



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/189446/>

Version: Accepted Version

Article:

Tirtakusuma, Ricky, Szoltysek, Katarzyna, Milne, Paul et al. (2022) Epigenetic regulator genes direct lineage switching in MLL/AF4 leukaemia. *Blood*. blood.2021015036. pp. 1875-1890. ISSN: 1528-0020

<https://doi.org/10.1182/blood.2021015036>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

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 Ptasińska (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) Mino 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):

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

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

35 ¹³Illumina Cambridge Ltd., Great Abington, UK

36 ¹⁴Department of Pediatric Oncology/Hematology, Charité Universitätsmedizin Berlin, Berlin,
37 Germany

38 ¹⁵St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria

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

41 ¹⁷Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK

42 ¹⁸Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton UK

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

45 ²⁰Department of Medicine, Section of Hematology/Oncology, University of Chicago, Chicago,
46 USA

47 ***Authors contributed equally**

48 ****Co-senior authors**

49 **†Co-corresponding Authors.**

50

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

52 **Keywords:** *MLL/AF4*; *KMT2A/AFF1*; acute lymphoblastic leukaemia; nucleosome

53 remodelling and deacetylation complex (NuRD); chromatin remodelling

54 **Corresponding authors**

55 Dr Simon Bomken
56 Wolfson Childhood Cancer Research Centre
57 Translational and Clinical Research Institute
58 Level 6 Herschel Building
59 Brewery Lane
60 Newcastle University
61 Newcastle upon Tyne
62 NE1 7RU, UK

63

64 Tel: +44 (0)191 2082231

65 E mail: s.n.bomken@ncl.ac.uk

66

67 Professor Olaf Heidenreich
68 Princess Maxima Center for Pediatric Oncology
69 Heidelberglaan 25
70 3584 CS Utrecht
71 The Netherlands

72

73 Tel: +31 (0)88 972 7272

74 E mail: o.t.heidenreich@prinsesmaximacentrum.nl

75

76	
77	Abstract word count - 169
78	Manuscript word count - 4054
79	
80	Figures - 7
81	Tables - 0
82	References - 62

83 Keypoints

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

88 Abstract

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

104

105

106

107 Introduction

108 Translocation of Mixed Lineage Leukaemia (*MLL*) with one of over 130 alternative partner
109 genes is a recurrent cytogenetic finding in both acute myeloid (AML) and lymphoblastic
110 leukaemias (ALL) and is generally associated with poor prognosis^{1,2}. Amongst the most
111 common translocations is t(4;11)(q21;q23), forming the *MLL/AF4* (also known as
112 *KMT2A/AFF1*) fusion gene. Uniquely amongst *MLL* rearrangements (*MLLr*), *MLL/AF4* is
113 almost exclusively associated with pro-B cell acute lymphoblastic leukaemia and is
114 prototypical of infant acute lymphoblastic leukaemia (ALL) where it carries a very poor
115 prognosis¹. However, despite this general lymphoid presentation, *MLL/AF4* leukaemias have
116 an intriguing characteristic - that of lineage switched relapses. Lineage switch acute
117 leukaemias (LSALs) lose their lymphoid specific features and gain myeloid phenotype upon
118 relapse³⁻⁵. Alternatively, *MLL/AF4* leukaemias may harbour distinct lymphoid and myeloid
119 populations at the same time, thus classifying as mixed phenotype acute leukaemias
120 (MPALs) of the bilineage subtype⁶.

121 Lineage plasticity has been associated with the loss of original therapeutic targets^{7,8}. In
122 order to understand the molecular basis of lineage promiscuity and switching, we examined
123 a unique cohort of *MLL/AF4*-positive LSAL presentation/relapse pairs and MPALs. We
124 demonstrate that disruption of the epigenetic machinery, including the nucleosome
125 remodelling and deacetylation complex (NuRD), is associated with the loss of lymphoid
126 restriction. Lineage switch is then enacted through redistribution of transcription factor
127 binding and chromatin reorganisation. These findings provide novel insight into factors which
128 may prove critical to the effective implementation of lineage specific, epitope-directed
129 therapies such as chimeric antigen receptor T-cell (CAR-T) cell or bi-specific T-cell engaging
130 antibody (BiTE) approaches.

131

132 **Methods**

133 ***Patient samples and data***

134 Patients were diagnosed by local haematology specialists according to contemporary clinical
135 diagnostic criteria based on morphology and immunophenotypic analysis. All patient
136 samples were collected at the point of diagnosis, remission following treatment or relapse
137 and stored with written informed consent for research in one of six centres (Newcastle
138 Haematology Biobank, Newcastle, UK; University Hospital Schleswig-Holstein, Kiel,
139 Germany; Dmitry Rogachev National Medical Research Center of Pediatric Hematology,
140 Oncology and Immunology, Moscow, Russia; Haematological Malignancy Diagnostic
141 Service, Leeds, UK; Princess Maxima Center for Pediatric Oncology, Utrecht, The
142 Netherlands; Cincinnati Children's Hospital Medical Center, Cincinnati, USA). Mononuclear
143 cells were isolated from bone marrow or peripheral blood by density centrifugation followed
144 by immediate extraction of DNA or RNA, or cryopreservation in the presence of 10% v/v
145 DMSO.

146 Samples were requested and used in accordance with the ethical approvals granted to each
147 of the local/institutional ethical review boards (NRES Committee North East - Newcastle &
148 North Tyneside 1, UK, reference 07/H0906/109+5; Medical Faculty Christian-Albrechts
149 University, Kiel, reference A 103/08; Dmitry Rogachev National Medical Research Center,
150 Moscow, references MB2008: 22.01.2008, MB2015: 22.01.2015, ALL-REZ-2014:
151 28.01.2014; Haematological Malignancy Research Network, Yorkshire, UK, reference
152 04/Q1205/69; Haematological Malignancy Diagnostic Service, Leeds, UK, reference
153 14/WS/0098; Erasmus MC METC, Netherlands, reference MEC-2016-739; IRB of Cincinnati
154 Children's Hospital, USA, reference 2010-0658) and in accordance with the Declaration of
155 Helsinki. Each patient/sample was allocated an anonymised reference and no identifiable
156 information was shared.

157 Additional methods are described in Supplemental Methods.

159 **Results**

160 ***Characterisation of MLL/AF4 acute leukaemias with lineage switch***

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

173 ***Lineage switch leukaemia is associated with transcriptional reprogramming***

174 We hypothesized that lineage switch would be linked with changes in gene expression.
175 Since the changes in transcriptome composition may include altered regulation of both
176 transcription and mRNA maturation⁹, we compared gene expression and splicing between
177 lymphoid and myeloid populations from lineage switch and MPAL patients. Cluster analysis
178 of differential gene expression robustly separated both population types (Figure 2A). We
179 identified 1374 up- (adj. $p < 0.01$, Log Fold change > 2) and 1323 down-regulated genes in the
180 AML lineage switches and the myeloid populations of MPAL patients linked to reduced
181 lymphoid and increased myeloid gene expression (Figure 2B, Table S2). Changed gene
182 expression included the loss of lymphoid genes such as *PAX5*, *EBF1*, *CD19*, *CD20* (*MS4A1*)
183 and *CD22*, diminished gene expression of immunoglobulin genes and genes involved in the
184 VDJ recombination (*RAG1*, *RAG2*, *DNTT*), and a gain of myeloid gene including *CLEC12A*,
185 *PRAM1*, *CSF3R* and members of the *CEBP* transcription factor family (Figures

186 2C,D,S2A,B)¹⁰⁻¹². Moreover, almost 30% of direct *bona fide* target genes of *MLL/AF4*
187 including *PROM1*, encoding the stem cell marker CD133, *IKZF2* and *HOXA7* showed lower
188 expression in myeloid cells despite sharing the same *MLL/AF4* isotype (Figures S3A-D,
189 Table S2)¹³⁻¹⁵. These data show that lineage switch also involves differential *MLL/AF4*-driven
190 gene expression.

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

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

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

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

216 The substantial gene expression changes, including those affecting epigenetic regulators
217 and lineage-determining transcription factors, prompted us to link transcriptional changes to
218 altered genome-wide chromatin accessibility. High resolution DNaseI hypersensitive site
219 (DHS) mapping combined with digital footprinting analysis using the Wellington algorithm¹⁶
220 uncovered multiple differentially accessible genes including the hematopoietic surface
221 marker genes CD33 and CD19 and transcription factors (Figures 4A-C,S4A,B). These
222 alterations occurred both at locations distal and proximal to transcriptional start sites (TSS)
223 indicating the involvement of enhancers and promoters (Figures 4D,S4C). Digital footprinting
224 is now generally accepted to highlight factors important for regulating specific cell fates¹⁷⁻¹⁹.
225 These analyses showed that changes in chromatin accessibility after lineage switch were
226 linked to an altered pattern of transcription factor binding site occupancy (Figures 4E,S4D)
227 with a loss of occupancy of consensus binding sites for lymphoid transcription factors
228 including EBF or PAX5 and a corresponding increased occupancy of binding motifs for
229 myeloid factors including C/EBP family members (Figures 4E,F). We also observed a
230 redistribution of footprinted sites for transcription factors controlling both lymphoid and
231 myeloid maturation such as RUNX, AP-1 and ETS family members to alternative cognate
232 motifs (Figures 4E,S4D)^{20,21}. This finding is exemplified by decreased accessibility of a
233 region located 1 kb upstream of the *CD19* TSS with concomitant loss of EBF binding site
234 occupancy at this element (Figure 4C). In conclusion, the transition from lymphoid to myeloid
235 immunophenotype is associated with genome-wide alterations in chromatin accessibility and
236 transcription factor binding site occupancy.

237 **The mutational landscape of lineage switch**

238 Next, we examined the mutational landscape of lineage switched *MLL/AF4* leukaemias by
239 performing exome sequencing on the entire cohort. In agreement with previously reported
240 mutation rates in *MLL*-rearranged leukaemias, presentation ALLs displayed a relatively quiet
241 mutational landscape with a median of 25 nonsynonymous somatic single nucleotide
242 variants (SNVs) or insertions/deletions (indels) (Figures S5A,B, Table S6)^{10,22}. Most of them
243 were sub-clonal with less than 30% of the reads. The group of AML relapses showed on
244 average 92 SNVs and indels. However, this increase was due to the more heterogeneous
245 composition of the relapse group: two cases (LS07AML and LS08AML) carried mutated
246 DNA polymerase genes resulting in increased mutational burden. We observed this
247 phenotype in only two out of ten relapses, arguing against this phenomenon being a general
248 requirement for the lineage switch.

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

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

264 To identify factors contributing to the lineage plasticity in *MLL/AF4*-positive leukaemic cells,
265 we compared all genes demonstrating differential expression, alternative splicing or mutation
266 in the AML relapse (Figure 6A). This comparison highlighted eight genes common to all
267 lineage-switched patients. One common gene was *CHD4*, which codes for the
268 ATPase/helicase subunit of the Nucleosome Remodelling and Deacetylation complex
269 (NuRD), a multiprotein transcriptional co-repressor complex with both histone deacetylase
270 and ATP-dependent chromatin remodelling activity. NuRD is critical for lymphoid lineage
271 determination by interacting with the transcription factor IKZF1²⁴⁻²⁶. *CHD4* shows significantly
272 lower expression in lineage switched AML when compared to ALL presentation and is
273 differentially spliced in the MPAL cases (Figure 3E, 6B). Finally, whilst *CHD4* mutations have
274 been reported in <1.5% *MLL*-germline childhood ALL cases²⁷, as with the R1068H mutation
275 found in the relapse of patient LS01, these variants commonly affect highly conserved
276 residues in the helicase/ATPase domains and are predicted to disrupt its activity (Figure
277 6C,S5C)²⁸⁻³⁰. In contrast, recurrent mutations in other NuRD complex members have not
278 been described in ALL and no other NuRD complex member was clonally mutated in our
279 cohort (Table S6).

280 We therefore hypothesised that *CHD4* was important in maintaining lineage fidelity in
281 *MLL/AF4*-positive ALL. To test this idea, we performed knockdown experiments in the
282 *MLL/AF4*-expressing and CD33-negative ALL cell line SEM, where we also included *ACAP1*,
283 *DHX36*, *NCOA2*, *PHF3* and *PPP1R7* as five additional genes with potentially deleterious
284 mutations in patient LS01 (Figures S6A). Reverse engineering of a mutual gene network
285 from 216 ALL and AML gene expression data sets identified *CHD4* and *PHF3*, a co-factor in
286 RNA Pol II-mediated transcription³¹, as the most relevant network components of the
287 mutated genes investigated (*PHF3* – 21 edges, $p=0.010$; *CHD4* – 12 edges, $p=0.0005$)
288 (Figure S6B, Table S7)^{32,33}.

289 Only knockdown of *CHD4* and of *PHF3* robustly induced expression of the myeloid surface
290 marker CD33 with a combined knockdown resulting in an even stronger CD33 expression
291 (Figures S6A,C). Moreover, knockdown of either *CHD4* or *PHF3* also increased CD33 levels
292 in RS4;11, another *MLL/AF4* ALL cell line, but not in the two *MLL*-germline ALL cell lines 697
293 and REH (Figure S6D), indicating that loss of *CHD4* or *PHF3* may only affect CD33 in the
294 context of *MLL/AF4*. Finally, the combined knockdown of *CHD4* and *PHF3* in PDX from
295 diagnostic ALL cells significantly increased the fraction of CD33+ cells from 8% to more than
296 25% (Figure S6E). These combined data suggest that *CHD4* and *PHF3* restrict *MLL/AF4*-
297 positive leukaemic cells to a lymphoid phenotype.

298 In order to examine the role of additional mutations of chromatin modifiers found in our
299 cohort, we investigated the impact of the PRC1 members *PCGF6* and *AUTS2*, genes with
300 known roles in B lymphoid malignancy³⁴ and mutated in LS07RAML and LS08RAML (Figure
301 5A). While knockdown of *AUTS2* did not change CD33 levels, depletion of *PCGF6* increased
302 CD33 surface expression in SEM cells, further supporting the notion of epigenetic factors in
303 regulating lineage determination in ALL (Figure S6F).

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

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

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

328 Given that the relapse-initiating cell may arise within an uncommitted, *MLL/AF4* translocated
329 HSPC population, we assessed the impact of *CHD4* and *PHF3* function loss in a human
330 cord blood model, which harbours a chimeric *MLL/AF4* fusion³⁵. Knockdown of either *CHD4*
331 or *PHF3* under lymphoid culture conditions significantly impaired lymphoid differentiation
332 potential, whilst co-knockdown of *CHD4* and *PHF3* disrupted differentiation entirely (Figures
333 6F,G, Table S9). Transcriptomic analysis of the sorted populations revealed that *CD33*
334 positive cells exhibited a metagene expression pattern similar to *MLLr* AML, while the
335 pattern describing *CD19+* cells was most similar to *MLLr* ALL, confirming that changes in
336 surface marker expression were associated with the corresponding changes in the
337 transcriptomic profiles (Figure S6G).

338 Taken together, our data show the important role of *CHD4* and *PHF3* in the epigenetic
339 control of lymphoid lineage maintenance in *MLL/AF4*-positive leukaemia. In particular,
340 dysregulation of *CHD4*/NuRD is mediated by mutation, down-regulation of expression and
341 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 \geq 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 data³⁶. 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 cells³⁷, it can alternatively originate within the HSPC compartment. Indeed, the identification of *MLL/AF4*-expressing MPP-like cells shows that lineage switched relapse can

394 originate from very immature haemopoietic progenitor populations. This finding agrees with
395 recently published data pointing at MPP cells as the origin of *MLL/AF4* leukaemia³⁸ and is in
396 line with transcriptomic similarities between t(4;11) ALL and Lin-CD34+CD38-CD19- fetal
397 liver cells, again suggesting an HSPC as the cell of origin²³. Furthermore, the identification of
398 *MLL/AF4* within HSPC populations is consistent with the recent identification of an early
399 lymphoid progenitor, ELP-like signature specifically in *MLL*-rearranged ALL³⁹. Nevertheless,
400 and in agreement with previously published findings for MPALs⁶, the data derived from the
401 present cohort strongly support a non-lineage committed progenitor compartment as one
402 source for lineage switched relapse. However, we can not exclude additional cells-of-origin
403 of *MLL/AF4* ALL.

404 Irrespective of the cellular origin of the relapse, lineage switching is associated with a major
405 rewiring of gene regulatory networks. At the level of transcriptional control, the decision for
406 lymphoid development relies not only on the activation of a lymphoid transcriptional program,
407 but also on the silencing of a default myeloid program⁴⁰. That decision is enacted by
408 lymphoid master regulators including *EBF1*, *PAX5* and *IKAROS*, which represent genes
409 commonly mutated in precursor B-ALL and do not just upregulate B-cell specific genes, but
410 also repress the myeloid program⁴⁰⁻⁴⁴. *Pax5*^{-/-} pro-B cells which lack lymphoid potential,
411 whilst capable of erythro-myeloid differentiation *in vitro*, still maintain expression of early B
412 cell transcription factors *EBF1* and *E2A (TCF3)*⁴⁰. In contrast, we show that lineage
413 switching *MLL/AF4* pro-B leukaemic relapse is associated with a significant reduction in
414 expression of these earliest B lymphoid transcription factors, which links to changes in the
415 *MLL/AF4* transcriptional programme, ultimately establishing a myeloid differentiation fate.
416 Unfortunately, we were not able to directly prove changes in transcription factor binding and
417 associated changes in histone modifications due to the lack of available primary patient
418 material. However, high resolution DHS-seq clearly demonstrated changes in chromatin
419 accessibility and loss of occupation of the corresponding transcription factor binding sites.

420 The opposite scenario is observed when myeloid transcription factors are expressed in B-
421 lymphoid cells⁴⁵. Here, overexpression of C/EBP α efficiently reprograms such cells into
422 macrophages by suppressing lymphoid genes. *CEBPA* is strongly upregulated after *CHD4*
423 knockdown (Figure 6E) and is likely to be a driving force behind the lineage switch. Taken
424 together, these published and newly presented data confirm that (i) the balance between
425 lymphoid and myeloid transcription factors is instructive for lineage choice, and (ii) the down-
426 regulation of the myeloid program is essential for the maintenance of the lymphoid fate.

427 How can the mutation of global chromatin regulators cause a switch in cell fate? Similar to
428 the *Pax5* knockout, loss of IKAROS DNA-binding activity prevents lymphoid differentiation²⁶.
429 NuRD co-operates directly with IKAROS to repress HSC self-renewal and myeloid
430 differentiation, permitting early lymphoid development^{26,46,47}. Lineage switch was either
431 associated with heterozygous mutation, reduced expression or, in the case of two MPALs,
432 alternative splicing of *CHD4* and other NuRD components. These gene dosage effects are
433 consistent with reports showing that complete loss of *CHD4* impairs normal and leukaemic
434 proliferation^{48,49}, myeloid and lymphoid differentiation of HSPCs and causes exhaustion of
435 HSC pools⁴⁶, indicating that basal *CHD4* expression is required for maintaining AML.
436 Moreover, a partial inhibition of *CHD4* supported induction of pluripotency in iPSCs, while a
437 complete deletion eliminated cell proliferation, demonstrating that lowering *CHD4* expression
438 may facilitate lineage promiscuity⁵⁰.

439 Recent studies have identified core NuRD and PRC1 complex members as being direct
440 targets of *MLL/AF4* binding^{51,52}. Moreover, NuRD components including *CHD4* were shown
441 to be part of an *MLL* supercomplex⁵³. We therefore hypothesise that epigenetic regulator
442 genes are recruited by lineage specific factors during *MLL/AF4* leukemogenesis and mediate
443 fundamental lineage specific decision-making processes, in this case the repression of the
444 myeloid lineage program. Multiple routes to their dysregulation may result in escape from
445 this lineage restriction and may be enacted at different stages of haematopoiesis. However,
446 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 clone⁵⁴.

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 trogocytosis⁵⁷⁻⁵⁹. 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 relapse⁶⁰. 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 *MLLr* 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

462 **Acknowledgements**

463 We thank Jon Coxhead and Raf Hussain at the Newcastle University core genomics facility
464 as well as Marc van Tuil at the Princess Maxima Center Diagnostic department for
465 development of sequencing strategies. We acknowledge the Newcastle University Flow
466 Cytometry Core Facility and Tomasz Poplonski at the Princess Maxima Center Flow
467 Cytometry core facility for their assistance with the generation of flow cytometry data and cell
468 sorting strategies as well as the Newcastle University Bioinformatics Support Unit for helping
469 to develop the analysis approach for sequencing data. We thank Monique den Boer, Frank
470 van Leeuwen and Ronald Stam for critically reading the manuscript.

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

475 **Author contributions**

476 Conceptualization, O.H., S.B., C.B.; Methodology, O.H., C.B., R.T., K.S., P.M., S.B., A.P.,
477 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.,
478 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.,
479 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.,
480 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.,
481 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.,
482 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.,
483 C.B., R.T., K.S.; Supervision, O.H., S.B., J.M.A., J.V., C.B., ; Funding Acquisition, O.H., J.V.,
484 S.B., C.B., P.N.C., J.M.A., E.Z.

485 Data availability

486 Exome sequencing data and genome sequencing data presented in this manuscript have
487 been deposited in the NCBI Sequence Read Archive (SRA) under project numbers
488 PRJNA547947 and PRJNA547815 respectively. Immunoglobulin and TCR sequencing data
489 have been deposited in NCBI SRA under project number PRJNA511413. RNA sequencing
490 data and DNase hypersensitivity sequencing data were deposited in Gene Expression
491 Omnibus under accession numbers GSE132396 and GSE130142 respectively. All deposited
492 data will be publically available following publication of the manuscript. Requests for
493 additional specific data/materials should be made to Olaf Heidenreich
494 (O.T.Heidenreich@prinsesmaximacentrum.nl).

495 Conflict of interest statement

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

498 Financial support

499 This study was supported by a Cancer Research UK Centre Studentship (C27826/A17312)
500 and Newcastle University Overseas Research Scholarship to RT, a CRUK program grant to
501 JV and OH (C27943/A12788), a Kika programme grant to OH and JV (329), grants from the
502 North of England Children's Cancer Research Fund to OH, JV and SB, by Bloodwise grants
503 12055 and 15005 to OH and by a grant from the Kay Kendall Leukaemia Fund (KKL1142) to
504 OH. SB was supported by an NIHR Academic Clinical Lectureship (CL-2012-01-002), the Sir
505 Bobby Robson Foundation Clinical Fellowship and a Medical Research Council Clinician
506 Scientist Fellowship (MR/S021590/1). Work in CB/PNC's lab was funded by a programme
507 grant from Bloodwise (15001). Work in JMA's lab was funded by a programme grant from
508 Bloodwise (13044). EZ was supported by an RFBR grant (№17-29-06052). Research in the
509 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

513 **References**

- 514 1. Meyer C, Burmeister T, Groger D, et al. The MLL recombinome of acute leukemias in
515 2017. *Leukemia*. 2018;32(2):273-284.
- 516 2. Moorman AV, Ensor HM, Richards SM, et al. Prognostic effect of chromosomal
517 abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the
518 UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol*. 2010;11(5):429-
519 438.
- 520 3. Germano G, Pigazzi M, del Giudice L, et al. Two consecutive immunophenotypic
521 switches in a child with MLL-rearranged acute lymphoblastic leukemia. *Haematologica*.
522 2006;91(5 Suppl):ECR09.
- 523 4. Jiang JG, Roman E, Nandula SV, Murty VV, Bhagat G, Alobeid B. Congenital MLL-
524 positive B-cell acute lymphoblastic leukemia (B-ALL) switched lineage at relapse to acute
525 myelocytic leukemia (AML) with persistent t(4;11) and t(1;6) translocations and JH gene
526 rearrangement. *Leuk Lymphoma*. 2005;46(8):1223-1227.
- 527 5. Rossi JG, Bernasconi AR, Alonso CN, et al. Lineage switch in childhood acute
528 leukemia: an unusual event with poor outcome. *Am J Hematol*. 2012;87(9):890-897.
- 529 6. Alexander TB, Gu Z, Iacobucci I, et al. The genetic basis and cell of origin of mixed
530 phenotype acute leukaemia. *Nature*. 2018;562(7727):373-379.
- 531 7. Le Magnen C, Shen MM, Abate-Shen C. Lineage Plasticity in Cancer Progression
532 and Treatment. *Annu Rev Cancer Biol*. 2018;2:271-289.
- 533 8. Quintanal-Villalonga A, Chan JM, Yu HA, et al. Lineage plasticity in cancer: a shared
534 pathway of therapeutic resistance. *Nat Rev Clin Oncol*. 2020;17(6):360-371.
- 535 9. Chen L, Kostadima M, Martens JHA, et al. Transcriptional diversity during lineage
536 commitment of human blood progenitors. *Science*. 2014;345(6204):1251033.
- 537 10. Andersson AK, Ma J, Wang J, et al. The landscape of somatic mutations in infant
538 MLL-rearranged acute lymphoblastic leukemias. *Nat Genet*. 2015;47(4):330-337.
- 539 11. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected
540 transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-
541 309.
- 542 12. Zangrando A, Dell'orto MC, Te Kronnie G, Basso G. MLL rearrangements in pediatric
543 acute lymphoblastic and myeloblastic leukemias: MLL specific and lineage specific
544 signatures. *BMC Med Genomics*. 2009;2:36.
- 545 13. Gessner A, Thomas M, Castro PG, et al. Leukemic fusion genes MLL/AF4 and
546 AML1/MTG8 support leukemic self-renewal by controlling expression of the telomerase
547 subunit TERT. *Leukemia*. 2010;24(10):1751-1759.
- 548 14. Somervaille TC, Matheny CJ, Spencer GJ, et al. Hierarchical maintenance of MLL
549 myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather
550 than adult stem cells. *Cell Stem Cell*. 2009;4(2):129-140.
- 551 15. Wilkinson AC, Ballabio E, Geng H, et al. RUNX1 is a key target in t(4;11) leukemias
552 that contributes to gene activation through an AF4-MLL complex interaction. *Cell Rep*.
553 2013;3(1):116-127.
- 554 16. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel
555 method for the accurate identification of digital genomic footprints from DNase-seq data.
556 *Nucleic Acids Res*. 2013;41(21):e201.
- 557 17. Bonifer C, Cockerill PN. Chromatin Structure Profiling Identifies Crucial Regulators of
558 Tumor Maintenance. *Trends Cancer*. 2015;1(3):157-160.
- 559 18. Kreher S, Bouhlel MA, Cauchy P, et al. Mapping of transcription factor motifs in
560 active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma. *Proc Natl
561 Acad Sci U S A*. 2014;111(42):E4513-4522.
- 562 19. Howell ED, Yzaguirre AD, Gao P, et al. Efficient hemogenic endothelial cell
563 specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-
564 regulated TGFbeta target genes. *Genes Dev*. 2021;35(21-22):1475-1489.

- 565 20. Hohaus S, Petrovick MS, Voso MT, Sun Z, Zhang DE, Tenen DG. PU.1 (Spi-1) and
566 C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor
567 receptor alpha gene. *Mol Cell Biol.* 1995;15(10):5830-5845.
- 568 21. Leddin M, Perrod C, Hoogenkamp M, et al. Two distinct auto-regulatory loops
569 operate at the PU.1 locus in B cells and myeloid cells. *Blood.* 2011;117(10):2827-2838.
- 570 22. Dobbins SE, Sherborne AL, Ma YP, et al. The silent mutational landscape of infant
571 MLL-AF4 pro-B acute lymphoblastic leukemia. *Genes Chromosomes Cancer.*
572 2013;52(10):954-960.
- 573 23. Agraz-Doblas A, Bueno C, Bashford-Rogers R, et al. Unraveling the cellular origin
574 and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-
575 wide analysis. *Haematologica.* 2019;104(6):1176-1188.
- 576 24. Arends T, Dege C, Bortnick A, et al. CHD4 is essential for transcriptional repression
577 and lineage progression in B lymphopoiesis. *Proc Natl Acad Sci U S A.* 2019;116(22):10927-
578 10936.
- 579 25. Ng SY, Yoshida T, Zhang J, Georgopoulos K. Genome-wide lineage-specific
580 transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic
581 stem cells. *Immunity.* 2009;30(4):493-507.
- 582 26. Zhang J, Jackson AF, Naito T, et al. Harnessing of the nucleosome-remodeling-
583 deacetylase complex controls lymphocyte development and prevents leukemogenesis. *Nat*
584 *Immunol.* 2011;13(1):86-94.
- 585 27. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699
586 paediatric leukaemias and solid tumours. *Nature.* 2018;555(7696):371-376.
- 587 28. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of
588 nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet.*
589 2011;88(4):440-449.
- 590 29. Sifrim A, Hitz MP, Wilsdon A, et al. Distinct genetic architectures for syndromic and
591 nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet.*
592 2016;48(9):1060-1065.
- 593 30. Novillo A, Fernandez-Santander A, Gaibar M, et al. Role of Chromodomain-Helicase-
594 DNA-Binding Protein 4 (CHD4) in Breast Cancer. *Front Oncol.* 2021;11:633233.
- 595 31. Appel LM, Franke V, Bruno M, et al. PHF3 regulates neuronal gene expression
596 through the Pol II CTD reader domain SPOC. *Nat Commun.* 2021;12(1):6078.
- 597 32. Kang H, Chen IM, Wilson CS, et al. Gene expression classifiers for relapse-free
598 survival and minimal residual disease improve risk classification and outcome prediction in
599 pediatric B-precursor acute lymphoblastic leukemia. *Blood.* 2010;115(7):1394-1405.
- 600 33. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell
601 gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA.*
602 2010;304(24):2706-2715.
- 603 34. Ferreira BI, Garcia JF, Suela J, et al. Comparative genome profiling across subtypes
604 of low-grade B-cell lymphoma identifies type-specific and common aberrations that target
605 genes with a role in B-cell neoplasia. *Haematologica.* 2008;93(5):670-679.
- 606 35. Lin S, Luo RT, Ptasinska A, et al. Instructive Role of MLL-Fusion Proteins Revealed
607 by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell.* 2016;30(5):737-
608 749.
- 609 36. Bolotin DA, Poslavsky S, Davydov AN, et al. Antigen receptor repertoire profiling
610 from RNA-seq data. *Nat Biotechnol.* 2017;35(10):908-911.
- 611 37. le Viseur C, Hotfilder M, Bomken S, et al. In childhood acute lymphoblastic leukemia,
612 blasts at different stages of immunophenotypic maturation have stem cell properties. *Cancer*
613 *Cell.* 2008;14(1):47-58.
- 614 38. Malouf C, Ottersbach K. The fetal liver lymphoid-primed multipotent progenitor
615 provides the prerequisites for the initiation of t(4;11) MLL-AF4 infant leukemia.
616 *Haematologica.* 2018;103(12):e571-e574.
- 617 39. Khabirova E, Jardine L, Coorens THH, et al. Single-cell transcriptomics reveals a
618 distinct developmental state of KMT2A-rearranged infant B-cell acute lymphoblastic
619 leukemia. *Nat Med.* 2022;28(4):743-751.

- 620 40. Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage
621 depends on the transcription factor Pax5. *Nature*. 1999;401(6753):556-562.
- 622 41. Mullighan CG, Kennedy A, Zhou X, et al. Pediatric acute myeloid leukemia with
623 NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene
624 expression distinct from MLL-rearranged leukemias. *Leukemia*. 2007;21(9):2000-2009.
- 625 42. Boer JM, van der Veer A, Rizopoulos D, et al. Prognostic value of rare IKZF1
626 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international
627 collaborative study. *Leukemia*. 2016;30(1):32-38.
- 628 43. Witkowski MT, Hu Y, Roberts KG, et al. Conserved IKAROS-regulated genes
629 associated with B-progenitor acute lymphoblastic leukemia outcome. *J Exp Med*.
630 2017;214(3):773-791.
- 631 44. Pongubala JM, Northrup DL, Lancki DW, et al. Transcription factor EBF restricts
632 alternative lineage options and promotes B cell fate commitment independently of Pax5. *Nat*
633 *Immunol*. 2008;9(2):203-215.
- 634 45. van Oevelen C, Collombet S, Vicent G, et al. C/EBPalpha Activates Pre-existing and
635 De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and
636 Myelopoiesis. *Stem Cell Reports*. 2015;5(2):232-247.
- 637 46. Yoshida T, Hazan I, Zhang J, et al. The role of the chromatin remodeler Mi-2beta in
638 hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev*.
639 2008;22(9):1174-1189.
- 640 47. Lu X, Chu CS, Fang T, et al. MTA2/NuRD Regulates B Cell Development and
641 Cooperates with OCA-B in Controlling the Pre-B to Immature B Cell Transition. *Cell Rep*.
642 2019;28(2):472-485 e475.
- 643 48. Sperlazza J, Rahmani M, Beckta J, et al. Depletion of the chromatin remodeler CHD4
644 sensitizes AML blasts to genotoxic agents and reduces tumor formation. *Blood*.
645 2015;126(12):1462-1472.
- 646 49. Heshmati Y, Turkoz G, Harisankar A, et al. The chromatin-remodeling factor CHD4 is
647 required for maintenance of childhood acute myeloid leukemia. *Haematologica*.
648 2018;103(7):1169-1181.
- 649 50. Mor N, Rais Y, Sheban D, et al. Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex
650 Facilitates Deterministic Induction of Naive Pluripotency. *Cell Stem Cell*. 2018;23(3):412-425
651 e410.
- 652 51. Kerry J, Godfrey L, Repapi E, et al. MLL-AF4 Spreading Identifies Binding Sites that
653 Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in
654 Leukemia. *Cell Rep*. 2017;18(2):482-495.
- 655 52. Harman JR, Thorne R, Jamilly M, et al. A KMT2A-AFF1 gene regulatory network
656 highlights the role of core transcription factors and reveals the regulatory logic of key
657 downstream target genes. *Genome Res*. 2021.
- 658 53. Nakamura T, Mori T, Tada S, et al. ALL-1 is a histone methyltransferase that
659 assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell*.
660 2002;10(5):1119-1128.
- 661 54. Jacoby E, Nguyen SM, Fontaine TJ, et al. CD19 CAR immune pressure induces B-
662 precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic
663 plasticity. *Nat Commun*. 2016;7:12320.
- 664 55. Gardner R, Wu D, Cherian S, et al. Acquisition of a CD19-negative myeloid
665 phenotype allows immune escape of MLL-rearranged B-ALL from CD19 CAR-T-cell therapy.
666 *Blood*. 2016;127(20):2406-2410.
- 667 56. Rayes A, McMasters RL, O'Brien MM. Lineage Switch in MLL-Rearranged Infant
668 Leukemia Following CD19-Directed Therapy. *Pediatr Blood Cancer*. 2016;63(6):1113-1115.
- 669 57. Sotillo E, Barrett DM, Black KL, et al. Convergence of Acquired Mutations and
670 Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy. *Cancer*
671 *Discov*. 2015;5(12):1282-1295.
- 672 58. Rabilloud T, Potier D, Pankaew S, Nozais M, Loosveld M, Payet-Bornet D. Single-
673 cell profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-
674 negative relapse after CAR-T therapy. *Nat Commun*. 2021;12(1):865.

- 675 59. Hamieh M, Dobrin A, Cabriolu A, et al. CAR T cell trogocytosis and cooperative
676 killing regulate tumour antigen escape. *Nature*. 2019;568(7750):112-116.
- 677 60. Liao W, Kohler ME, Fry T, Ernst P. Does lineage plasticity enable escape from CAR-
678 T cell therapy? Lessons from MLL-r leukemia. *Exp Hematol*. 2021;100:1-11.
- 679 61. Lambie AJ, Myers RM, Taraseviciute A, et al. KMT2A Rearrangements Are
680 Associated with Lineage Switch Following CD19 Targeting CAR T-Cell Therapy. *Blood*.
681 2021;138(Supplement 1):256-256.
- 682 62. Leahy AB, Devine KJ, Li Y, et al. Impact of high-risk cytogenetics on outcomes for
683 children and young adults receiving CD19-directed CAR T-cell therapy. *Blood*.
684 2022;139(14):2173-2185.

685

686

687

688

689

690

691

692 **Figure Legends**

693 **Figure 1. Characterisation of *MLL/AF4* lineage switch cases.** (A) Morphological change
694 from lymphoblastic leukaemia (left panel) to acute monoblastic/monocytic leukaemia (right
695 panel) in patient LS01. The scale bar represents 20 μm . (B) Sanger sequencing of *MLL/AF4*
696 and reciprocal *AF4/MLL* fusions in LS01 presentation ALL (upper panel) and relapse AML
697 (lower panel) identifies a common breakpoint with identical filler sequence in ALL and AML
698 samples.

699 **Figure 2. Transcriptional reprogramming in lineage switch and MPAL cases.**

700 (A) Heatmap showing the top 100 differentially expressed genes between ALL and AML
701 from six lineage switch (LS01, LS03, LS04, LS05, LS06, LS10) and two MPAL cases,
702 ranked by Wald statistics. (B) Enrichment of myeloid growth and differentiation signature in
703 relapsed samples (left panel) identified by GSEA analyses, also pointing to downregulation
704 of genes highly correlated with acute lymphoblastic leukemia (middle and right panel). Gene
705 set enrichment analyses have been performed based on data derived from six lineage
706 switch samples. FDR – false discovery rate, NES – normalised enrichment score. (C)
707 Differential expression of lineage specific and (D) immunoglobulin recombination machinery
708 genes in lineage switch and MPAL cases. Error bars show standard error of the mean (SEM)
709 for lineage switch cases and ranges for two MPAL cases.

710 **Figure 3. Alternative splicing in lineage switch and MPAL cases.**

711 (A) Pie charts showing the classification of non-differential (non-DEEj) and differential (DEEj) exon-exon junctions.
712 Shown are the percentages of splicing events assigned to a particular mode of splicing.
713 Complex splicing event corresponds to several (two or more) alternative splicing incidents
714 occurring simultaneously in the same sample. (B) Volcano plots demonstrating differential
715 usage of exon-exon junctions in the transcriptome of AML/myeloid versus ALL/lymphoid
716 cells of lineage switch (LS01, LS03 & LS04) or MPAL patients. The vertical dashed lines
717 represent two-fold differences between the AML and ALL cells, and the horizontal dashed
718 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

746 unique to AML relapse as compared to ALL relapse as shown in (D). (F) EBF1 and C/EBP
747 binding motifs demonstrate differential motif density in presentation ALL and relapse AML.

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

760 **Figure 6. Epigenetic modulatory genes influence lineage specific expression profiles.**
761 (A) Intersection between identified hits of clonal mutations (VAF>30%), differentially
762 expressed genes and alternatively spliced, differentially used exon-exon junctions (adj.p-
763 value<0.01) in lineage switched myeloid relapse/myeloid fraction of MPALs, present in the
764 analysed cohort. (B) Fold change in expression of NuRD complex members (*CHD4*, *MTA1*,
765 *RBBP4*, *MBD3*) and *PHF3* following lineage switched relapse (left panel) and in MPAL cases
766 (right panel). (C) CHD4 structure; the R1068H mutation (red) is located in the critical
767 helicase domain of CHD4 at a highly conserved residue. An * (asterisk) indicates positions
768 which have a single, fully conserved residue, a : (colon) indicates conservation between
769 groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, a .
770 (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5
771 in the Gonnet PAM 250 matrix. (D) Identification of regions of differential chromatin
772 accessibility before and after knockdown of *CHD4* and *PHF3* depicted in red in *MLLr* SEM

773 cells (left panel) and non-*MLLr* 697 cells (right panel). For all reads the fold change in ATAC-
774 peak height was calculated relative to the control (shNTC) and ATAC-peaks from knock-
775 down cells were plotted according to their fold-change along-side the control signals. *CHD4*
776 ChIP density plots from SEM cells (depicted in blue) were plotted alongside the
777 corresponding DNA regions of the shNTC control. Differentially expressed genes associated
778 with changing ATAC peaks (log₂FC analysed vs shNTC) identified in each cellular variant
779 are represented by heatmaps included at the right side of each panel (for SEM and 697
780 cells). (E) UCSC genome browser screenshots representing differential chromatin
781 accessibility (ATAC-seq) and gene expression level (RNA-seq) in the myeloid *CEBPA* and
782 the lymphoid *RAG2* loci following *CHD4* and *PHF3* knockdown in *MLLr* SEM cells and non-
783 *MLLr* 697 cells. ChIP-seq traces representing normal *CHD4* occupancy in non-*MLLr* B-ALL
784 (REH cells), *MLLr* B-ALL (SEM cells) and *MLLr* AML cells (MV-4;11) are shown as a
785 reference at the bottom of each screenshot. TSS – transcriptional start site is depicted for
786 each gene. (F) Expression of the lineage specific cell surface markers CD19 (lymphoid) and
787 CD33 (myeloid) following culture of *MLL/Af4* transformed hCD34+ cord blood progenitor
788 cells in lymphoid permissive conditions. Knockdown of *PHF3*, *CHD4* or the combination
789 disrupts the dominant lymphoid differentiation pattern seen in non-targeting control (shNTC).
790 (G) *PHF3* knockdown is capable of influencing the surface marker expression after longer
791 incubation period (33 days); *CHD4* knockdown impaired cellular survival upon longer *in vitro*
792 culture (data not shown).

793 **Figure 7. Haematopoietic stem/progenitor populations carry *MLL/AF4*.** (A) Summary of
794 *MLL/AF4* positivity and 12 SNVs exclusive for the AML relapse, within different populations
795 analysed in patient LS01RAML Circles with solid colour indicate VAF >30%, light colour and
796 dashed line indicates VAF 5-30%. Remaining genes (yellow circle) represent the 10 other
797 SNVs (out of 12 SNVs) which showed the same pattern in the frequency of mutation
798 acquisition (described in Table S10). (B) Summary of the proposed model of the origin of
799 lineage switched relapse. Evaluation of B-cell receptor repertoires on ALL (presentation) and

800 AML (relapse) lineage switch, and MPAL cases identified three different patterns. Pattern 1 -
801 with clonotypes on the ALL only. Pattern 2 - B-cell receptor-containing clones on ALL and
802 AML, but distinct to each other. Pattern 3 - B-cell receptor-containing clones shared between
803 ALL and AML. (C) BCR clones frequencies identified in whole-exome seq data with
804 application of MiXCR software in all analysed LSAL and MPAL patients.

805

806

Figure 1

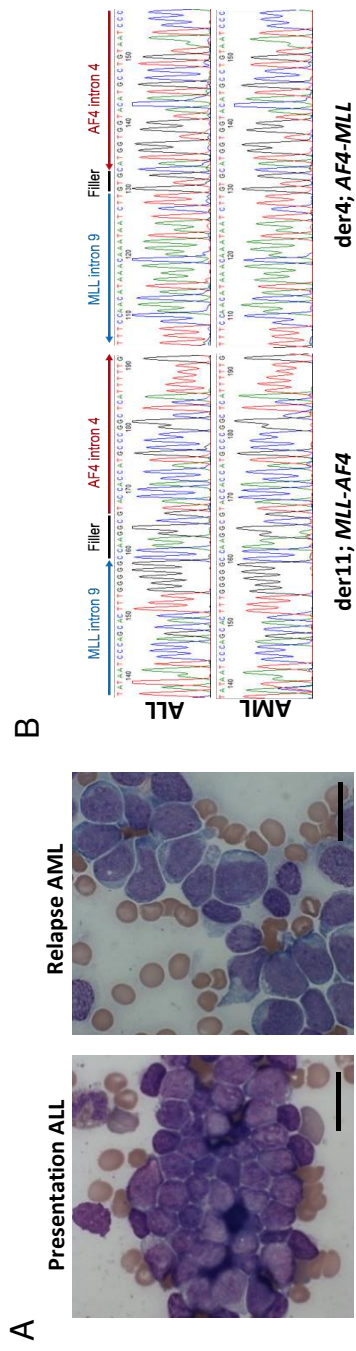


Figure 2

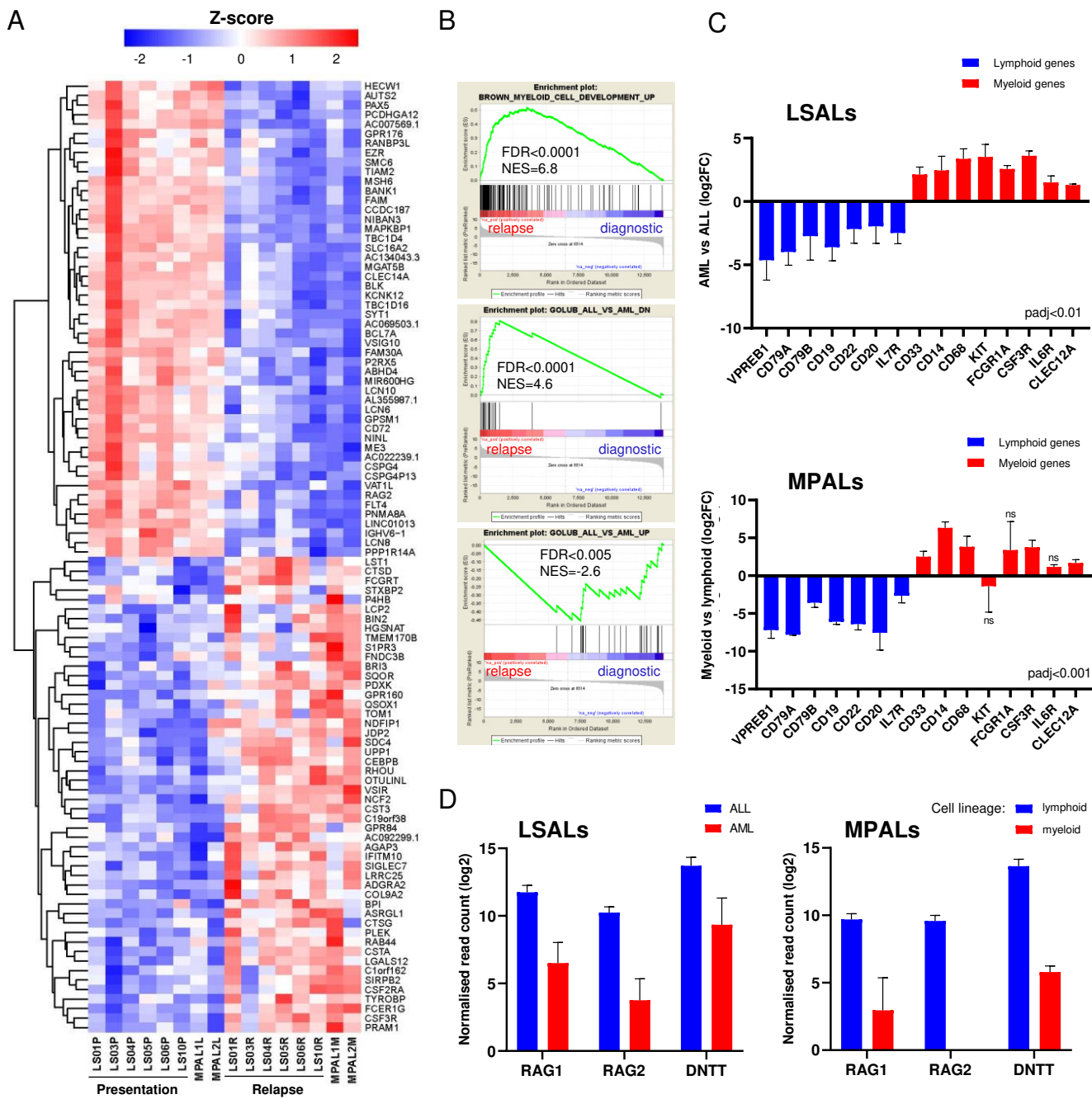


Figure 3

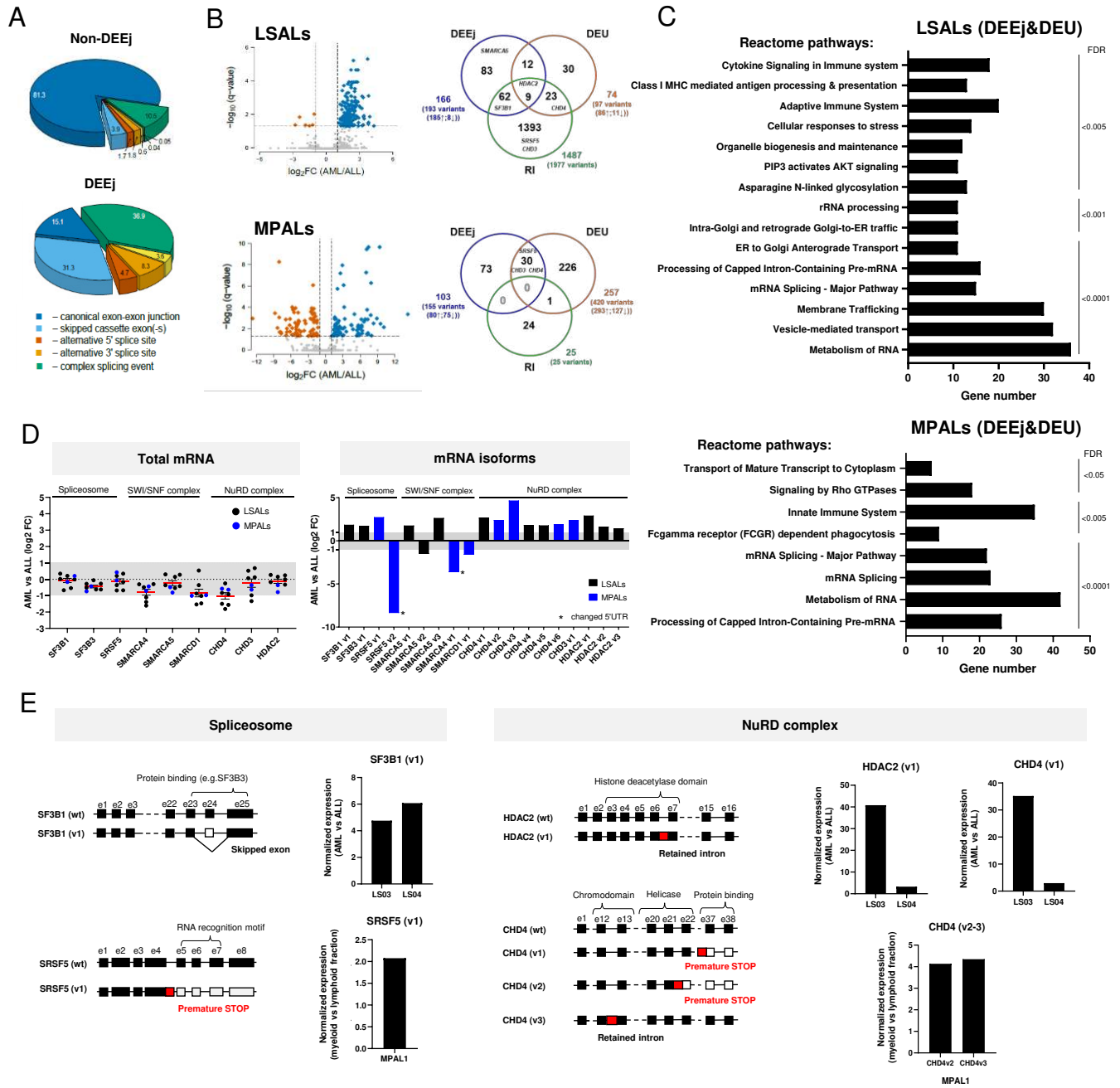


Figure 4

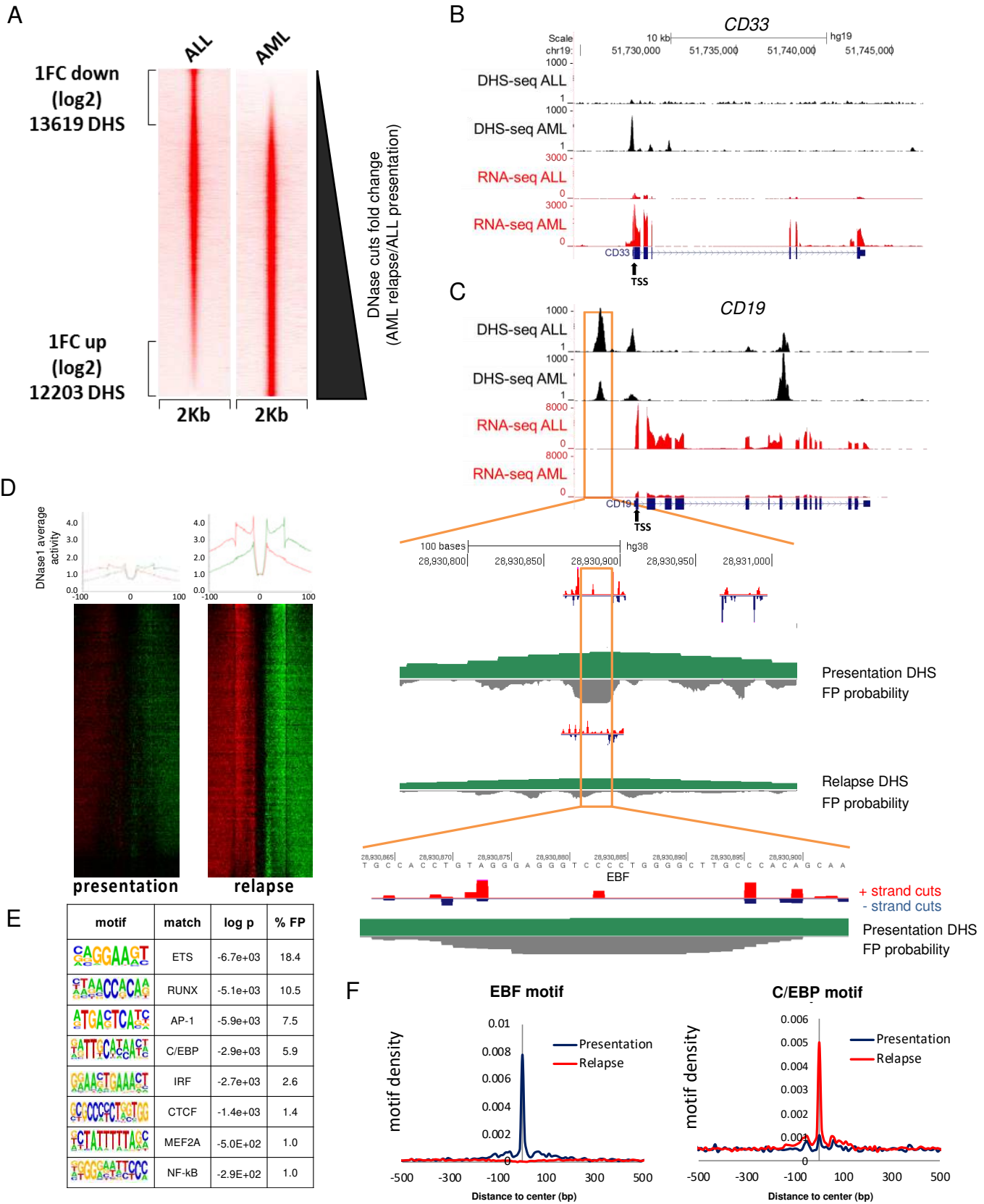


Figure 5

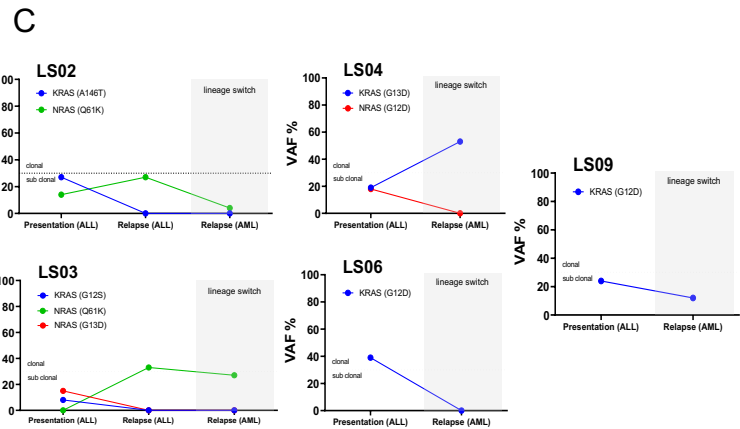
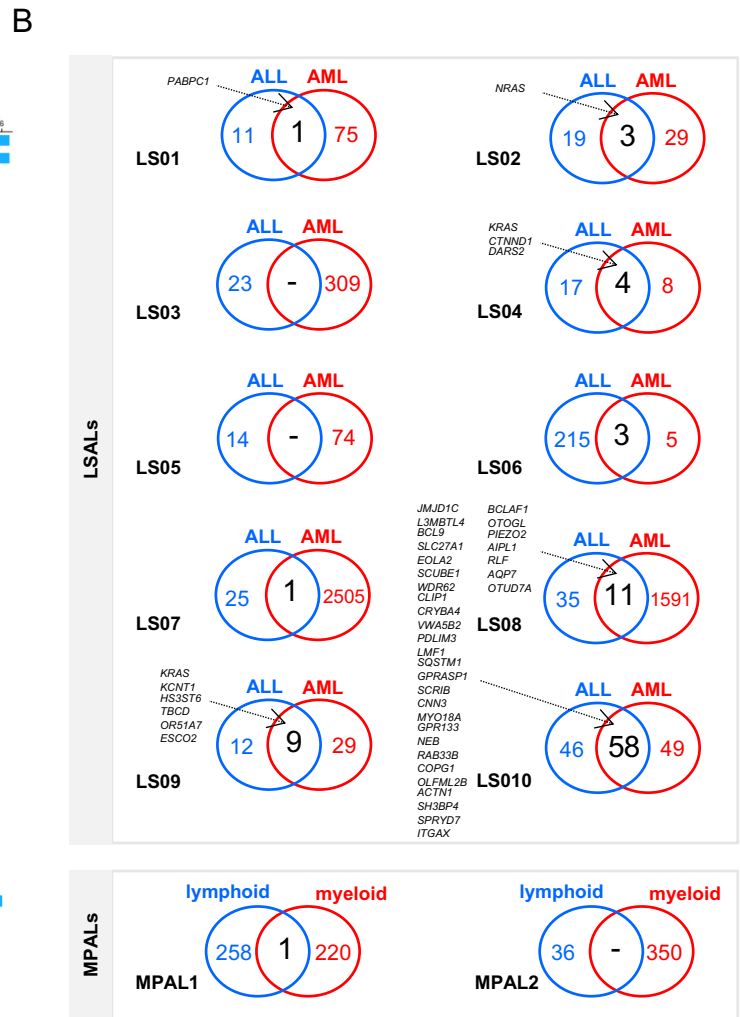
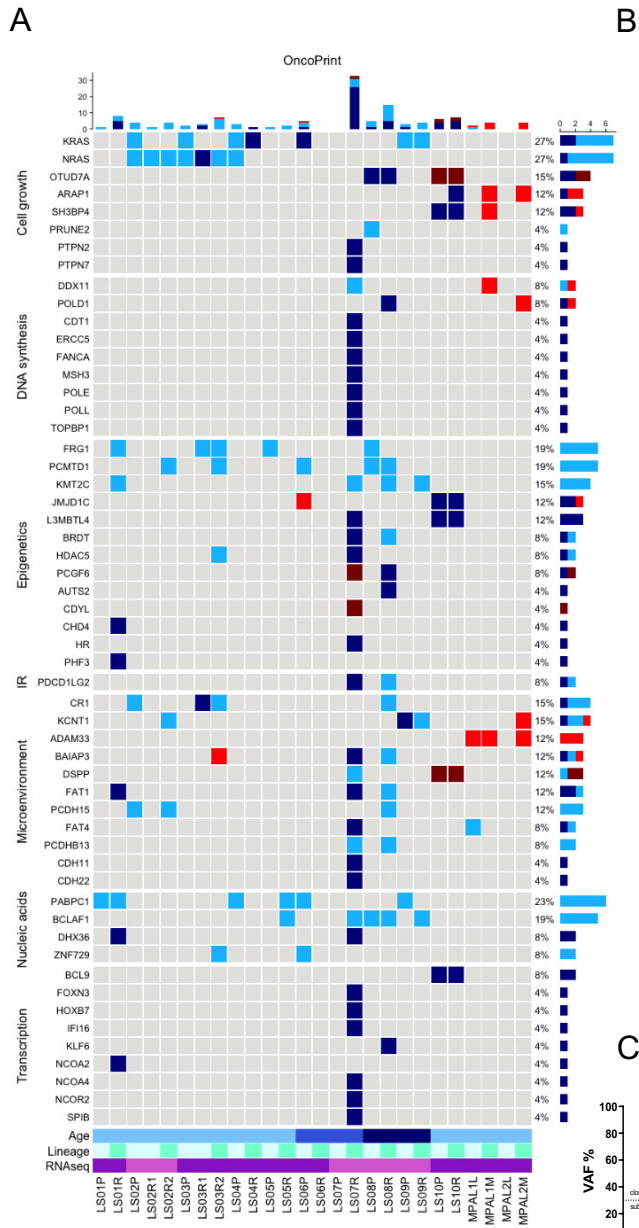


Figure 6

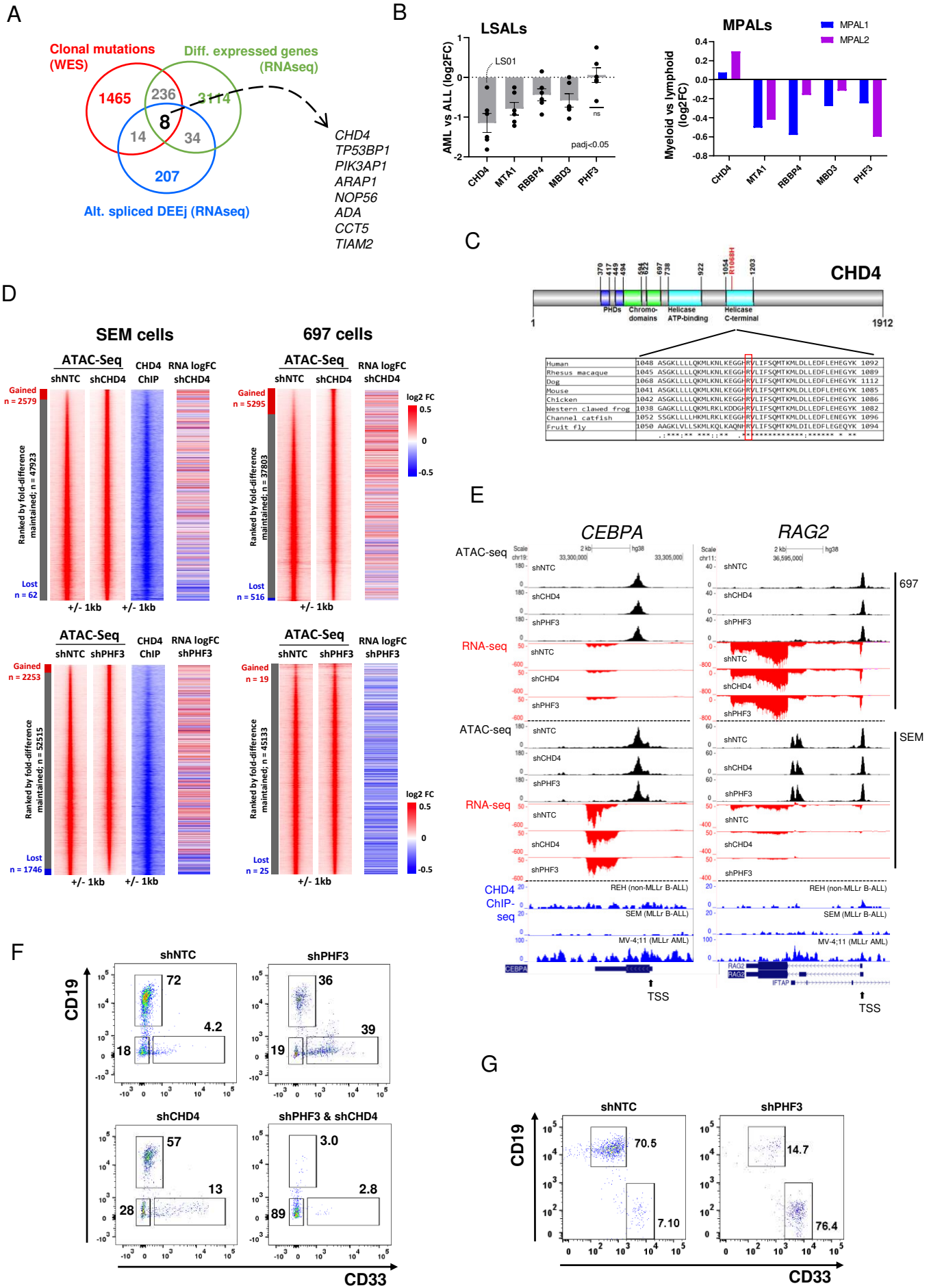


Figure 7

