AF4* lineage switch cases. (A) Morphological change from lymphoblastic leukaemia (left panel) to acute monoblastic/monocytic leukaemia (right panel) in patient LS01. 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.

Figure 2. 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 Wald statistics. (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) Differential expression of lineage specific and (D) immunoglobulin recombination machinery genes 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 3. Alternative splicing in lineage switch and MPAL cases. (A) 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. (B) 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 (left panel). Venn diagrams (right

719 panel) showing distribution of splice variants identified as significantly changed in AML (or
720 myeloid fraction of MPAL patients), including exon-exon junctions (DEEj), differential exon
721 usage (DEU) and retained introns (RI). (C) Enrichment analysis of affected signalling
722 pathways by the exon-exon junctions (DEEj) and differential exome usage (DEU) in the
723 LSAL AML relapse and myeloid compartment of MPAL patients. Pathway enrichment
724 analysis has been performed with https://biit.cs.ut.ee/gprofiler/gost under the highest
725 significance threshold, with multiple testing correction (g:SCS algorithm). (D) Fold log2
726 change of total transcript levels among genes affected by alternative splicing (left panel), and
727 of differentially spliced variants in lineage switched and myeloid compartments of MPAL
728 patients (right panel). (E) Schematic representation of impact of alternative splicing on
729 mRNA composition and open reading frames (ORFs) of selected genes. Column graphs on
730 the right indicate corresponding fold changes of variant expression between AML (or
731 myeloid) and ALL (or lymphoid) populations in two tested lineage switch patients (LS03 and
732 LS04) and one MPAL.

733 **Figure 4. Chromatin re-organisation and differential transcription factor binding**
734 **underpins lineage switching.** (A) DNaseI hypersensitive site sequencing identifies 13,619
735 sites with a log2 fold reduction and 12,203 sites with a log2 fold increase following lineage
736 switch to AML. Relative peak heights in the AML sample were plotted against those of the
737 ALL sample. (B) University of California, Santa Cruz (UCSC) genome browser screenshot
738 displaying differential expression at lineage specific loci (lower red tracks) accompanied by
739 altered DNaseI hypersensitivity (upper black tracks) proximal to the transcriptional start site
740 (TSS) of *CD33*. (C) UCSC genome browser screenshot for *CD19* zoomed in on an ALL-
741 associated DHS with EBF occupation as indicated by high resolution DHS-seq and
742 Wellington analysis. FP - footprint. (D) Heat maps showing distal DHS regions specific for
743 AML relapse on a genomic scale. Red and green indicate excess of positive and negative
744 strand cuts, respectively, per nucleotide position. Sites are sorted from top to bottom in order
745 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.

Figure 5. Molecular characterisation of lineage switch *MLL/AF4* leukaemias. (A) Whole exome sequencing (WES) data showing genes recurrently mutated within the analysed cohort and genes clonally mutated in relapse cases belonging to the same function protein complexes (e.g. DNA polymerases, epigenetic complexes, transcriptional regulators). Data are presented according to the disease timepoint/cell lineage and age of the patient. Depicted are major single nucleotide variants (SNVs)/indels that were found in >30% of reads and minor SNVs/indels present in <30% reads. (B) Comparison of total mutation load (SNVs and indels) identified in patients at presentation (ALL) and relapse (AML) disease stage or lymphoid and myeloid fraction in MPALs. Listed are common SNVs predicted (by Condel scoring) to have deleterious effect. (C) Evolution of KRAS/NRAS mutation carrying cells during lineage switch process. Clonal vs sub-clonal mutations were defined based on variant allelic frequencies (VAFs) of identified hit at setup cutoff equal to 30%.

Figure 6. Epigenetic modulatory genes influence lineage specific expression profiles. (A) Intersection between identified hits of clonal mutations (VAF>30%), differentially expressed genes and alternatively spliced, differentially used exon-exon junctions (adj.p-value<0.01) in lineage switched myeloid relapse/myeloid fraction of MPALs, present in the analysed cohort. (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) CHD4 structure; the R1068H mutation (red) 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. (D) Identification of regions of differential chromatin accessibility before and after knockdown of *CHD4* and *PHF3* depicted in red in *MLL*r SEM

cells (left panel) and non-*MLL*r 697 cells (right panel). For all reads the fold change in ATAC-peak height was calculated relative to the control (shNTC) and ATAC-peaks from knock-down cells were plotted according to their fold-change along-side the control signals. CHD4 ChIP density plots from SEM cells (depicted in blue) were plotted alongside the corresponding DNA regions of the shNTC control. Differentially expressed genes associated with changing ATAC peaks (log2FC analysed vs shNTC) identified in each cellular variant are represented by heatmaps included at the right side of each panel (for SEM and 697 cells). (E) UCSC genome browser screenshots representing differential chromatin accessibility (ATAC-seq) and gene expression level (RNA-seq) in the myeloid *CEBPA* and the lymphoid *RAG2* loci following *CHD4* and *PHF3* knockdown in *MLL*r SEM cells and non-*MLL*r 697 cells. ChIP-seq traces representing normal CHD4 occupancy in non-*MLL*r B-ALL (REH cells), *MLL*r B-ALL (SEM cells) and *MLL*r AML cells (MV-4;11) are shown as a reference at the bottom of each screenshot. TSS – transcriptional start site is depicted for each gene. (F) Expression of the 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) *PHF3* knockdown is capable of influencing the surface marker expression after longer incubation period (33 days); *CHD4* knockdown impaired cellular survival upon longer *in vitro* culture (data not shown).

Figure 7. Haematopoietic stem/progenitor populations carry *MLL/AF4*. (A) Summary of *MLL/AF4* positivity and 12 SNVs exclusive for the AML relapse, within different populations analysed in patient LS01RAML 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 (described in Table S10). (B) Summary of the proposed model of the origin of lineage switched relapse. 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. Pattern 2 - B-cell receptor-containing clones on ALL and AML, but distinct to each other. Pattern 3 - B-cell receptor-containing clones shared between ALL and AML. (C) BCR clones frequencies identified in whole-exome seq data with application of MiXCR software in all analysed LSAL and MPAL patients.

Figure 1

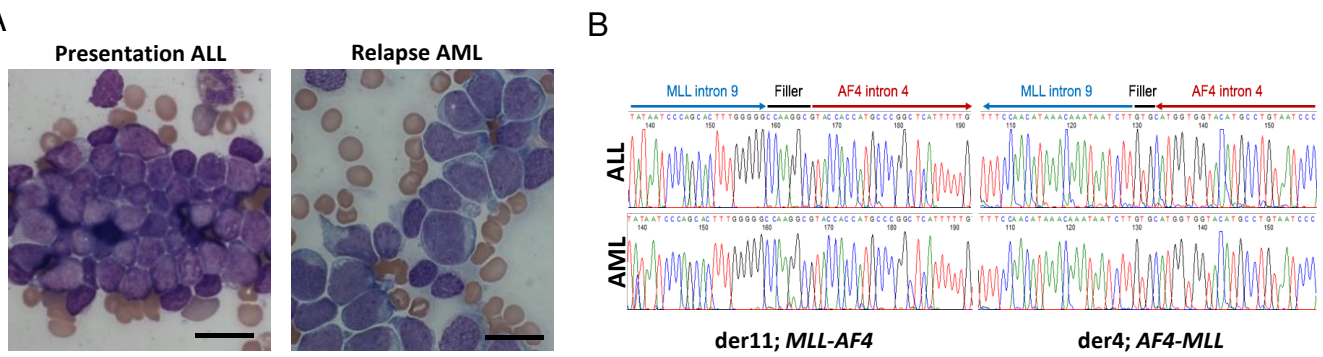


Figure 2

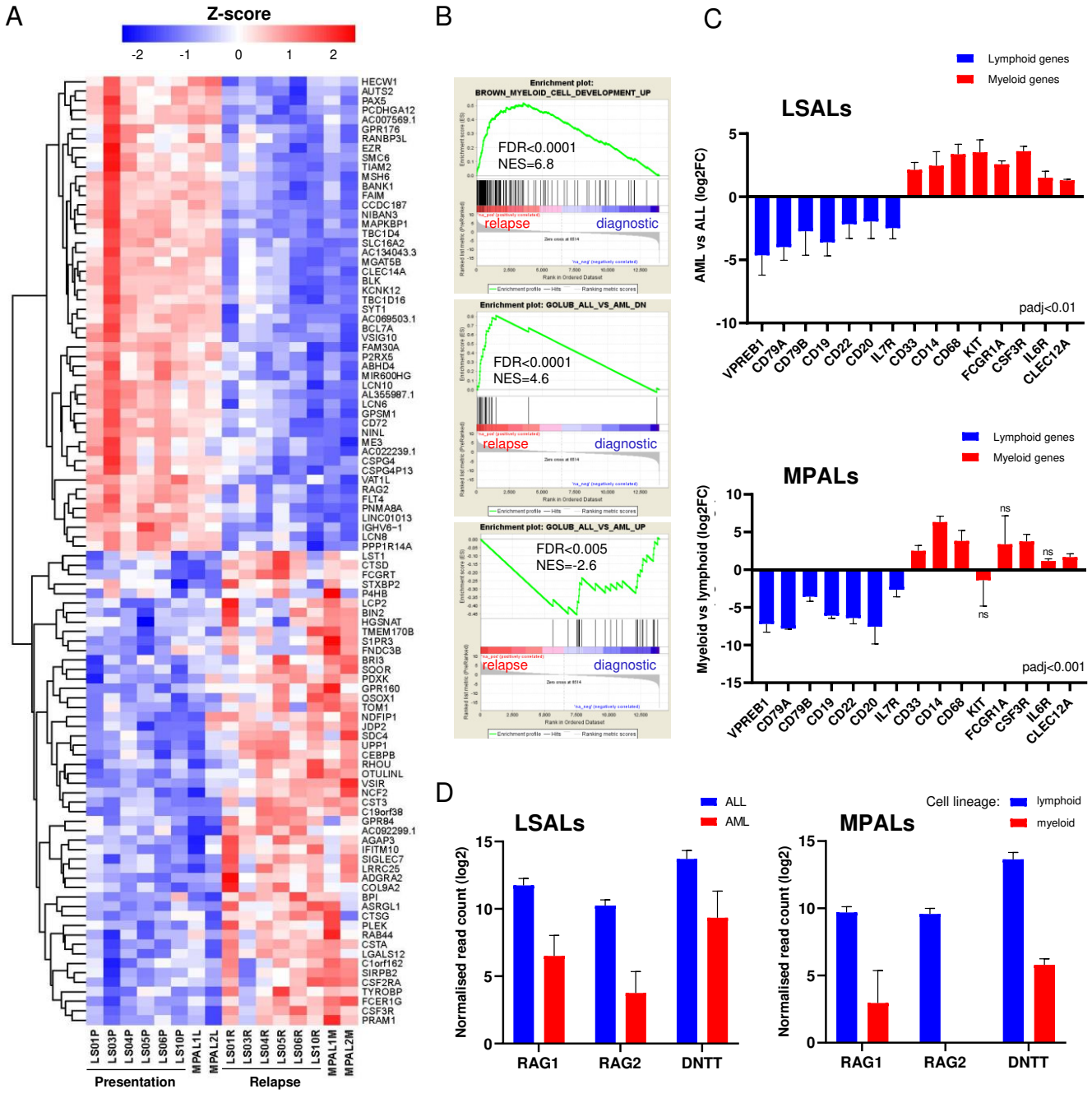


Figure 3

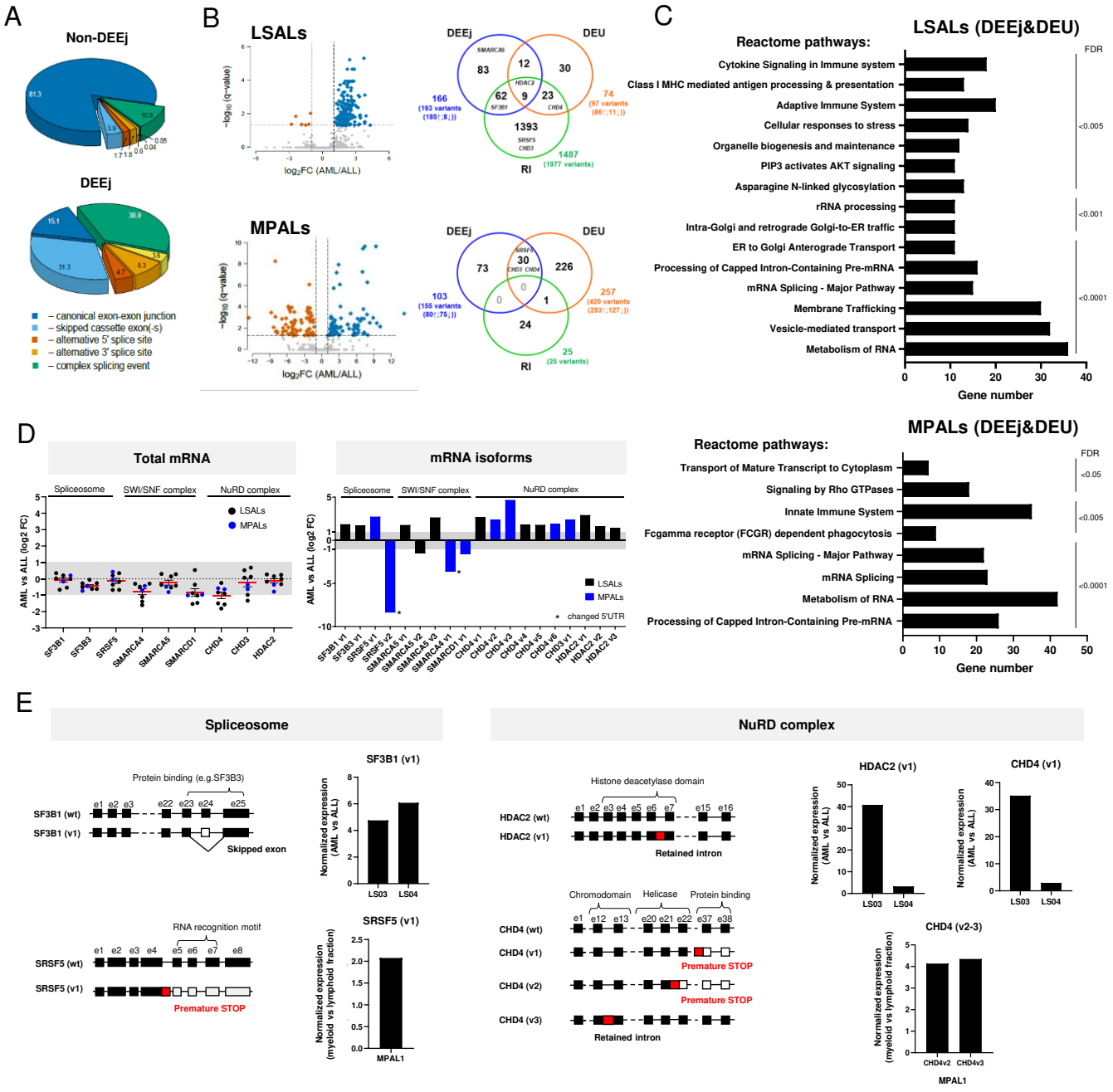
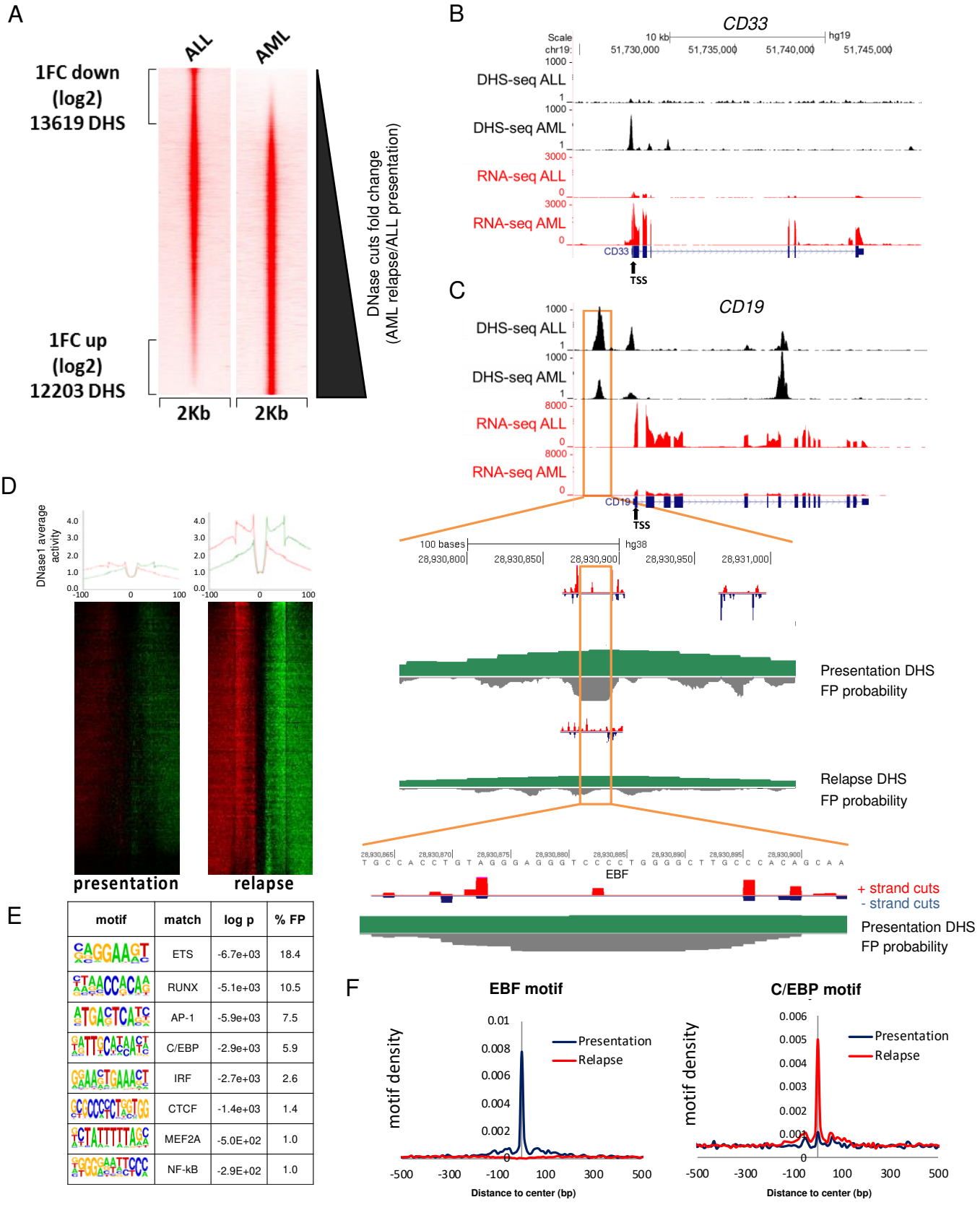


Figure 4



motif	match	log p	% FP
	ETS	-6.7e+03	18.4
	RUNX	-5.1e+03	10.5
	AP-1	-5.9e+03	7.5
	C/EBP	-2.9e+03	5.9
	IRF	-2.7e+03	2.6
	CTCF	-1.4e+03	1.4
	MEF2A	-5.0E+02	1.0
	NF-kB	-2.9E+02	1.0

Figure 5

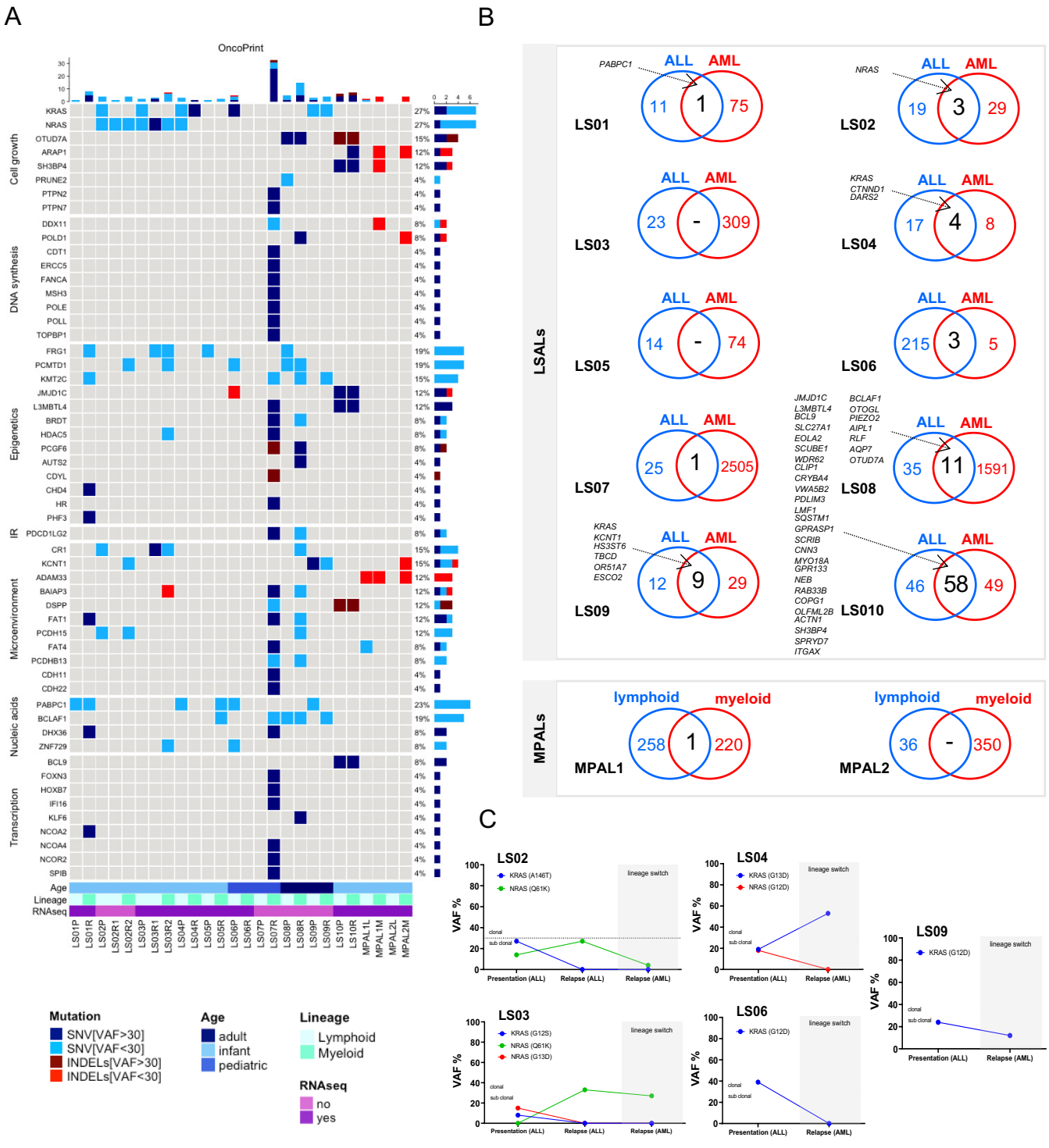
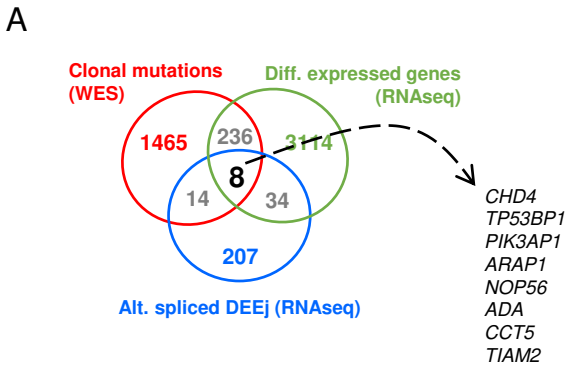
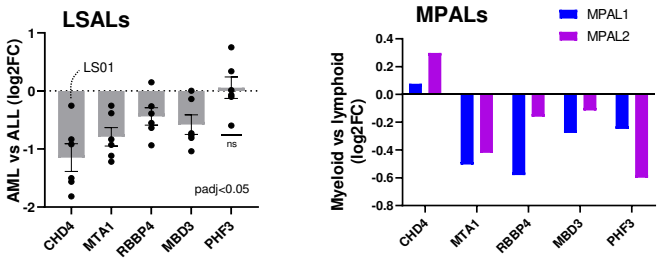
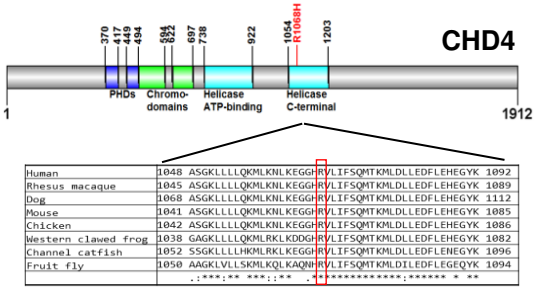
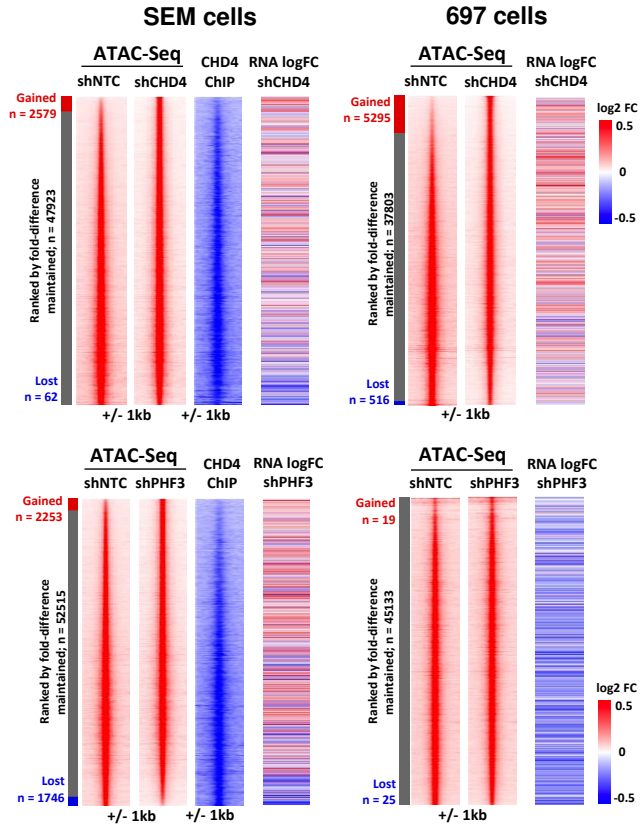


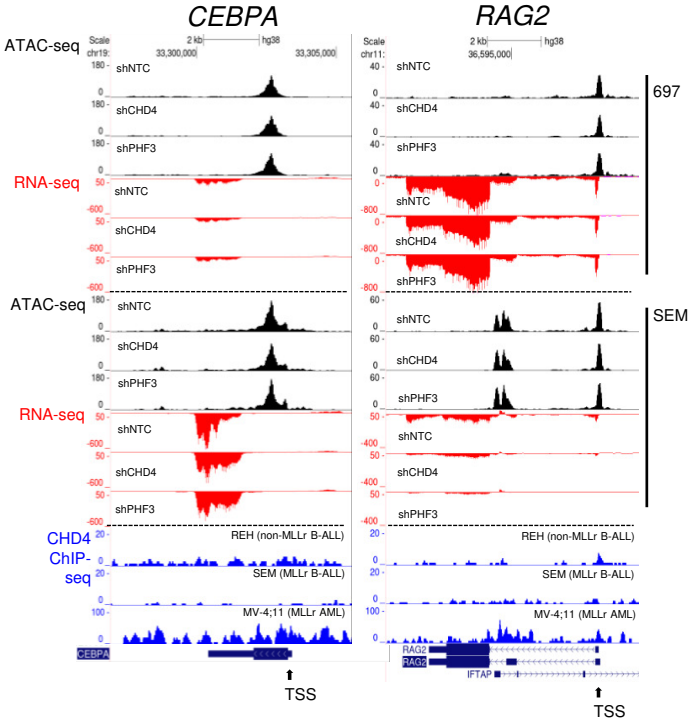
Figure 6

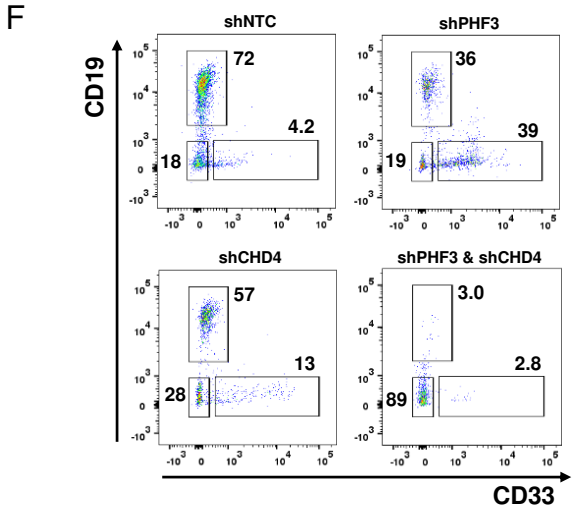












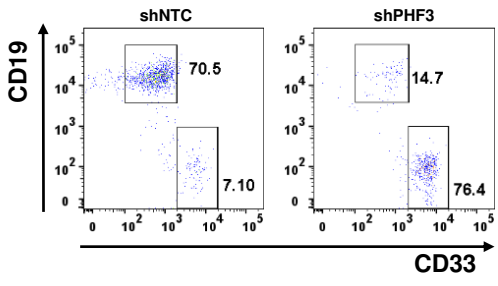


Figure 7