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

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

Author Affiliations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Summary

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

Introduction

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

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

91 **Results**

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

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

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

113 **Cellular origin of lineage switched relapse**

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

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

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

143 **Lineage switch relapse can originate in HSPC compartments**

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

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

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

168 We next investigated the underlying molecular events associated with a lineage switch. To
169 this end we analysed differential gene expression across eight cases, including six LSALs for
170 which RNA was available at both presentation and relapse, and sorted lymphoid and myeloid
171 blast populations of two MPALs. This analysis identified in total 1374 up- (adj. p<0.01, Log
172 Fold change >2) and 1323 down-regulated genes in the AML lineage switches and the myeloid
173 populations of MPAL patients (Table S4). The most substantially down-regulated genes
174 include lymphoid genes such as *PAX5* and *RAG2*, but also the polycomb PRC1 like complex
175 component *AUTS2* and SWI/SNF complex component *BCL7A,* while up-regulated genes
176 comprise several myeloid genes such as cathepsins, cystatins, *PRAM1* and *CSF3R* (Figure
177 4A). These fingings were consistent irrespective of which cellular origin of relapse the BCR
178 rearrangement analysis supported (pattern 1, 2 or 3),

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

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

194 Somervaille *et al.*, 2009; Gessner *et al.*, 2010; Wilkinson *et al.*, 2013; Kerry *et al.*, 2017). For
195 instance, significant changes in gene expression across the *HOXA* cluster, notably with a 50-
196 fold reduction in *HOXA7* expression, represent a major additional disruption of *MLL*r
197 leukaemogenic transcriptional regulation (Figures S5A-C) (Ayton and Cleary, 2003;
198 Somervaille *et al.*, 2009; Gessner *et al.*, 2010). Furthermore, a very significant portion of target
199 genes including *PROM1*, *IKZF2* and chromatin modifying factors are highly enriched amongst
200 genes that show lower expression in myeloid lineages (Figures S5D, E, Table S4). In total,
201 996 out of 5208 bona fide direct target genes of *MLL-AF4* changed expression in the AML
202 relapse (Wilkinson *et al.*, 2013). These data suggest that the process of lineage switch is
203 associated with a major reorganisation of the *MLL-AF4* transcriptional network and pose the
204 question of which epigenetic regulators are involved in the lineage determination associated
205 with this fusion gene.

206 **Reorganisation of chromatin accessibility and transcription factor binding upon**
207 **lineage switch**

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

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

220 of occupancy of motifs bound by C/EBP, IRF and NF-κB family members (Figure 5E-F). The
221 gain in C/EBP motif binding was associated with the expression of a C/EBPA regulated
222 myeloid transcriptional gene set (Figure 5G). We also observed a redistribution of occupancy
223 of transcription factors controlling both lymphoid and myeloid maturation such as RUNX, AP-
224 1 and ETS members to alternative cognate motifs (Figures 5E and S6B) (Hohaus *et al.*, 1995;
225 Nerlov and Graf, 1998; Leddin *et al.*, 2011). This finding is exemplified by decreased
226 accessibility of a region located 1 kb upstream of the *CD19* TSS with concomitant loss of EBF
227 binding at this element (Figure 5C). Differential motif enrichments were associated with
228 changes in RNA expression and chromatin accessibility at the genes encoding the
229 corresponding cognate transcription factors *EBF1*, *PAX5*, *LEF1* (B lymphoid determinants),
230 *NFKB2* and *CEBPB/D/E* (myeloid determinants), particularly in regions flanking TSSs (Figures
231 5H, S6B).

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

235 **Lineage switch changes alternative splicing patterns**

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

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

258 Pathway analysis revealed an enrichment of alternatively spliced genes in immune pathways
259 including antigen processing, membrane trafficking and FCGR-dependent phagocytosis
260 reflecting the change from a lymphoid to a myeloid state (Figures 6C, S7E). Furthermore, it
261 highlighted RNA processing and maturation including mRNA splicing, processing of capped
262 intron-containing pre-mRNAs and rRNA processing. Indeed, myeloid populations expressed
263 4-6-fold higher alternatively spliced *SF3B1* and *SRSF5* levels than their matched lymphoid
264 populations (Figure 6D, E). In addition, we noted a significant number of genes encoding
265 epigenetic modulators including *KDM5C, HDAC2* and several *CHD* members being
266 differentially spliced in AML relapse or myeloid subpopulations of MPALs (Figure 6D, E).

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

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

277 **The mutational landscape of lineage switch**

278 Next, we examined the mutational landscape of lineage switched *MLL-AF4* leukaemias by
279 performing exome sequencing on the entire cohort. In keeping with published data on newly
280 presenting *MLL*r acute leukaemias (Dobbins *et al.*, 2013; Andersson *et al.*, 2015), exome
281 sequencing of presentation ALL samples confirmed a relatively quiet mutational landscape in
282 infant ALL cases, with median of 13 nonsynonymous somatic single nucleotide variants
283 (SNVs) or insertions/deletions (indels) predicted to be deleterious to protein function (Table
284 S7). Many of these were present in less than 30% of reads and considered sub-clonal. The
285 most commonly mutated genes at presentation were *NRAS* (3 cases) and *KRAS* (5 cases)
286 (Figure 7A) as described previously (Andersson *et al.*, 2015). In contrast, relapse AML
287 samples contained a median of 46 deleterious somatic SNVs and indels (Table S7). This
288 increase can be mainly attributed to three samples (LS03RAML, LS07RAML and LS08RAML)
289 that carried deleterious mutations in DNA polymerase genes in the respective major clones
290 linked to hypermutator phenotypes (Church *et al.*, 2013; Erson-Omay *et al.*, 2015; Mur *et al.*,
291 2020) (Figure 7A). However, we observed this phenotype only in three out of ten relapses,
292 arguing against this phenomenon being a general requirement for the lineage switch in
293 relapse. Similarly, many of the predominantly subclonal mutations identified in presentation
294 ALL samples, including half of *RAS* mutations, were subsequently lost at relapse, indicating
295 alternative subclones as the origin of relapse (Figure 7A). Finally, both MPALs harboured
296 several mutations that were exclusively found in either the lymphoid or myeloid subpopulation
297 indicating the presence of subclones with a lymphoid and myeloid bias (Figure 7A).

298 Next, we examined the global mutation patterns of patient LS01 in greater detail and focussed
299 on SNVs with a VAF of ≥0.3. ALL and AML contained 104 and 3196 SNVs above this
300 threshold, respectively, with only 22 shared SNVs between both samples. The most prevalent
301 type of SNV was the C to T transition in the DNA of both ALL and AML samples (Figure 7B).
302 However, the contribution of underlying single base substitution (SBS) signatures differed
303 between diagnosis and relapse. Three different signatures (SBS16, 5 and 1) explained about
304 60% of the SNPs found in ALL, while SBS1 seemed to explain more than 50% of all SNPs in
305 the AML (Figure 7B) suggesting a mutational clock as the main driver of the evolution of
306 relapse (Alexandrov et al., 2015; Blokzijl et al., 2016). Further inspection of the pattern
307 revealed a mutational signature mainly comprising C to T transitions and to a lesser degree C
308 to G transversions in NCG triplets, raising the possibility that thiopurine maintenance treatment
309 may have increased the mutational burden, resulting in lineage switched relapse in this patient
310 (Li *et al.*, 2020).

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

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

327 The two earliest mutations within LS01 relapse were identified in the Nucleosome Remodelling
328 and Deacetylation complex (NuRD) member *CHD4* and the plant homeodomain finger
329 containing *PHF3*. PHF3 is a member of a family of transcriptional regulators that have been
330 suggested to link the deposition of histone marks to the regulation of transcription (Kinkelin *et*
331 *al.*, 2013). PHF3 itself has been recently identified as an inhibitor of transcription elongation
332 by competing with TFIIS for binding to the C-terminal domain of RNA polymerase II (Appel *et*
333 *al.*, 2020). NuRD is a multiprotein transcriptional co-repressor complex with both histone
334 deacetylase and ATP-dependent chromatin remodelling activity. It is a critical factor in the
335 lymphoid lineage determination in part directed by the transcription factor IKZF1 (Kim *et al.*,
336 1999; Yoshida *et al.*, 2006; Ng *et al.*, 2009; Zhang *et al.*, 2011; Arends *et al.*, 2019). Both the
337 CHD4 R1068H and PHF3 K1119I mutations affect highly conserved residues (Figures 8A,
338 S8A) that are predicted by the Condel classifier (González-Pérez and López-Bigas, 2011) to
339 disrupt protein function. Specifically, the CHD4 R1068H mutation has previously been linked
340 to defects in cardiac development (Sifrim *et al.*, 2016).

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

348 The relevance of *CHD4* and *PHF3* in regulating lymphoid versus myeloid lineage choice was
349 further supported by ARACNE analysis using published ALL (Harvey *et al.*, 2010; Kang *et al.*,
350 2010) and AML (Gentles *et al.*, 2010) expression datasets (n=216) which we used to reverse-
351 engineer a mutual information network (Margolin *et al.*, 2006). This network was trimmed to

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

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

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

379 2015). *AUTS2* is both mutated in LS08RAML and highly expressed in all lymphoid populations
380 examined (Figures 4A, 7A). STRING network (Szklarczyk *et al.*, 2019) analysis demonstrated
381 close functional associations within PRC1 and NuRD complexes and their shared associations
382 (Figure S8E). Mutual interactions include CBX2, a PRC1 complex member, which shows a
383 similar expression pattern between lineage switch and MPAL cases (Figure S8F). While
384 knockdown of *AUTS2* did not change CD33 levels, depletion of *PCGF6* also increased CD33
385 surface expression in SEM cells, further supporting the notion of epigenetic factors regulating
386 lineage determination in ALL (Figures S8G). Furthermore, GSEA also indicated impaired
387 function of PRC1 and PRC2 complexes in the AML relapse compared with the presentation
388 ALL with down-regulation of their respective target genes (Figure S8H).

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

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

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

407 Discussion

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

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

431 within *HOXA* cluster genes (Ayton and Cleary, 2003; Somervaille *et al.*, 2009; Gessner *et al.*,
432 2010)., which likely results in a wider reorganisation of malignant haematopoietic
433 transcriptional networks, ultimately leading to a myeloid differentiation fate.

434 Similar to the *Pax5* knockout (Nutt *et al.*, 1999), loss of IKAROS DNA-binding activity prevents
435 lymphoid differentiation (Zhang *et al.*, 2011). NuRD co-operates directly with IKAROS to
436 repress HSC self-renewal and subsequent myeloid differentiation, permitting early lymphoid
437 development (O'Neill *et al.*, 2000; Williams *et al.*, 2004; Yoshida *et al.*, 2008; Zhang *et al.*,
438 2011; Hosokawa *et al.*, 2013; Lu *et al.*, 2019). We found that the abrogation of this pathway
439 through multiple mechanisms was central to the lineage switch from ALL to AML. Lineage
440 switch was either associated with mutation, reduced expression or, in the case of two MPALs,
441 alternative splicing of *CHD4* and other NuRD components. Long term knockdown of *CHD4*
442 was not tolerated in our cord blood culture. This is in line with reports showing that complete
443 loss of *CHD4* impairs leukaemic proliferation (Sperlazza *et al.*, 2015; Heshmati *et al.*, 2018),
444 both myeloid and lymphoid differentiation of HSPCs and causes exhaustion of HSC pools
445 (Yoshida *et al.*, 2008), indicating that basal CHD4 expression is required for maintaining AML.
446 Moreover, our observation that a 60% *CHD4* knockdown is associated with the activation of
447 pluripotency gene signatures is in line with the finding that a partial inhibition of *CHD4*
448 supported induction of pluripotency in iPSCs, while a complete deletion eliminated cell
449 proliferation (Mor *et al.*, 2018).

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

458 however, the associated loss of B cell surface markers (*e.g.*, CD19) provides an alternative
459 mechanism for relapse following CAR-T cell or blinatumomab therapy (Gardner *et al*., 2016;
460 Rayes, McMasters and O'Brien, 2016) in addition to mutations, alternative splicing (Sotillo *et*
461 *al.*, 2015; Orlando *et al.*, 2018; Rabilloud *et al.*, 2021) and T cells trogocytosis (Hamieh *et al.*,
462 2019). Whilst these therapies target lineage specific surface markers, lineage-switched (pre-
463)leukaemic progenitor populations escape epitope recognition and provide a potential clonal
464 source for the relapse. Given the increasing use of advanced immunological therapies, a
465 detailed understanding of the molecular processes underlying lineage determination and
466 switching will be critical for developing new strategies to avoid this route to clinical relapse.

467 **Acknowledgements**

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

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

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

493 **Author contributions**

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

502 **Declaration of interests**

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

505 **Figure titles and legends**

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

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

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

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

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

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

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

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

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

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

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

617 **Figure 8. Epigenetic modulatory genes influence lineage specific expression profiles.**
618 (A) CHD4 scheme; the R1068H mutation is located in the critical helicase domain of CHD4 at
619 a highly conserved residue. An * (asterisk) indicates positions which have a single, fully
620 conserved residue, a : (colon) indicates conservation between groups of strongly similar
621 properties - scoring > 0.5 in the Gonnet PAM 250 matrix, a . (period) indicates conservation
622 between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.
623 (B) Fold change in expression of NuRD complex members (*CHD4*, *MTA1*, *RBBP4*, *MBD3*)
624 and *PHF3* following lineage switched relapse (left panel) and in MPAL cases (right panel). (C)
625 Expression of lineage specific genes following knockdown of *PHF3*, *CHD4* or the combination,
626 relative to non-targeting control construct in the *MLL-AF4* positive ALL cell line, SEM. (D) Flow
627 cytometric analysis of the surface CD33 expression following knockdown of *PHF3*, *CHD4* or
628 the combination in the SEM cell line. shNTC – non-targeting control. (E) Gene set enrichment
629 analysis of RNA sequencing data derived from: knockdown of *PHF3* and *CHD4* in the SEM
630 cell line (left panel) and lineage switch leukaemia cases (right panel). Shown is negative
631 correlation of shCHD4/shPHF3 and relapse samples with Jaatinen haematopoietic stem cell
632 signature (upper panel) and Haddad B-lymphocyte progenitor signature (lower panel). (F)
633 Expression of lineage specific cell surface markers CD19 (lymphoid) and CD33 (myeloid)
634 following culture of *MLL-Af4* transformed hCD34+ cord blood progenitor cells in lymphoid
635 permissive conditions. Knockdown of *PHF3*, *CHD4* or the combination disrupts the dominant
636 lymphoid differentiation pattern seen in non-targeting control (shNTC). (G) Assessment of
637 *PHF3* knockdown influence on the surface marker expression after longer incubation period
638 (33 days); *CHD4* knockdown impaired cellular survival upon longer *in vitro* culture.

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

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

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

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

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

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

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

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

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

708 **Figure S8. The functional role of epigenetic modifiers on leukaemia lineage switching.**
709 (A) PHF3 scheme; the K1119I mutation is located at a highly conserved residue. An *
710 (asterisk) indicates positions which have a single, fully conserved residue, a : (colon) indicates
711 conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM
712 250 matrix, a . (period) indicates conservation between groups of weakly similar properties -
713 scoring < 0.5 in the Gonnet PAM 250 matrix. (B) Mutual information sub-networks illustrate
714 high centrality of mutated genes *PHF3* (left panel) and *CHD4* (right panel) within the AML/ALL
715 transcriptional network. Size and colour of nodes reflect differential expression between AML
716 and ALL. (C) Myeloid marker CD33 expression level upon knockdown of *CHD4*, *PHF3*, and
717 non-targeting control (NTC) in *MLL*r (t(4;11)) cell line RS4;11 and non-*MLL*r cell lines, 697 and
718 REH cells. (D) CD19 and CD33 surface expression level change upon knockdown of *CHD4*
719 and *PHF3* in the PDX sample generated from the first relapse of patient LS03 (ALL relapse).
720 (E) Protein network analysis of differentially expressed NuRD and PRC1 members, showing
721 interactions between proteins (and clusters, depicted as blue and red colour); interaction
722 strength is shown as increasing thickness of the joining line (https://string-db.org/). (F) Fold
723 change expression level changes of PRC1 members following lineage switched relapse and
724 in MPAL cases. (G) Expression of lineage specific surface markers following *PCGF6*

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

732 **Table titles and legends**

733 **Supplementary Table 1.** Clinical characteristics of patient cases

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

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

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

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

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

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

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

755

756

757 **STAR Methods**

758

759 **KEY RESOURCES TABLE**

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

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

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

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

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

CONTACT FOR REAGENT AND RESOURCE SHARING

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient samples and data

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

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

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

790 **Cell Lines**

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

795 *MLL-Af4* transduced cord blood cells (Lin *et al.*, 2016) were cultured in IMDM with 10% FBS,
796 2 mM Glutamine supplemented with recombinant human SCF, IL-3, IL-6, FLT-3L, and TPO
797 (10 ng/ml each) at 37°C in a humidified 5% CO_2 incubator. To prime towards lymphoid
798 differentiation, the cells were co-cultured with MS-5 (RRID:CVCL_2128, murine) stroma cells

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

801 ***In Vivo* Mouse Studies**

802 *In vivo* studies were conducted in accordance with the UK Animals (Scientific Procedures) Act
803 1986 under project licences PPL60/4552 and PPL60/4222 following institutional ethical
804 review.

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

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

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

817

818 **METHOD DETAILS**

819 **Long-distance inverse PCR experiments**

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

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

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

832 **PCR-based clonality immunoglobulin gene targets**

833 Clonality analysis based on immunoglobulin gene rearrangements was performed according
834 to the BIOMED-2 Concerted Action BMH-CT98-3936 using standardised protocols and primer
835 sets (van Dongen *et al.*, 2003). DNA was amplified by multiplex PCR assays IGH (VH–JH and
836 DH–JH) and IGK (Vκ–Jκ and Kde primers) using AmpliTaq Gold (Applied Biosystems) in
837 duplicate reactions. The PCR reaction consists of 100 ng DNA, 1 X ABI Gold Buffer, 10 pmol
838 of each primer, 200 µM final concentration dNTP, 1.5 mM final concentration $MgCl_2$, and 1 U
839 Taq enzyme in 50 µl volume. The PCR amplification parameters were: one cycle at 95°C for
840 10 min, thirty five cycles of 95°C for 45 s, 60°C for 45 s, and 72°C for 1.5 min, followed by one
841 cycle at 72°C for 10 min. The duplicate samples were analysed by GeneScan analysis on a
842 Life Technologies 3100 platform.

843 **Exome sequencing**

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

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

856 Raw reads were aligned to human reference genome (hg19 or hg38 for lineage switch or
857 MPAL patients, respectively) using Burrows-Wheeler Aligner (BWA) 0.7.12 (Li and Durbin,
858 2009) and were processed using the Genome Analysis Toolkit (GATK, v3.8 or 4.1) best
859 practices recommended workflow for variant discovery analysis (Mckenna *et al.*, 2010;
860 DePristo *et al.*, 2011; van der Auwera *et al.*, 2013). MuTect (v1.1.7) and MuTect2 (4.1) were
861 used to identify somatic variants for each matched sample pair (Cibulskis *et al.*, 2013).
862 Variants were annotated using Ensembl Variant Effect Predictor (VEP, version 90) (Cibulskis
863 *et al.*, 2013).

864 RNA sequencing

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

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

879 **Alternative splicing analysis**

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

888 **Whole genome sequencing**

889 Presentation, remission and relapse DNA samples from case LS01 were sequenced by
890 Illumina UK and analysed using the remission sample as the matching normal. Sequencing
891 reads were aligned to the human GRCh37.1 reference genome using ISAAC (Raczy *et al.*,
892 2013). Identification of somatic SNVs and small somatic indels (<50 bp) was performed by
893 Strelka (Saunders *et al.*, 2012). Large structural variants (including deletions, inversions,
894 duplications and insertions all >50bp and translocations) were called by Manta (Chen *et al.*,
895 2016).

896 **Non-negative matrix factorization analysis**

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

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

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

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

936 **DHS library generation, sequencing, and mapping**

937 DHS analysis was performed as described previously (Ptasinska *et al.*, 2014; Cauchy *et al.*,
938 2015). DNase I (Worthington, Cat# LS006328) digestion was performed using ~5 million
939 patient sample cells using 8 units (presentation) or 14 units (relapse) for 3 min at 22°C in a 1
940 mM $CaCl_2$ supplemented buffer. Nuclear proteins were digested with 1 mg/ml Proteinase K
941 overnight at 37°C. DNase I digestion products were size-selected on an agarose gel, cutting
942 below 150 bp. High-throughput sequencing libraries were prepared from 10 ng of size-selected
943 material, using the Kapa Hyperprep kit as per manufacturer's instruction. Libraries were
944 sequenced with 50 bp single-end reads on an Illumina HiSeq 2500 sequencer according to
945 manufacturer's instructions.

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

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

956 **Flow cytometry and cell sorting**

957 One million cells (of all tested cell lines, cord blood or LS03ALL PDX cells) were used for flow
958 cytometry analysis of CD19 (BV421) and CD33 (APCCy7). The antibody details are listed in
959 the Key Resources Table. Cells were collected and washed in a 3 ml wash buffer containing
960 0.2% BSA in PBS, re-suspended in 100 µl wash buffer and 5 µl each antibody was added,
961 followed by incubation at RT for 20 min in the dark. Cells were washed and re-suspended in
962 a 0.5 ml wash buffer. Flow cytometry was performed on FACSCanto II from Becton Dickinson
963 (BD) and data analysed with FlowJo (Treestar).

964 Haematopoietic hierarchy analysis was performed from 10^7 cells of primary and PDX samples.
965 The cells were collected and incubated for 30 min in a wash buffer containing 0.5% FBS, 2
966 mM EDTA in PBS. Following the incubation, the cells were washed and resuspended in a
967 wash buffer. Fluorescence activated cell sorting (FACS) was performed using a FACS Aria
968 Fusion (BD) or ASTRIOS EQ (Beckman Coulter). In order to separate early progenitors or
969 lymphoid and myeloid fractions, MPAL patient samples were sorted from 10^7 or $3*10^6$ cells for
970 MPAL1 and MPAL2 respectively. Sorted cells were collected into 1.5 ml microfuge tubes
971 containing 500 µl RPMI with 10% fetal calf serum or SFEM II media in case of cell lines or
972 primary samples, respectively. In order to analyse single cell derived clones, sorted early
973 progenitor cells derived from MPAL samples were deposited onto 384-well plates containing
974 75ul of SFEM II media (supplemented with SCF, FLT3L, IL6, IL3 and TPO). Single cells were
975 grown into 60-90% confluency colonies followed by RNA/DNA extraction and MLL-AF4
976 detection by PCR.

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

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

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

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

994 **Nested multiplex PCR and targeted sequencing**

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

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

1002 cycle at 72°C for 10 min. The products were diluted 500-fold, and 1 µl used as the template
1003 for the second PCR reactions (25 µl reaction). The second/nested multiplex reaction
1004 parameters were: one cycle at 98°C for 2 min, twenty cycles of 98°C for 10 s, 65°C for 30 s,
1005 and 72°C for 30 s, followed by one cycle at 72°C for 10 min. The amplicons that were taken
1006 forward for next-generation targeted sequencing had additional CS1
1007 (ACACTGACGACATGGTTCTACA) and CS2 (TACGGTAGCAGAGACTTGGTCT) Fluidigm
1008 tag sequences on the nested PCR primer forward and reverse, respectively. The amplicons
1009 were barcoded using Fluidigm Access Array Barcode Library for Illumina Sequencers
1010 (Cat#100-4876) by taking 0.8 µl multiplex PCR products (multiplex group A-C), 4 µl Fluidigm
1011 barcode primer (400 nM final concentration), 10 µl of 2X Phusion Master Mix (NEB,
1012 Cat#M0531L), and 3.6 µl H_2O (20 µl reaction). The barcoding PCR reaction parameters were:
1013 one cycle at 98°C for 2 min, six cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 1 min,
1014 followed by one cycle at 72°C for 10 min. The products were run on 2% agarose gel and
1015 extracted using the QIAquick Gel Extraction Kit (Qiagen, Cat#28706). The purified products
1016 were submitted for paired-end 300 bp sequencing on MiSeq (Illumina), resulting in >1,000
1017 coverage per gene.

1018 ***MLL-Af4* stem cell expression analysis**

1019 Following myeloid or lymphoid culture of *MLL/Af4* transduced CD34+ cord blood cells,
1020 CD19+CD33-, CD19-CD33+ and CD19-CD33- populations were flow sorted, lysed and RNA
1021 extracted using RNeasy Micro Kit (Qiagen), according to the manufacturer's instructions. Input
1022 RNA was equilibrated to a starting input cell number of 300 cells per population before cDNA
1023 and sequencing library production were performed using SMARTSeqv4 (Clontech) and
1024 NexteraXT (Illumina) kits, according to manufacturer's instructions. The resultant libraries
1025 were submitted for paired-end 150 bp sequencing on a NEXTSeq500 (Illumina). For each
1026 sample, transcript abundance was quantified from raw reads with Salmon (version 0.8.2)
1027 (Patro *et al.*, 2017) using the reference human transcriptome (hg38) defined by GENCODE
1028 release 27. An R package Tximport (version 1.4.0) (Soneson, Love and Robinson, 2015) was

1029 used to estimate gene-level abundance from Salmon's transcript-level counts. Gene-level
1030 differential expression analysis was performed using DESeq2 (version 1.16.1) (Love, Huber
1031 and Anders, 2014) prior NMF analysis as described above.

1032 ***PHF3*, *CHD4*, *PCGF6*, and *AUTS2* shRNAs**

1033 shRNAs against PHF3 (TRCN0000019118, TRCN0000019114, TRCN0000274376), CHD4
1034 (TRCN0000380981, TRCN0000021363, TRCN0000021360), PCGF6 (TRCN0000229804,
1035 TRCN0000073109, TRCN0000073109), AUTS2 (TRCN0000119058, TRCN0000304019,
1036 TRCN0000304081), or non-targeting control (ATCTCGCTTGGGCGAGAGTAAG) were
1037 cloned into pLKO5d.SFFV.miR30n (Schwarzer *et al.*, 2017). Each target gene was linked with
1038 different fluorescent protein, including dTomato (PHF3 and AUTS2), eGFP (CHD4 and
1039 PCGF6), and RFP657 (non-targeting control).

1040 The vector contained BsmBI at the cloning sites. The oligonucleotides were designed with the
1041 appropriate complementary overhang sequences of BsmBI-cleaved vector. They were ligated
1042 and transformed into STBL3 chemically competent cells (Invitrogen) according to
1043 manufacturer's instructions and plated on agar plates containing 100 µg/ml ampicillin
1044 incubated for 16 h at 37°C. Single colonies were inoculated for DNA preparation, and extracted
1045 using EndoFree Plasmid Maxi Kit (Qiagen, Cat#12362). All clones were verified by sequencing
1046 prior to lentivirus production.

1047 **Lentivirus Production and Transduction**

1048 Viral particles were produced using calcium phosphate precipitation method on 293T cells by
1049 co-transfecting an envelope plasmid pMD2.G, packaging plasmid pCMVΔR8.91, and the
1050 shRNA vector pLKO5d.SFFV.miR30n. The cells were grown in 100 mm tissue culture dishes
1051 at a concentration of 1-2 x 10^6 cells in 10 ml medium the day prior to co-transfection. On the
1052 following day when the cells had reached 30-50% confluence, 20 µg of shRNA vector, 15 µg
1053 of pCMVΔR8.91, and 5 µg of pMD2.G were mixed. The volume of the mixture was adjusted

1054 with HEPES buffer solution (2.5 mM HEPES containing deionized water at pH7.3) to 250 µl.
1055 A volume of 250 µl of 0.5 M $CaCl_2$ was added to the mixture. This solution was added to 500
1056 µl of 2X HeBS (0.28 M NaCl, 0.05 M HEPES and 1.5 mM Na_2HPO_4 in deionized water at pH
1057 7.00) and mixed by vortexing. It was incubated at RT for 30-40 min to allow the formation of
1058 the calcium phosphate precipitate, before adding dropwise on the 293T cells. After 16 h, the
1059 cells were washed with 10 ml PBS and added with 10 ml culture media. They were incubated
1060 at 37°C in a humidified atmosphere with 5% CO_2 for the next two days. Lentivirus particles
1061 were collected by centrifuging the supernatant at 400 g for 10 min at 4°C and filtered through
1062 Acrodisc Syringe 0.45 filters. They were stored in aliquots at -80°C. Cell lines and PDX
1063 samples were transduced with lentivirus as previously described (Bomken *et al.*, 2013).

1064 Lymphoid differentiation of transduced *MLL-Af4* cord blood cells

1065 *MLL-Af4* cord blood cells (5 x 10^5 cells per transduction, Lin *et al.*, 2016) were transduced
1066 with short hairpin constructs as described and co-cultured with MS-5 stroma cells in
1067 lymphoid culture conditions described above. Single and triple transduced populations were
1068 identified using the construct specific fluorophores and lineage specific surface markers
1069 assessed as a proportion of the total transduced leukocyte population.

1070 qRT-PCR

1071 One million cells were collected and RNA was extracted using RNeasy Mini Kit (Qiagen,
1072 Cat#74106). cDNA was synthesised from 1 µg RNA using RevertAid H Minus First Strand
1073 cDNA Synthesis Kit (Thermo Fisher Scientific, Cat#K1632). The product was diluted by
1074 adding 80 µl H_2O (Livak and Schmittgen, 2001). Sequences of primers used in this study were
1075 designed using Primer Express (Applied Biosystems) software.

1076 **QUANTIFICATION AND STATISTICAL ANALYSIS**

1077 **Reverse Engineering of Transcriptional Networks**

1078 A mutual information network was developed such that the connections between genes within
1079 the network represented an inferred likelihood of a causal relationship within ALL/AML. The
1080 network was further refined by retaining only those nodes (genes) and edges (causal
1081 connections) which are significantly associated, in expression terms, to the distinction
1082 between normal myeloid and lymphoid cells. Thus, the centrality of the mutations within the
1083 network reflects the estimated extent of their causal influence upon a lymphoid/myeloid
1084 distinction within the framework of primary ALL/AML.

1085 Mutual Information networks were reverse engineered from 216 published AML/ALL
1086 Affymetrix HGU133p2 expression profiles (GSE11877 & GSE24006) processed using Rma
1087 (affy package R/Bioconductor) using the ARACNe2 algorithm (Margolin *et al.*, 2006) set to
1088 adaptive partitioning with a DPI tolerance of 0.1 and a p-value threshold of 0.01 and 1000
1089 bootstraps. The probes/genes were filtered to remove those that were insufficiently variable
1090 as per the Aracne method. Probes had to be greater than 3 fold expression and 300 delta
1091 between the min and max excluding the 5% most extreme values to be included. Probes
1092 without >20 intensity in greater than 20% of samples were also removed. All probes which
1093 fulfilled these requirements were included in the construction of this network (n=9780).

1094 Nodes within the network were annotated with a metagene value calculated using Non-
1095 Negative Matrix Factorisation (k=2) and reflecting the differences between myeloid and
1096 lymphoid cells (GSE24759) and the other. A metagene cutoff of 0.13 (based on a calculated
1097 alpha of 0.1) was applied to trim the network. Cytoscape (Shannon *et al.*, 2003) was used to
1098 merge nodes where probe sets represented common genes and to calculate centrality
1099 statistics for each of the mutated genes of interest. Nodes of interest (i.e. mutated nodes) were
1100 ranked according to their centrality (e.g. degree). Nodes were sized and coloured according

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

DATA AND SOFTWARE AVAILABILITY

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

1112 **References**

1113 Agraz-Doblas, A., Bueno, C., Bashford-Rogers, R., Roy, A., Schneider, P., Bardini, M., Ballerini, P.,
1114 Cazzaniga, G., Moreno, T., Revilla, C., Agraz-Doblas, A., *et al*. (2019). Unravelling the cellular origin and
1115 clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis.
1116 Haematologica, 104, 1176-1188. 10.3324/haematol.2018.206375.

1117 Alexander, T.B., Gu, Z., Iacobucci, I., Dickerson, K., Choi, J.K., Xu, B., Payne-Turner, D., Yoshihara, H.,
1118 Loh, M.L., Horan, J., *et al*. (2018) The genetic basis and cell of origin of mixed phenotype acute
1119 leukaemia. Nature 562, 373-379. 10.1038/s41586-018-0436-0.

1120 Alexandrov, L. B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., Stratton, M.R.
1121 (2015) Clock-like mutational processes in human somatic cells. Nature Genetics 47, 1402–1407.
1122 10.1038/ng.3441.

1123 Andersson, A.K., Ma, J., Wang, J., Chen, X., Gedman, A.L., Dang, J., Nakitandwe, J., Holmfeldt, L.,
1124 Parker, M., Easton, J., *et al*. (2015). The landscape of somatic mutations in infant *MLL*-rearranged acute
1125 lymphoblastic leukemias. Nature Genetics 47, 330-337. 10.1038/ng.3230.

1126 Appel, L-M., Franke, V., Bruno, M., Grishkovskaya, I., Kasiliauskaite, A., Schoeberl, U.E., Puchinger,
1127 M.G., Kostrhon, S., Beltzung, E., Mechtler, K., *et al*. (2020) PHF3 regulates neuronal gene expression
1128 through the new Pol II CTD reader domain SPOC. bioRxiv, p. 2020.02.11.943159.
1129 10.1101/2020.02.11.943159.

1130 Arends, T., Dege, C., Bortnick, A., Danhorn, T., Knapp, J.R., Jia, H., Harmacek, L., Fleenor, C.J., Straign,
1131 D., Walton, K., *et al*. (2019). CHD4 is essential for transcriptional repression and lineage progression in
1132 B lymphopoiesis. Proc Natl Acad Sci U S A 116, 10927-10936. 10.1073/pnas.1821301116.

1133 Ayton, P. M., and Cleary, M. L. (2003). Transformation of myeloid progenitors by MLL oncoproteins is
1134 dependent on Hoxa7 and Hoxa9. Genes Dev 17, 2298–2307. 10.1101/gad.1111603.

1135 Ferreira, B.I., García, J.F., Suela, J., Mollejo, M., Camacho, F.I., Carro, A., Montes, S., Piris, M.A.,
1136 Cigudosa, J.C. (2008). Comparative genome profiling across subtypes of low-grade B-cell lymphoma
1137 identifies type-specific and common aberrations that target genes with a role in B-cell neoplasia.
1138 Haematologica 93, 670–679. 10.3324/haematol.12221.

1139 Blokzijl, F. de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E.,
1140 Prins, P. (2016) Tissue-specific mutation accumulation in human adult stem cells during life. Nature
1141 538, 260–264. doi: 10.1038/nature19768.

1142 Boer, J.M., van der Veer, A., Rizopoulos, D., Fiocco, M., Sonneveld, E., de Groot-Kruseman, H.A.,
1143 Kuiper, R.P., Hoogerbrugge, P., Horstmann, M., Zaliova, M., *et al*. (2016). Prognostic value of rare IKZF1
1144 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative
1145 study. Leukemia, 30, 32–38. 10.1038/leu.2015.199.

1146 Bolotin, D.A., Poslavsky, S., Davydov, A.N., Frenkel, F.E., Fanchi, L., Zolotareva, O.I., Hemmers, S.,
1147 Putintseva, E.V., Obraztsova, A.S., Shugay, M., *et al*. (2017). Antigen receptor repertoire profiling from
1148 RNA-seq data. Nat Biotechnol, 35, 908–911. 10.1038/nbt.3979.

1149 Bolouri, H., Farrar, J.E., Triche, T.Jr., Ries, R.E., Lim, E.L., Alonzo, T.A., Ma, Y., Moore, R., Mungall, A.J.,
1150 Marra, M.A., *et al*. (2018). The molecular landscape of pediatric acute myeloid leukemia reveals
1151 recurrent structural alterations and age-specific mutational interactions. Nat Med, 24, 103–112.

1152 10.1038/nm.4439 PMID - 29227476.

1153 Bomken, S., Buechler, L., Rehe, K., Ponthan, F., Elder, A., Blair, H., Bacon, C.M., Vormoor, J.,
1154 Heidenreich, O. (2013). Lentiviral marking of patient-derived acute lymphoblastic leukaemic cells
1155 allows in vivo tracking of disease progression. Leukemia, 27, 718-721. 10.1038/leu.2012.206

1156 Brunet, J.P., Tamayo, P., Golub, T.R., Mesirov, J.P. (2004). Metagenes and molecular pattern discovery
1157 using matrix factorization. Proc Natl Acad Sci U S A 101, 4164–4169. 10.1073/pnas.0308531101.

1158 Carulli, G., Marini, A., Ferreri, M.I., Azzarà, A., Ottaviano, V., Lari, T., Rocco, M., Giuntini, S., Petrini, M.
1159 (2012). B-cell acute lymphoblastic leukemia with t(4;11)(q21;q23) in a young woman: evolution into
1160 mixed phenotype acute leukemia with additional chromosomal aberrations in the course of therapy.
1161 Hematol Rep 4, e15. 10.4081/hr.2012.e15.

1162 Cauchy, P., James, S.R., Zacarias-Cabeza, J., Ptasinska, A., Imperato, M.R., Assi, S.A., Piper, J.,
1163 Canestraro, M., Hoogenkamp, M., Raghavan, M., *et al*. (2015). Chronic FLT3-ITD Signaling in Acute
1164 Myeloid Leukemia Is Connected to a Specific Chromatin Signature. Cell Rep 12, 821–836.
1165 10.1016/j.celrep.2015.06.069.

1166 Chen, L., Kostadima, M., Martens, J.H.A., Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C.,
1167 Bielczyk-Maczynska, E., Coe, S., *et al*. (2014). Transcriptional diversity during lineage commitment of
1168 human blood progenitors. Science 345, 1251033. 10.1126/science.1251033.

1169 Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S.,
1170 Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer
1171 sequencing applications. Bioinformatics 32, 1220–1222. 10.1093/bioinformatics/btv710.

1172 Church, D.N., Briggs, S.E., Palles, C., Domingo, E., Kearsey, S.J., Grimes, J.M., Gorman, M., Martin, L.,
1173 Howarth, K.M., Hodgson, S.V., *et al*. (2013). DNA polymerase ε and δ exonuclease domain mutations
1174 in endometrial cancer. Hum Mol Genet 22, 2820–2828. 10.1093/hmg/ddt131.

1175 Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson,
1176 M., Lander, E.S., Getz, G. (2013). Sensitive detection of somatic point mutations in impure and
1177 heterogeneous cancer samples. Nat Biotechnol 31, 213–219. 10.1038/nbt.2514.

1178 DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel,
1179 G., Rivas, M.A., Hanna, M., *et al*. (2011). A framework for variation discovery and genotyping using
1180 next-generation DNA sequencing data. Nat Genet 43, 491–498. 10.1038/ng.806.

1181 Dobbins, S.E., Sherborne, A.L., Ma, Y.P., Bardini, M., Biondi, A., Cazzaniga, G., Lloyd, A., Chubb, D.,
1182 Greaves, M.F., Houlston, R.S. (2013). The silent mutational landscape of infant MLL-AF4 pro-B acute
1183 lymphoblastic leukemia. Genes Chromosomes Cancer 52, 954–960. 10.1002/gcc.22090.

1184 Erson-Omay, E.Z., Çağlayan, A.O., Schultz, N., Weinhold, N., Omay, S.B., Özduman, K., Köksal, Y., Li, J.,
1185 Serin Harmancı, A., Clark, V., *et al*. (2015). Somatic POLE mutations cause an ultramutated giant cell
1186 high-grade glioma subtype with better prognosis. Neuro Oncol 17, 1356–1364.
1187 10.1093/neuonc/nov027.

1188 Gardner, R., Wu, D., Cherian, S., Fang, M., Hanafi, L.A., Finney, O., Smithers, H., Jensen, M.C., Riddell,
1189 S.R., Maloney, D.G., *et al*. (2016). Acquisition of a CD19-negative myeloid phenotype allows immune
1190 escape of *MLL*-rearranged B-ALL from CD19 CAR-T-cell therapy. Blood 127, 2406-2410.
1191 10.1182/blood-2015-08-665547.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1274 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
1275 RNA-seq data with DESeq2. Genome Biol 15, 550. 10.1186/s13059-014-0550-8.

1276 Lu, X., Chu, C.S., Fang, T., Rayon-Estrada, V., Fang, F., Patke, A., Qian, Y., Clarke, S.H., Melnick, A.M.,
1277 Zhang, Y., *et al*. (2019). MTA2/NuRD Regulates B Cell Development and Cooperates with OCA-B in
1278 Controlling the Pre-B to Immature B Cell Transition. Cell Rep 28, 472-485.
1279 10.1016/j.celrep.2019.06.029.

1280 Malouf, C., and Ottersbach, K. (2018). The fetal liver lymphoid-primed multipotent progenitor
1281 provides the prerequisites for the initiation of t(4;11) MLL-AF4 infant leukemia. Haematologica, 103,
1282 e571–e574. 10.3324/haematol.2018.191718.

1283 Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., Califano, A.
1284 (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian
1285 Cellular Context. BMC Bioinformatics 7, S7. doi: 10.1186/1471-2105-7-S1-S7.

1286 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler,
1287 D., Gabriel, S., Daly, M., DePristo, M.A. (2010). The genome analysis toolkit: A MapReduce framework
1288 for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303.
1289 10.1101/gr.107524.110.

1290 Meyer, C., Schneider B., Dingermann, T., Klingebiel, T., Marschalek, R. (2005). A LDI-PCR based method
1291 allows the identification of any MLL rearrangement. Blood 106, 2848.
1292 10.1182/blood.V106.11.2848.2848.

1293 Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De
1294 Braekeleer, E., De Braekeleer, M., *et al*. (2009). New insights to the MLL recombinome of acute
1295 leukemias. Leukemia 23, 1490–1499. 10.1038/leu.2009.33.

1296 Meyer, C., Hofmann, J., Burmeister, T., Gröger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M.,
1297 Renneville, A., Villarese, P., Macintyre, E., *et al*. (2013). The MLL recombinome of acute leukemias in
1298 2013. Leukemia 27, 2165–2176. 10.1038/leu.2013.135.

1299 Meyer, C., Burmeister, T., Gröger, D., Tsaur, G., Fechina, L., Renneville, A., Sutton, R., Venn, N.C.,
1300 Emerenciano, M., Pombo-de-Oliveira, M.S. (2018). The MLL recombinome of acute leukemias in 2017.
1301 Leukemia 32, 273–284. 10.1038/leu.2017.213.

1302 Moorman, A.V., Ensor, H.M., Richards, S.M., Chilton, L., Schwab, C., Kinsey, S.E., Vora, A., Mitchell,
1303 C.D., Harrison, C.J. (2010). Prognostic effect of chromosomal abnormalities in childhood B-cell
1304 precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99
1305 randomised trial. Lancet Oncol 11, 429–438. 10.1016/S1470-2045(10)70066-8.

1306 Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula,
1307 S., Krupalnik, V., *et al*. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates
1308 Deterministic Induction of Naive Pluripotency. Cell Stem Cell 23, 412-425.
1309 10.1016/j.stem.2018.07.004.

1310 Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K.,
1311 Mathew, S., Ma, J., Pounds, S.B., *et al*. (2007). Genome-wide analysis of genetic alterations in acute
1312 lymphoblastic leukaemia. Nature 446, 758–764. 10.1038/nature05690.

1313 Mur, P., García-Mulero, S., Del Valle, J., Magraner-Pardo, L., Vidal, A., Pineda, M., Cinnirella, G., Martín-
1314 Ramos, E., Pons, T., López-Doriga, A., *et al*. (2020). Role of POLE and POLD1 in familial cancer. Genet

1315 Med 22, 2089–2100. 10.1038/s41436-020-0922-2.

1316 Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent
1317 hematopoietic progenitors. Genes Dev 12, 2403–2412. 10.1101/gad.12.15.2403.

1318 Ng, S.Y., Yoshida, T., Zhang, J., Georgopoulos, K. (2009). Genome-wide lineage-specific transcriptional
1319 networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity 30,
1320 493–507. 10.1016/j.immuni.2009.01.014.

1321 Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N.,
1322 Yosef, N., Chang, C.Y., Shay, T., *et al*. (2011). Densely Interconnected Transcriptional Circuits Control
1323 Cell States in Human Hematopoiesis. Cell 144, 296–309. 10.1016/j.cell.2011.01.004.

1324 Nutt, S.L., Urbánek, P., Rolink, A., Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell
1325 development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination
1326 at the IgH locus. Genes Dev 11, 476–491. 10.1101/gad.11.4.476.

1327 Nutt, S.L., Heavey, B., Rolink, A.G., Busslinger, M. (1999). Commitment to the B-lymphoid lineage
1328 depends on the transcription factor Pax5. Nature 402, 14–20. 10.1038/35005514.

1329 O'Neill, D. W., Schoetz, S.S., Lopez, R.A., Castle, M., Rabinowitz, L., Shor, E., Krawchuk, D., Goll, M.G.,
1330 Renz, M., Seelig, H.P., Han, S., *et al*. (2000). An ikaros-containing chromatin-remodeling complex in
1331 adult-type erythroid cells. Mol Cell Biol 20, 7572–7582. 10.1128/mcb.20.20.7572-7582.2000.

1332 Orlando, E.J., Han, X., Tribouley, C., Wood, P.A., Leary, R.J., Riester, M., Levine, J.E., Qayed, M., Grupp,
1333 S.A., Boyer, M., *et al*. (2018). Genetic mechanisms of target antigen loss in CAR19 therapy of acute
1334 lymphoblastic leukemia. Nat Med 24, 1504–1506. 10.1038/s41591-018-0146-z.

1335 Park, M., Koh, K.N., Kim, B.E., Im, H.J., Jang, S., Park, C.J., Chi, H.S., Seo, J.J. (2011). Lineage Switch at
1336 Relapse of Childhood Acute Leukemia: A Report of Four Cases. J Korean Med Sci 26, 829–831.
1337 10.3346/jkms.2011.26.6.829.

1338 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C. (2017). Salmon provides fast and bias-aware
1339 quantification of transcript expression. Nat Methods 14, 417–419. 10.1038/nmeth.4197.

1340 Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., Hovi, L., LeBlanc, T.,
1341 Szczepanski, T., Ferster, A., *et al*. (2007). A treatment protocol for infants younger than 1 year with
1342 acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised
1343 trial. Lancet 370, 240–250. 10.1016/S0140-6736(07)61126-X.

1344 Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., Ott, S. (2013). Wellington: a novel method
1345 for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids
1346 Research 41, e201–e201. 10.1093/nar/gkt850.

1347 Piper, J., Assi, S.A., Cauchy, P., Ladroue, C., Cockerill, P.N., Bonifer, C., Ott, S. (2015). Wellington-
1348 bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors.
1349 BMC Genomics 16, 1000. 10.1186/s12864-015-2081-4.

1350 Pongubala, J.M., Northrup, D.L., Lancki, D.W., Medina, K.L., Treiber, T., Bertolino, E., Thomas, M.,
1351 Grosschedl, R., Allman, D., Singh, H. (2008). Transcription factor EBF restricts alternative lineage
1352 options and promotes B cell fate commitment independently of Pax5. Nat Immunol 9, 203–215.
1353 10.1038/ni1555.

1354 Ptasinska, A., Assi, S.A., Martinez-Soria, N., Imperato, M.R., Piper, J., Cauchy, P., Pickin, A., James, S.R.,
1355 Hoogenkamp, M., Willimason, D. *et al*. (2014). Identification of a Dynamic Core Transcriptional
1356 Network in t(8;21) AML that Regulates Differentiation Block and Self-Renewal. Cell Rep 8, 1974–1988.
1357 10.1016/j.celrep.2014.08.024 .

1358 Rabilloud, T.,Potier, D., Pankaew, S., Nozais, M., Loosveld, M., Payet-Bornet, D. (2021). Single-cell
1359 profiling identifies pre-existing CD19-negative subclones in a B-ALL patient with CD19-negative relapse
1360 after CAR-T therapy. Nature Commun 12, 865. 10.1038/s41467-021-21168-6.

1361 Raczy, C., Petrowski, R., Saunders, C.T., Chorny, I., Kruglyak, S., Margulies, E.H., Chuang, H.-Y., Källberg,
1362 M., Kumar, S.A., Liao, A . (2013). Isaac: Ultra-fast whole-genome secondary analysis on Illumina
1363 sequencing platforms. Bioinformatics 29, 2041–2043. 10.1093/bioinformatics/btt314.

1364 Rayes, A., McMasters, R.L., O'Brien, M.M. (2016). Lineage switch in MLL-rearranged infant leukemia
1365 following CD19-directed therapy. Pediatr Blood Cancer 63, 1113-510. 10.1002/pbc.25953

1366 Rehe, K. Wilson, K., Bomken, S., Williamson, D., Irving, J., den Boer, M.L., Stanulla, M., Schrappe, M.,
1367 Hall, A.G., Heidenreich, O., Vormoor, J.(2013). Acute B lymphoblastic leukaemia-propagating cells are
1368 present at high frequency in diverse lymphoblast populations. EMBO Mol Med 5, 38–51.
1369 10.1002/emmm.201201703.

1370 Risso, D. Ngai, J., Speed, T.P., Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of
1371 control genes or samples.Nature Biotechnol 32, 896–902. 10.1038/nbt.2931.

1372 Rongvaux, A. Willinger, T., Martinek, J., Strowig, T., Gearty, S.V., Teichmann, L.L., Saito, Y., Marches,
1373 F., Halene, S., Palucka, A.K. *et al*. (2014). Development and function of human innate immune cells in
1374 a humanized mouse model.Nature Biotechnol 32, 364–372. 10.1038/nbt.2858.

1375 Rossi, J.G., Andrea R Bernasconi, Cristina N Alonso, Patricia L Rubio, Marta S Gallego, Carolina A
1376 Carrara, Myriam R Guitter, Silvia Eandi Eberle, Mariela Cocce, Pedro A Zubizarreta, *et al*. (2012).
1377 Lineage switch in childhood acute leukemia: An unusual event with poor outcome. Am J Hematol 87,
1378 890–897. 10.1002/ajh.23266.

1379 Saldanha, A. J. (2004). Java Treeview - Extensible visualization of microarray data. Bioinformatics 20,
1380 3246–3248.10.1093/bioinformatics/bth349.

1381 Saunders, C.T. Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., Cheetham, R.K. (2012). Strelka:
1382 Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28,
1383 1811–1817. 10.1093/bioinformatics/bts271.

1384 Schwalbe, E.C., Williamson, D., Lindsey, J.C., Hamilton, D., Ryan, S.L., Megahed, H., Garami, M., Hauser,
1385 P., Dembowska-Baginska, B., Perek, D. *et al*. (2013). DNA methylation profiling of medulloblastoma
1386 allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. Acta
1387 Neuropathol 125, 359–371. 10.1007/s00401-012-1077-2.

1388 Schwarzer, A. Emmrich, S., Schmidt, F., Beck, D., Ng, M., Reimer, C., Adams, F.F., Grasedieck, S., Witte,
1389 D., Käbler, S. *et al*. (2017). The non-coding RNA landscape of human hematopoiesis and leukemia.
1390 Nature Commun 8, 218.10.1038/s41467-017-00212-4.

1391 Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B.,
1392 Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular
1393 interaction networks. Genome Research 13, 2498–2504. 10.1101/gr.1239303.

1394 Sifrim, A. Hitz, M.-P., Wilsdon, A., Breckpot, J., Al Turki, S.H., Thienpont, B., McRae, J., Fitzgerald,
1395 T.W.F., Singh, T., Swaminathan, G.J. *et al*. (2016). Distinct genetic architectures for syndromic and
1396 nonsyndromic congenital heart defects identified by exome sequencing. Nature Gen 48, 1060–1065.
1397 10.1038/ng.3627.

1398 Somervaille, T.C.P., Christina J Matheny, C.J., Gary J Spencer, G.J., Masayuki Iwasaki, M., John L Rinn,
1399 J.L., Daniela M Witten, D.M., Howard Y Chang, H.Y., Sheila A Shurtleff, S.A., James R Downing, J.R.,
1400 Michael L Cleary, M.L. (2009). Hierarchical maintenance of MLL myeloid leukemia stem cells employs
1401 a transcriptional program shared with embryonic rather than adult stem cells. Cell Stem Cell 4, 129–
1402 140. 10.1016/j.stem.2008.11.015.

1403 Soneson, C., Love, M.I., Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level
1404 estimates improve gene-level inferences. F1000 Research 4, 1521. 10.12688/f1000research.7563.

1405 Sotillo, E., Barrett, D.M., Black, K.L., Bagashev, A., Oldridge, D., Wu, G., Sussman, R., Lanauze, C., Ruella,
1406 M., Gazzara, M.R., *et al*. (2015). Convergence of Acquired Mutations and Alternative Splicing of CD19
1407 Enables Resistance to CART-19 Immunotherapy. Cancer Discovery 5, 1282–1295. 10.1158/2159-
1408 8290.CD-15-1020.

1409 Sperlazza, J., Rahmani, M., Beckta, J., Aust, M., Hawkins, E., Wang, S., Zhu, S., Podder, S., Dumur, C.,
1410 Archer, K., *et al*. (2015). Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to
1411 genotoxic agents and reduces tumor formation. Blood 126, 1462–1472. 10.1182/blood-2015-03-
1412 631606.

1413 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A.,
1414 Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-
1415 based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102,
1416 15545–15550. 10.1073/pnas.0506580102

1417 Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva,
1418 N.T., Morris, J.H., Bork, P., *et al*. (2019). STRING v11: protein-protein association networks with
1419 increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic
1420 Acids Research 47, D607–D613. 10.1093/nar/gky1131.

1421 Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W.M., Mesirov, J.P. (2007). Metagene
1422 projection for cross-platform, cross-species characterization of global transcriptional states. . Proc Natl
1423 Acad Sci U S A 104, 5959 LP – 5964. 10.1073/pnas.0701068104.

1424 van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan,
1425 T., Shakir, K., Roazen, D., Thibault, J., *et al*. (2013). From FastQ data to high-confidence variant calls:
1426 the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43, 11.10.1-11.10.33.
1427 10.1002/0471250953.bi1110s43.

1428 van Dongen, J.J.M., Langerak, A.W., Brüggemann, M., Evans, P.A.S., Hummel, M., Lavender, F.L.,
1429 Delabesse, E., Davi, F., Schuuring, E., García-Sanz, R. *et al*. (2003). Design and standardization of PCR
1430 primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene
1431 recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-
1432 CT98-3936. Leukemia 17, 2257–2317. 10.1038/sj.leu.2403202.

1433 Wang, J., Mi, J.-Q., Debernardi, A., Vitte, A.-L., Emadali, A., Meyer, J.A., Charmpi, K., Ycart, B., Callanan,
1434 M.B., Carroll, W.L., *et al*. (2015). A six gene expression signature defines aggressive subtypes and
1435 predicts outcome in childhood and adult acute lymphoblastic leukemia. Oncotarget 6, 16527–16542.
1436 10.18632/oncotarget.4113.

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

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

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

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

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

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

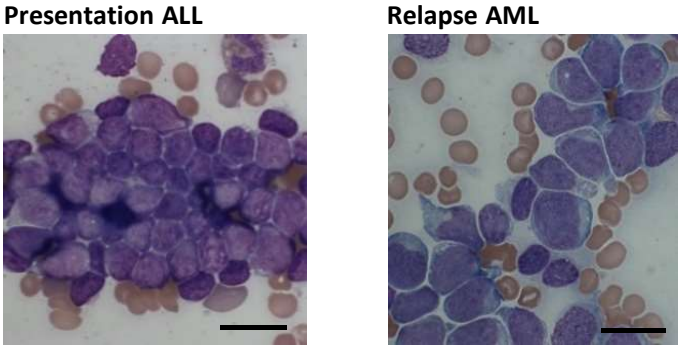
1454 Zhang, J., Jackson, A.F., Naito, T., Dose, M., Seavitt, J., Liu, F., Heller, E.J., Kashiwagi, M., Yoshida, T.,
1455 Gounari, F., et al. (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls
1456 lymphocyte development and prevents leukemogenesis. Nature Immunology 13, 86–94.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

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

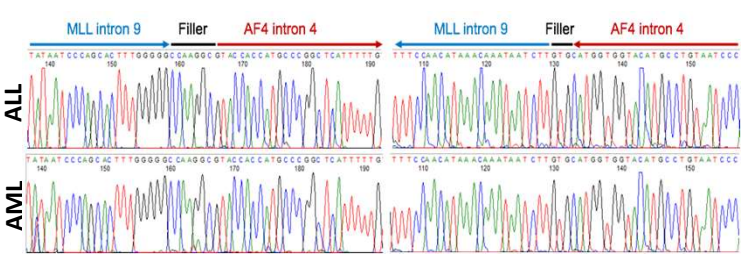
1461
1462

Figure 1

A

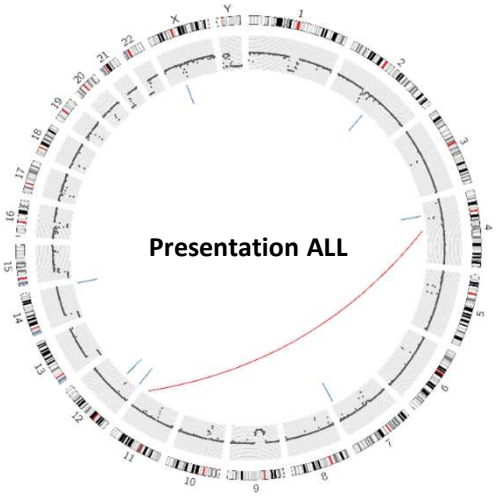


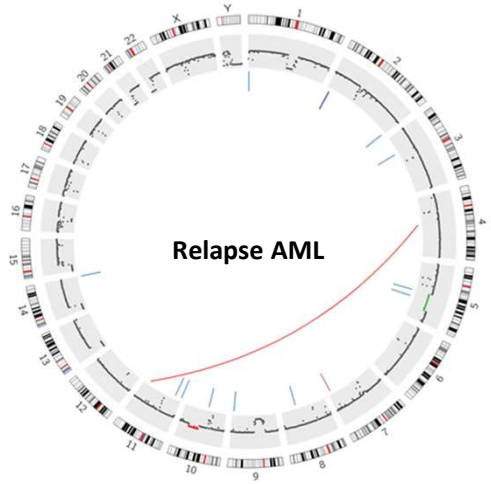
B



der11; *MLL-AF4* der4; *AF4-MLL*

C





D

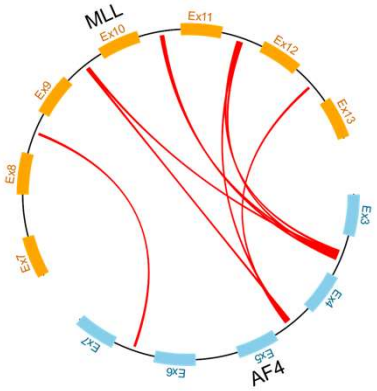
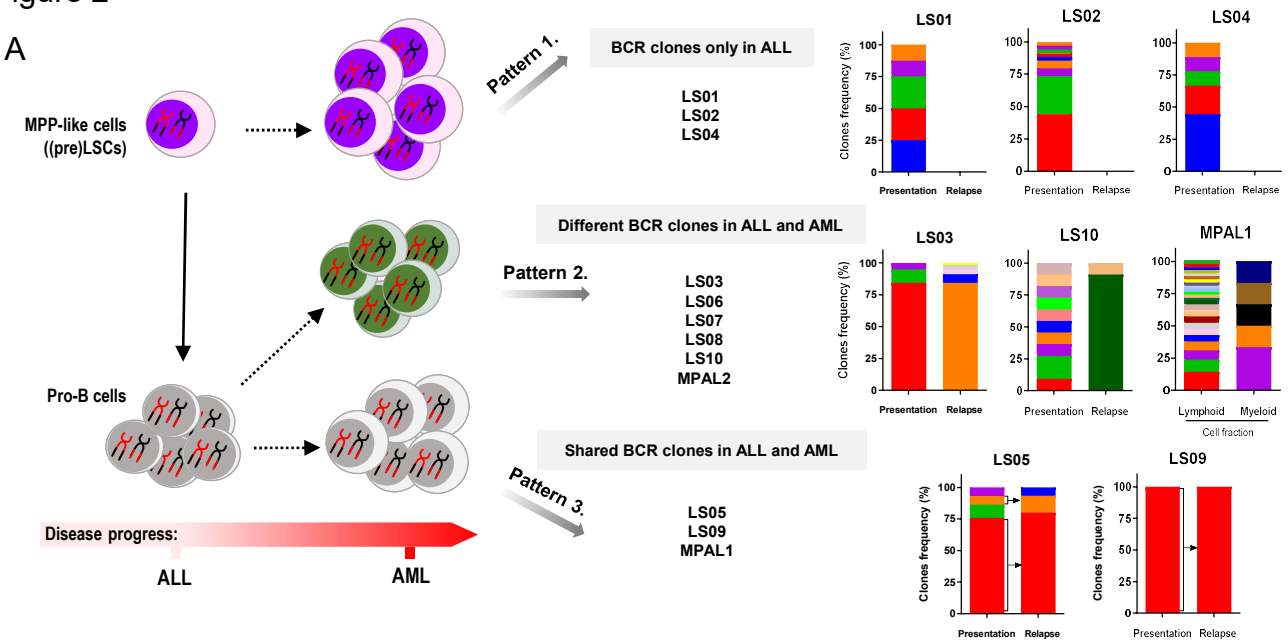
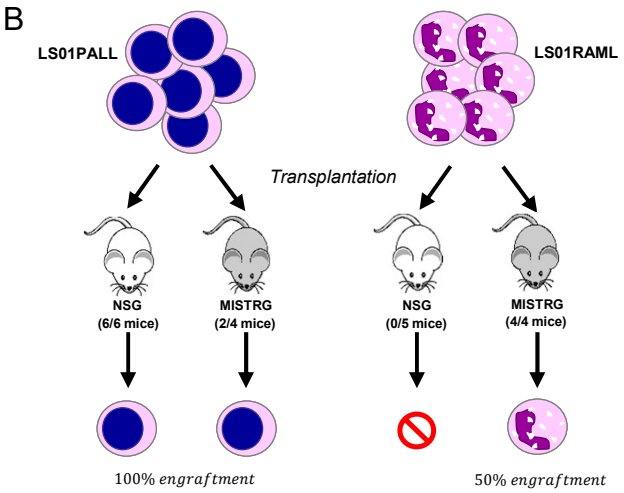


Figure 2


A
MPP-like cells
((pre)LSCs)
Pro-B cells
Disease progress:
ALL
AML
Pattern 1.
BCR clones only in ALL
LS01
LS02
LS04
Pattern 2.
Different BCR clones in ALL and AML
LS03
LS06
LS07
LS08
LS10
MPAL2
Pattern 3.
Shared BCR clones in ALL and AML
LS05
LS09
MPAL1
Clones frequency (%)
Presentation
Relapse
Lymphoid
Myeloid
Cell fraction


B
LS01PALL
LS01RAML
Transplantation
NSG
(6/6 mice)
MISTRG
(2/4 mice)
NSG
(0/5 mice)
MISTRG
(4/4 mice)
100% engraftment
50% engraftment

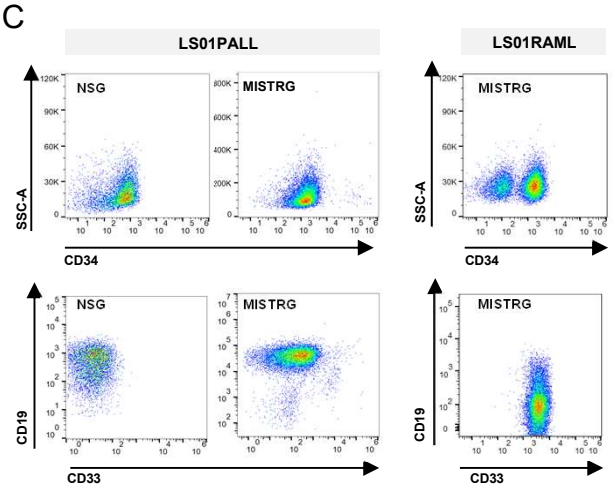

C
LS01PALL
LS01RAML
NSG
MISTRG
SSC-A
CD34
CD19
CD33

Figure 3

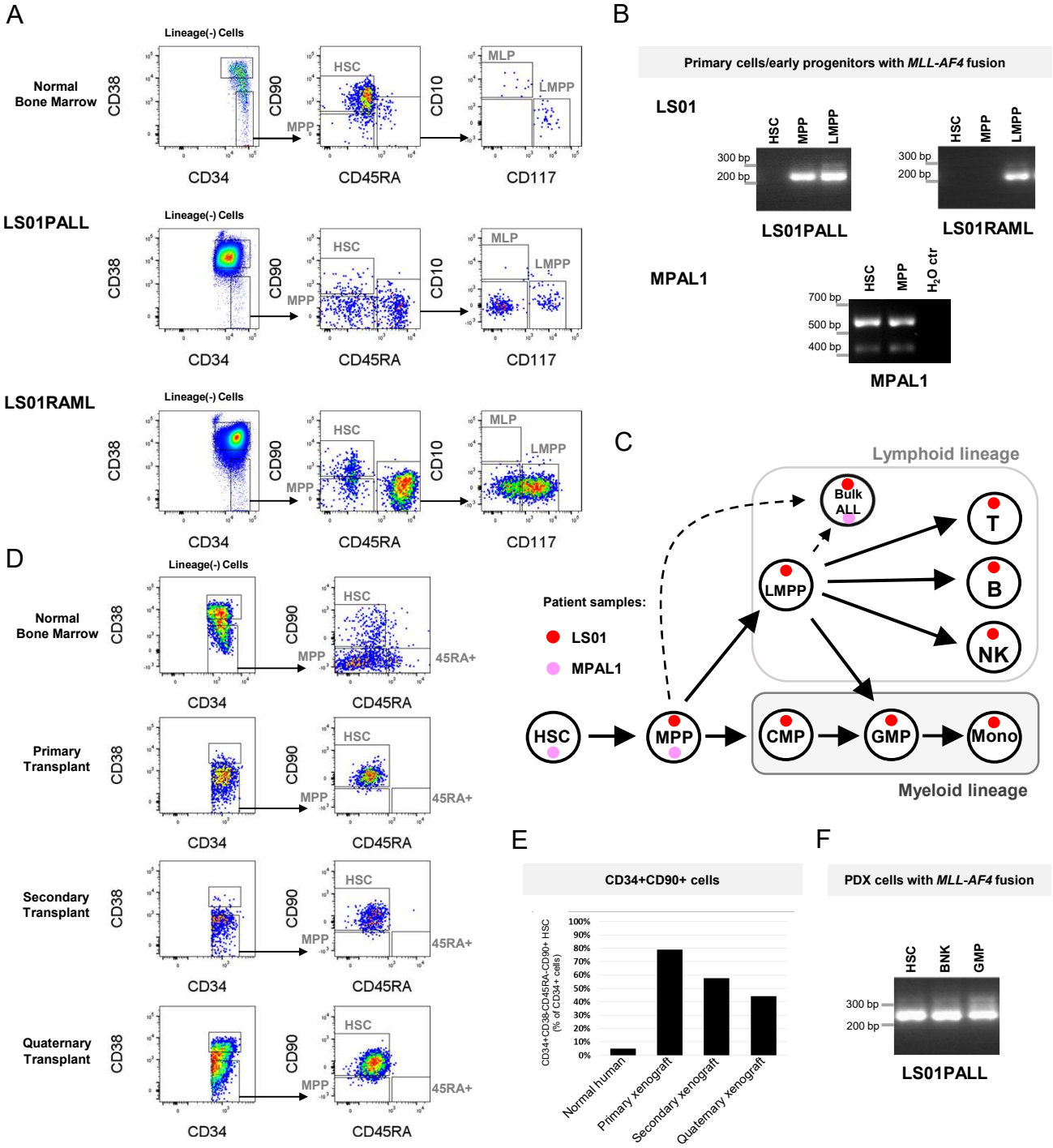

A
Lineage(-) Cells
Normal Bone Marrow
LS01PALL
LS01RAML
CD38
CD34
CD90
CD45RA
CD10
CD117
HSC
MPP
MLP
LMPP
B
Primary cells/early progenitors with MLL-AF4 fusion
LS01
HSC
MPP
LMPP
300 bp
200 bp
LS01PALL
LS01RAML
MPAL1
H2O ctr
700 bp
500 bp
400 bp
C
Lymphoid lineage
Bulk ALL
T
B
NK
LMPP
Patient samples:
LS01
MPAL1
HSC
MPP
CMP
GMP
Mono
Myeloid lineage
D
45RA+
Normal Bone Marrow
Primary Transplant
Secondary Transplant
Quaternary Transplant
E
CD34+CD90+ cells
CD34+CD38-CD45RA-CD90+ HSC (% of CD34+ cells)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Normal human
Primary xenograft
Secondary xenograft
Quaternary xenograft
F
PDX cells with MLL-AF4 fusion
HSC
BNK
GMP
300 bp
200 bp
LS01PALL

Figure 4

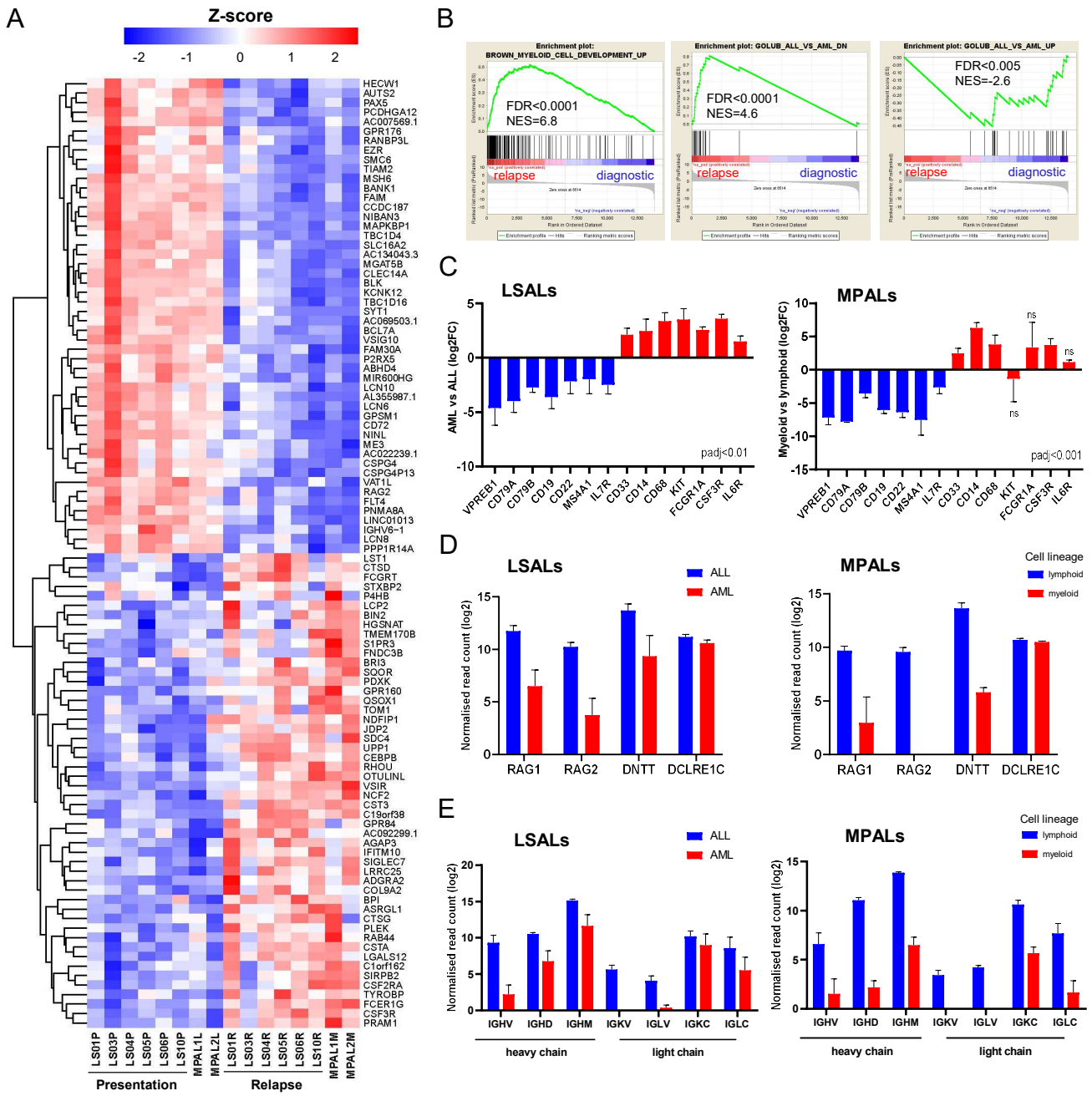
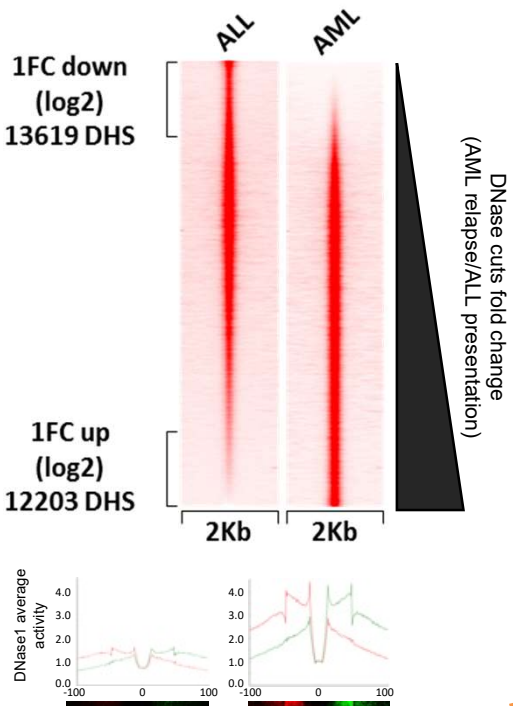
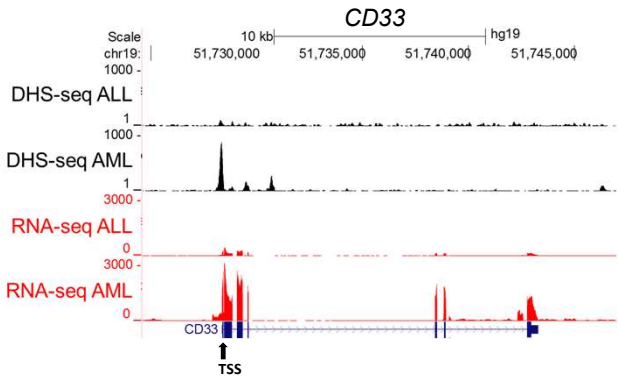

A
Z-score
-2
-1
0
1
2
LS01P
LS03P
LS04P
LS05P
LS06P
LS10P
MPAL1L
MPAL2L
LS01R
LS03R
LS04R
LS05R
LS06R
LS10R
MPAL1M
MPAL2M
Presentation
Relapse
B
Enrichment plot: BROWN_MYELOID_CELL_DEVELOPMENT_UP
FDR<0.0001
NES=6.8
relapse
diagnostic
Enrichment plot: GOLUB_ALL_VS_AML_DN
FDR<0.0001
NES=4.6
relapse
diagnostic
Enrichment plot: GOLUB_ALL_VS_AML_UP
FDR<0.005
NES=-2.6
relapse
diagnostic
C
LSALs
AML vs ALL (log2FC)
padj<0.01
MPALs
Myeloid vs lymphoid (log2FC)
padj<0.001
ns
VPREB1
CD79A
CD79B
CD19
CD22
MS4A1
IL7R
CD33
CD14
CD68
KIT
FCGR1A
CSF3R
IL6R
D
LSALs
ALL
AML
Normalised read count (log2)
RAG1
RAG2
DNTT
DCLRE1C
MPALs
Cell lineage
lymphoid
myeloid
E
LSALs
MPALs
IGHV
IGHD
IGHM
IGKV
IGLV
IGKC
IGLC
heavy chain
light chain

Figure 5

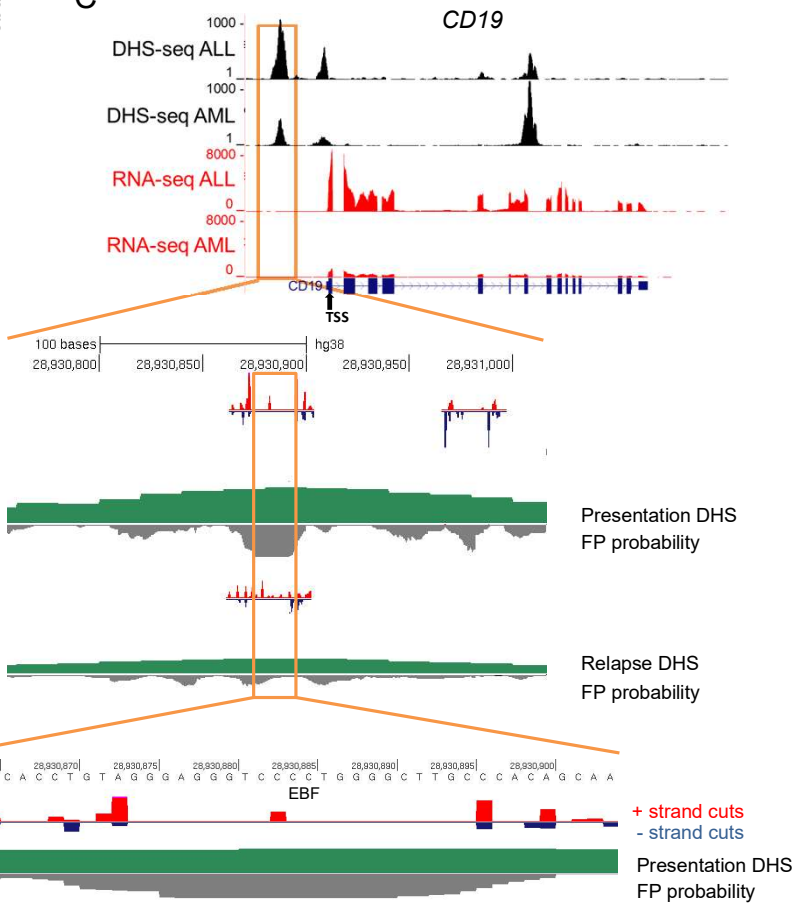
A



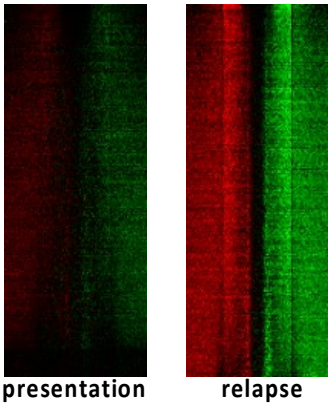
B



C



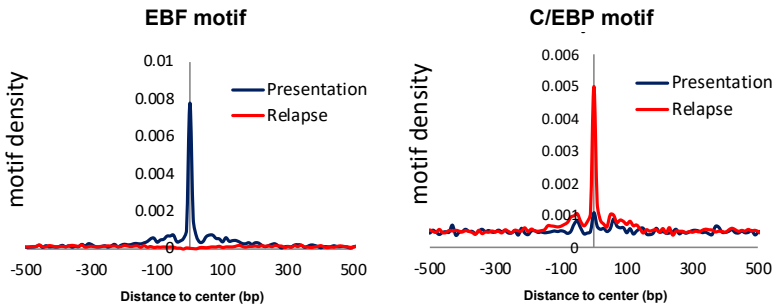
D



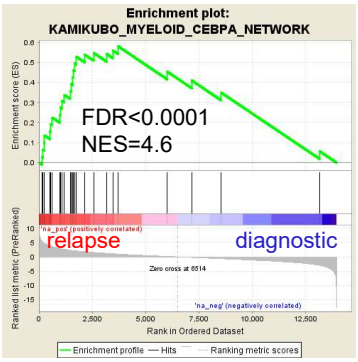
E

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

F



G



H

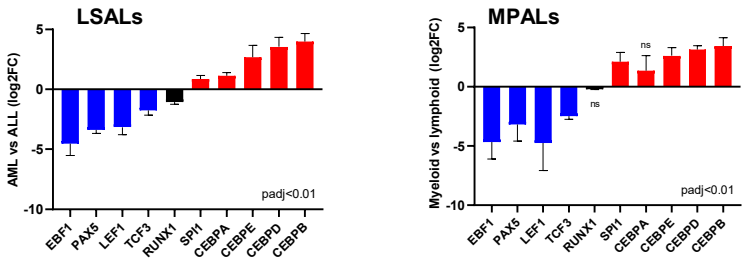


Figure 6

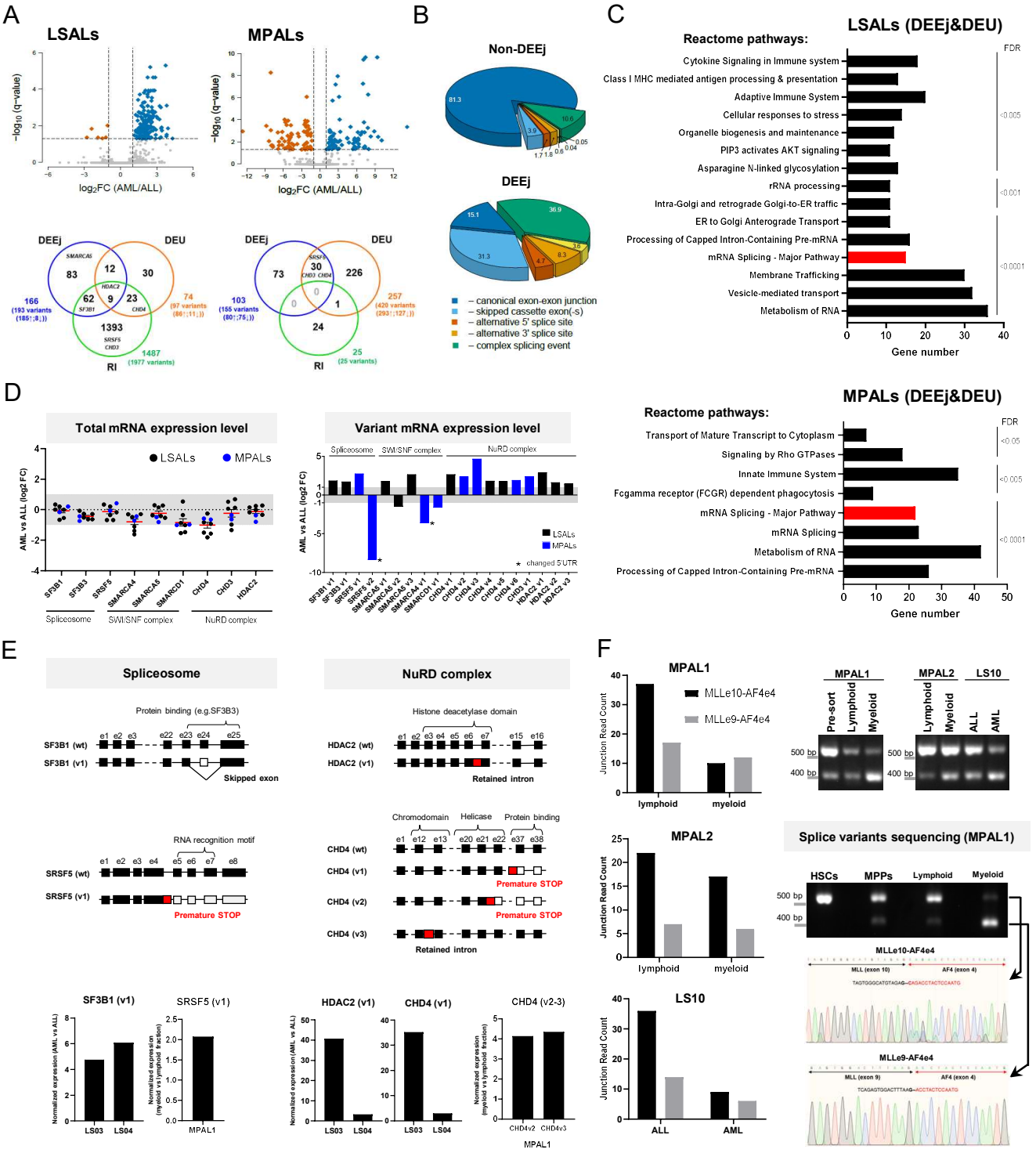

A
LSALs
MPALs
−log10 (q-value)
log2FC (AML/ALL)
DEEj
DEU
RI
B
Non-DEEj
DEEj
– canonical exon-exon junction
– skipped cassette exon(-s)
– alternative 5' splice site
– alternative 3' splice site
– complex splicing event
C
LSALs (DEEj&DEU)
Reactome pathways:
Cytokine Signaling in Immune system
Class I MHC mediated antigen processing & presentation
Adaptive Immune System
Cellular responses to stress
Organelle biogenesis and maintenance
PIP3 activates AKT signaling
Asparagine N-linked glycosylation
rRNA processing
Intra-Golgi and retrograde Golgi-to-ER traffic
ER to Golgi Anterograde Transport
Processing of Capped Intron-Containing Pre-mRNA
mRNA Splicing - Major Pathway
Membrane Trafficking
Vesicle-mediated transport
Metabolism of RNA
FDR
Gene number
MPALs (DEEj&DEU)
Transport of Mature Transcript to Cytoplasm
Signaling by Rho GTPases
Innate Immune System
Fcgamma receptor (FCGR) dependent phagocytosis
mRNA Splicing - Major Pathway
mRNA Splicing
Metabolism of RNA
Processing of Capped Intron-Containing Pre-mRNA
D
Total mRNA expression level
LSALs
MPALs
AML vs ALL (log2 FC)
Spliceosome
SWI/SNF complex
NuRD complex
Variant mRNA expression level
changed 5'UTR
E
Spliceosome
NuRD complex
Protein binding (e.g.SF3B3)
SF3B1 (wt)
SF3B1 (v1)
Skipped exon
Histone deacetylase domain
HDAC2 (wt)
HDAC2 (v1)
Retained intron
RNA recognition motif
SRSF5 (wt)
SRSF5 (v1)
Premature STOP
Chromodomain
Helicase
Protein binding
CHD4 (wt)
CHD4 (v1)
CHD4 (v2)
CHD4 (v3)
Normalized expression (AML vs ALL)
LS03
LS04
MPAL1
HDAC2 (v1)
CHD4 (v1)
CHD4 (v2-3)
CHD4v2
CHD4v3
F
MPAL1
MLLe10-AF4e4
MLLe9-AF4e4
Junction Read Count
lymphoid
myeloid
Pre-sort
Lymphoid
Myeloid
MPAL2
LS10
ALL
AML
500 bp
400 bp
Splice variants sequencing (MPAL1)
HSCs
MPPs
MLL (exon 10)
AF4 (exon 4)
MLL (exon 9)

Figure 7

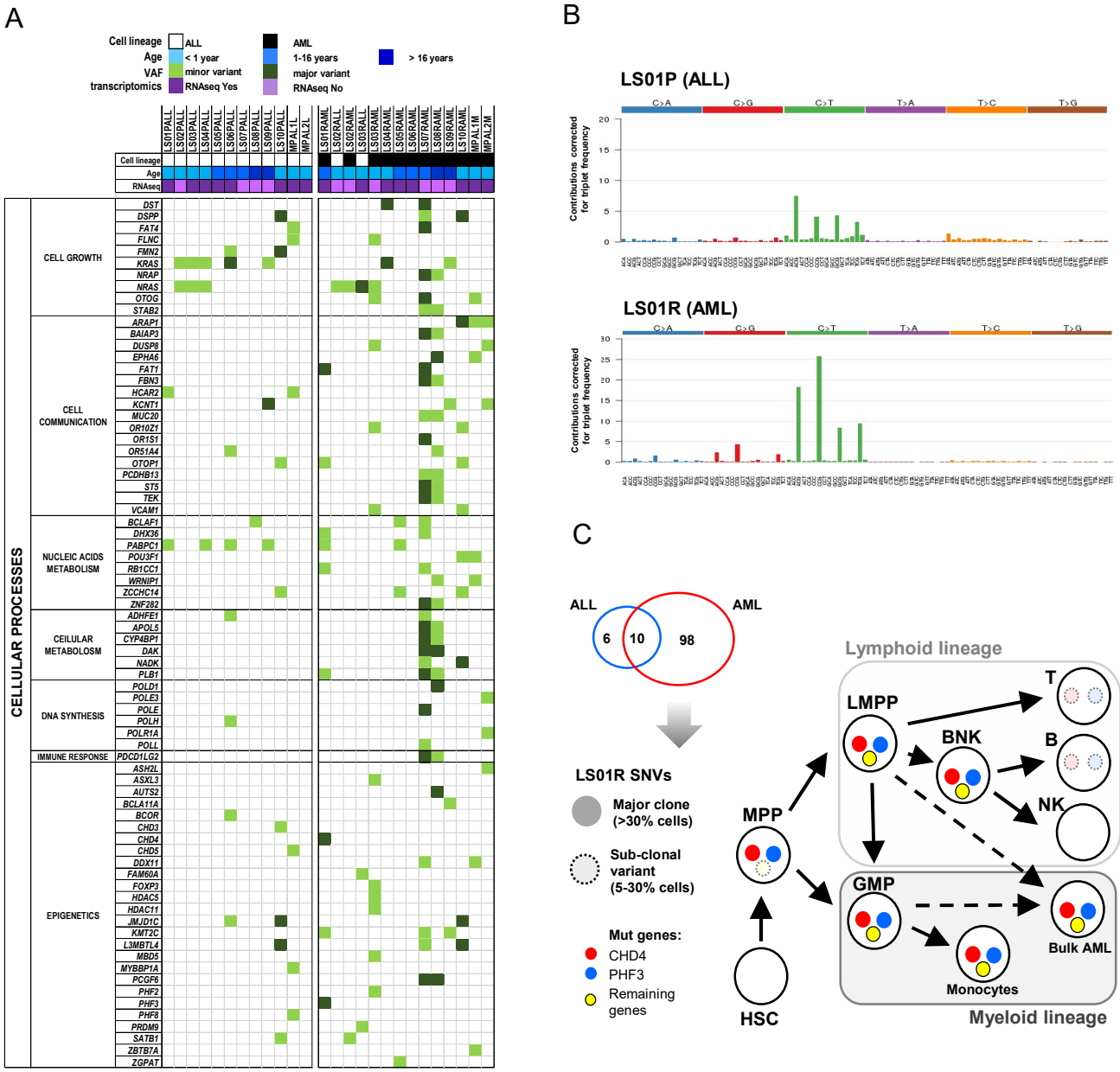

A
Cell lineage
ALL
AML
Age
< 1 year
1-16 years
> 16 years
VAF
minor variant
major variant
transcriptomics
RNAseq Yes
RNAseq No
LS01PALL
LS02PALL
LS03PALL
LS04PALL
LS05PALL
LS06PALL
LS07PALL
LS08PALL
LS09PALL
LS10PALL
MPAL1L
MPAL2L
LS01RAML
LS02RAML
LS02RAML
LS03RAML
LS03RAML
LS04RAML
LS05RAML
LS06RAML
LS07RAML
LS08RAML
LS09RAML
LS10RAML
MPAL1M
MPAL2M
CELLULAR PROCESSES
CELL GROWTH
DST
DSPP
FAT4
FLNC
FMN2
KRAS
NRAP
NRAS
OTOG
STAB2
CELL COMMUNICATION
ARAP1
BAIAP3
DUSP8
EPHA6
FAT1
FBN3
HCAR2
KCNT1
MUC20
OR10Z1
OR1S1
OR51A4
OTOP1
PCDHB13
ST5
TEK
VCAM1
NUCLEIC ACIDS METABOLISM
BCLAF1
DHX36
PABPC1
POU3F1
RB1CC1
WRNIP1
ZCCHC14
ZNF282
CEILULAR METABOLOSM
ADHFE1
APOL5
CYP4BP1
DAK
NADK
PLB1
DNA SYNTHESIS
POLD1
POLE3
POLE
POLH
POLR1A
POLL
IMMUNE RESPONSE
PDCD1LG2
EPIGENETICS
ASH2L
ASXL3
AUTS2
BCLA11A
BCOR
CHD3
CHD4
CHD5
DDX11
FAM60A
FOXP3
HDAC5
HDAC11
JMJD1C
KMT2C
L3MBTL4
MBD5
MYBBP1A
PCGF6
PHF2
PHF3
PHF8
PRDM9
SATB1
ZBTB7A
ZGPAT
B
LS01P (ALL)
C>A
C>G
C>T
T>A
T>C
T>G
Contributions corrected for triplet frequency
LS01R (AML)
C>A
C>G
C>T
T>A
T>C
T>G
Contributions corrected for triplet frequency
C
ALL
AML
6
10
98
LS01R SNVs
Major clone (>30% cells)
Sub-clonal variant (5-30% cells)
Mut genes:
CHD4
PHF3
Remaining genes
Lymphoid lineage
LMPP
T
BNK
B
NK
MPP
GMP
Bulk AML
Monocytes
HSC
Myeloid lineage

Figure 8

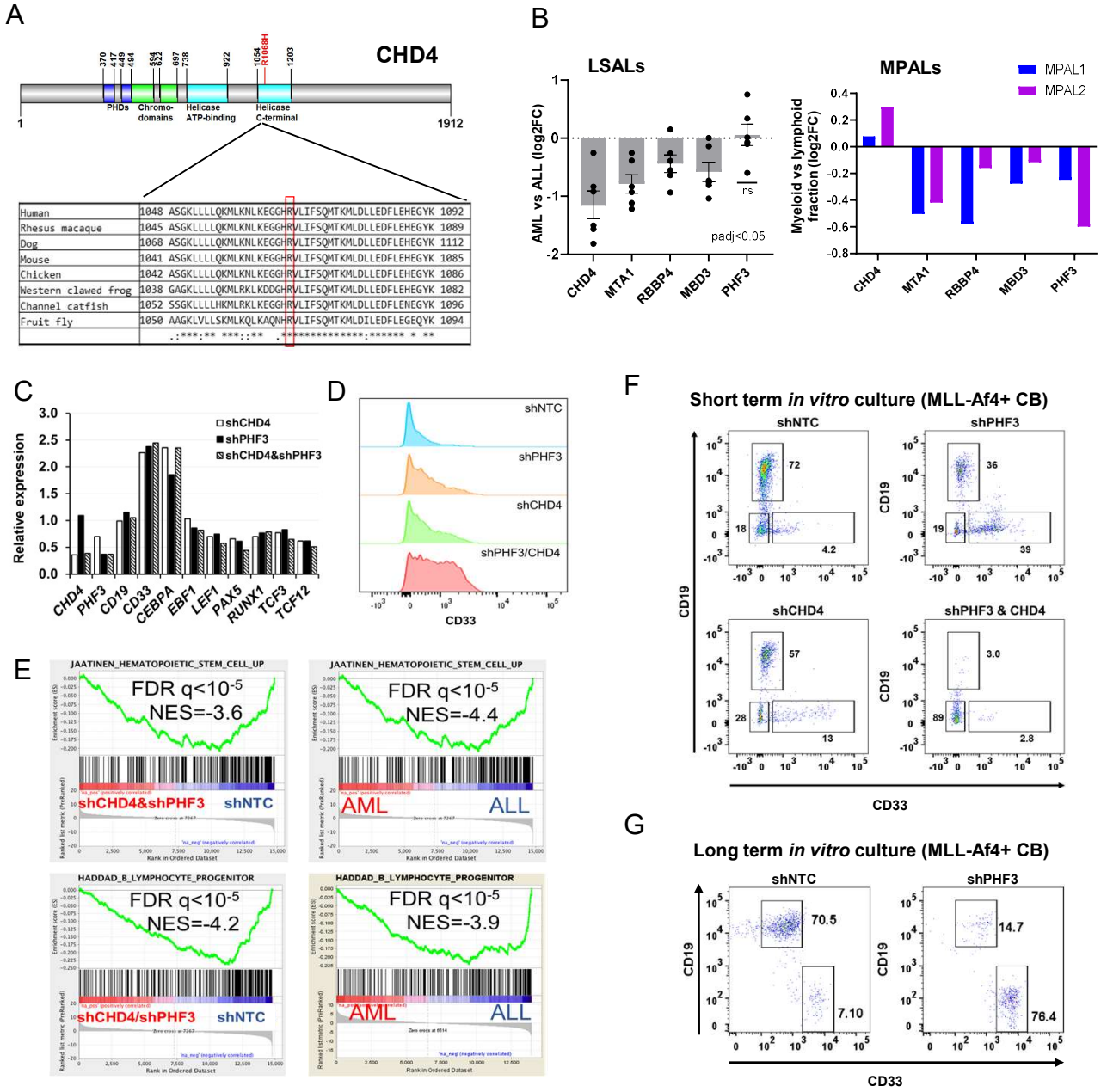

A
CHD4
370
417
449
494
584
622
697
738
922
1054
R1068H
1203
PHDs
Chromo-domains
Helicase ATP-binding
Helicase C-terminal
1
1912
Human 1048 ASGKLLLLQKMLKNLKEGGHRVLIFSQMTKMLDLLEDFLEHEGYK 1092
Rhesus macaque 1045 ASGKLLLLQKMLKNLKEGGHRVLIFSQMTKMLDLLEDFLEHEGYK 1089
Dog 1068 ASGKLLLLQKMLKNLKEGGHRVLIFSQMTKMLDLLEDFLEHEGYK 1112
Mouse 1041 ASGKLLLLQKMLKNLKEGGHRVLIFSQMTKMLDLLEDFLEHEGYK 1085
Chicken 1042 ASGKLLLLQKMLKNLKEGGHRVLIFSQMTKMLDLLEDFLEHEGYK 1086
Western clawed frog 1038 GAGKLLLLQKMLRKLKDDGHRVLIFSQMTKMLDLLEDFLEHEGYK 1082
Channel catfish 1052 SSGKLLLLHKMLRKLKEGGHRVLIFSQMTKMLDLLEDFLENEGYK 1096
Fruit fly 1050 AAGKLVLLSKMLKQLKAQNHRVLIFSQMTKMLDILEDFLEGEQYK 1094
B
LSALs
MPALs
MPAL1
MPAL2
AML vs ALL (log2FC)
Myeloid vs lymphoid fraction (log2FC)
ns
padj<0.05
CHD4
MTA1
RBBP4
MBD3
PHF3
C
Relative expression
shCHD4
shPHF3
shCHD4&shPHF3
CHD4
PHF3
CD19
CD33
CEBPA
EBF1
LEF1
PAX5
RUNX1
TCF3
TCF12
D
shNTC
shPHF3
shCHD4
shPHF3/CHD4
CD33
E
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP
FDR q<10-5
NES=-3.6
shCHD4&shPHF3
shNTC
FDR q<10-5
NES=-4.4
AML
ALL
HADDAD_B_LYMPHOCYTE_PROGENITOR
FDR q<10-5
NES=-4.2
shCHD4/shPHF3
shNTC
FDR q<10-5
NES=-3.9
AML
ALL
Rank in Ordered Dataset
F
Short term in vitro culture (MLL-Af4+ CB)
shNTC
shPHF3
shCHD4
shPHF3 & CHD4
CD19
CD33
72
18
4.2
36
19
39
57
28
13
3.0
89
2.8
G
Long term in vitro culture (MLL-Af4+ CB)
shNTC
shPHF3
CD19
CD33
70.5
7.10
14.7
76.4

Figure S1

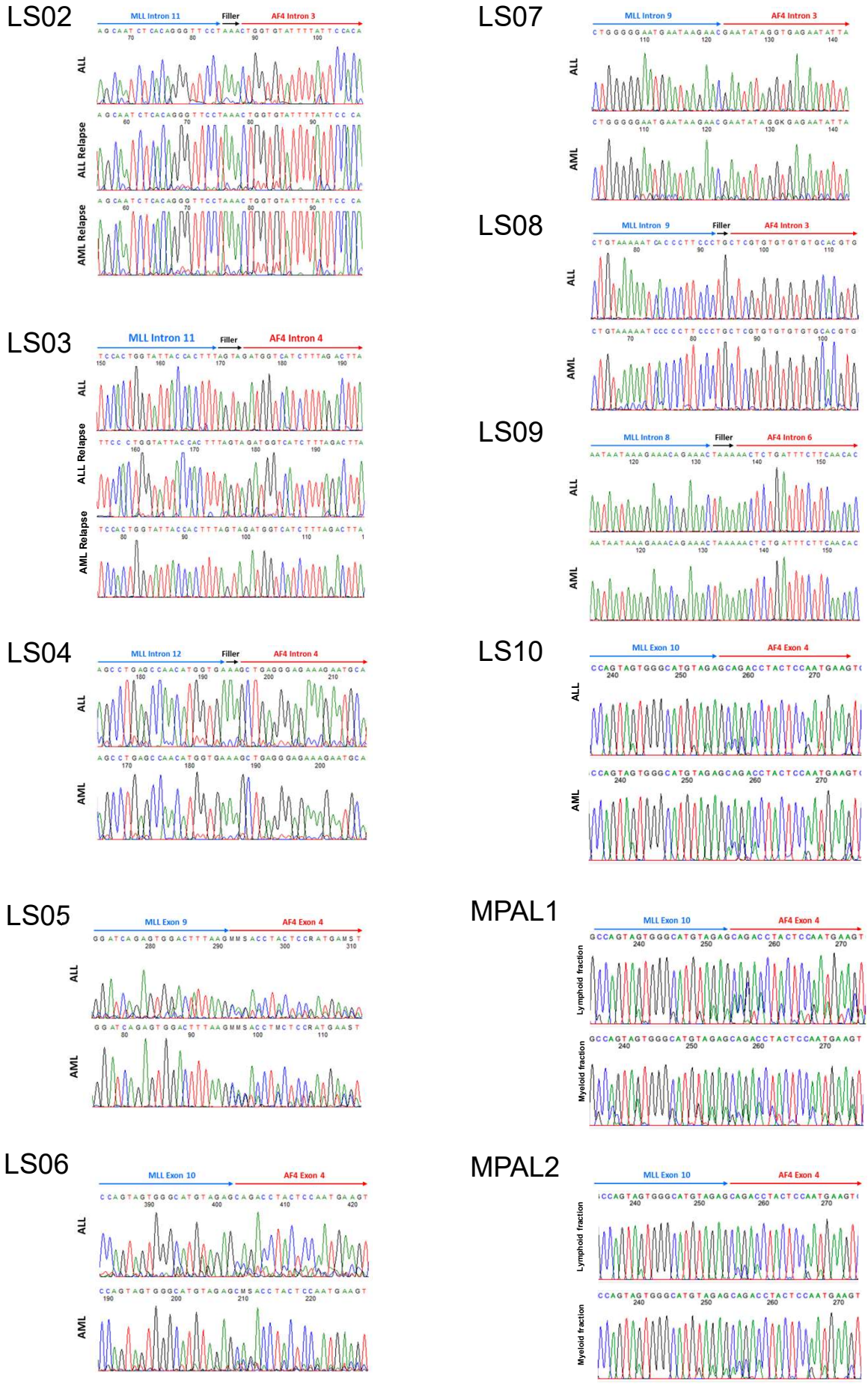

LS02
MLL Intron 11
Filler
AF4 Intron 3
ALL
ALL Relapse
AML Relapse
LS03
MLL Intron 11
Filler
AF4 Intron 4
ALL
ALL Relapse
AML Relapse
LS04
MLL Intron 12
Filler
AF4 Intron 4
ALL
AML
LS05
MLL Exon 9
AF4 Exon 4
ALL
AML
LS06
MLL Exon 10
AF4 Exon 4
ALL
AML
LS07
MLL Intron 9
AF4 Intron 3
ALL
AML
LS08
MLL Intron 9
Filler
AF4 Intron 3
ALL
AML
LS09
MLL Intron 8
Filler
AF4 Intron 6
ALL
AML
LS10
MLL Exon 10
AF4 Exon 4
ALL
AML
MPAL1
MLL Exon 10
AF4 Exon 4
Lymphoid fraction
Myeloid fraction
MPAL2
MLL Exon 10
AF4 Exon 4
Lymphoid fraction
Myeloid fraction

Figure S2

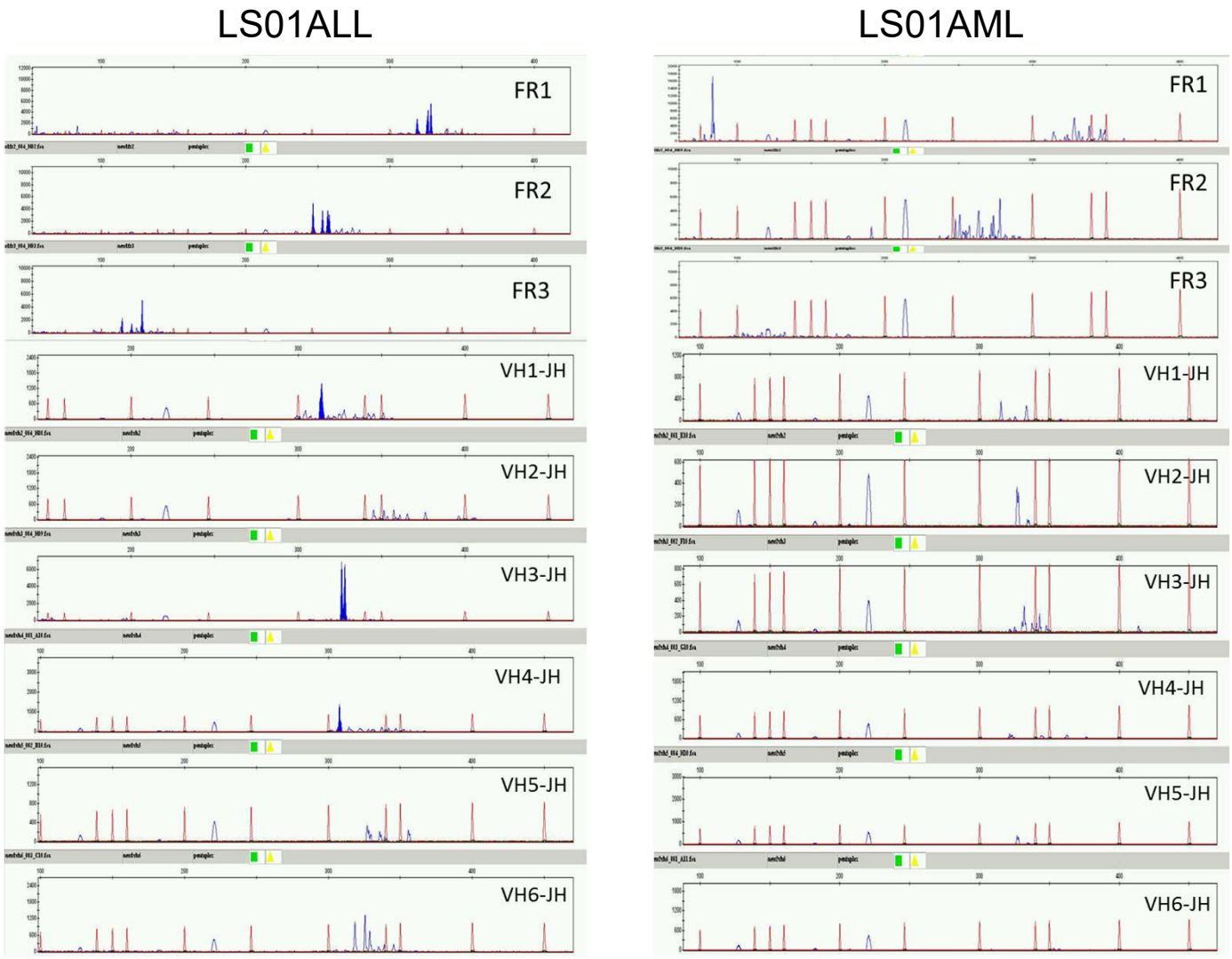

LS01ALL
FR1
FR2
FR3
VH1-JH
VH2-JH
VH3-JH
VH4-JH
VH5-JH
VH6-JH
LS01AML
FR1
FR2
FR3
VH1-JH
VH2-JH
VH3-JH
VH4-JH
VH5-JH
VH6-JH

Figure S3

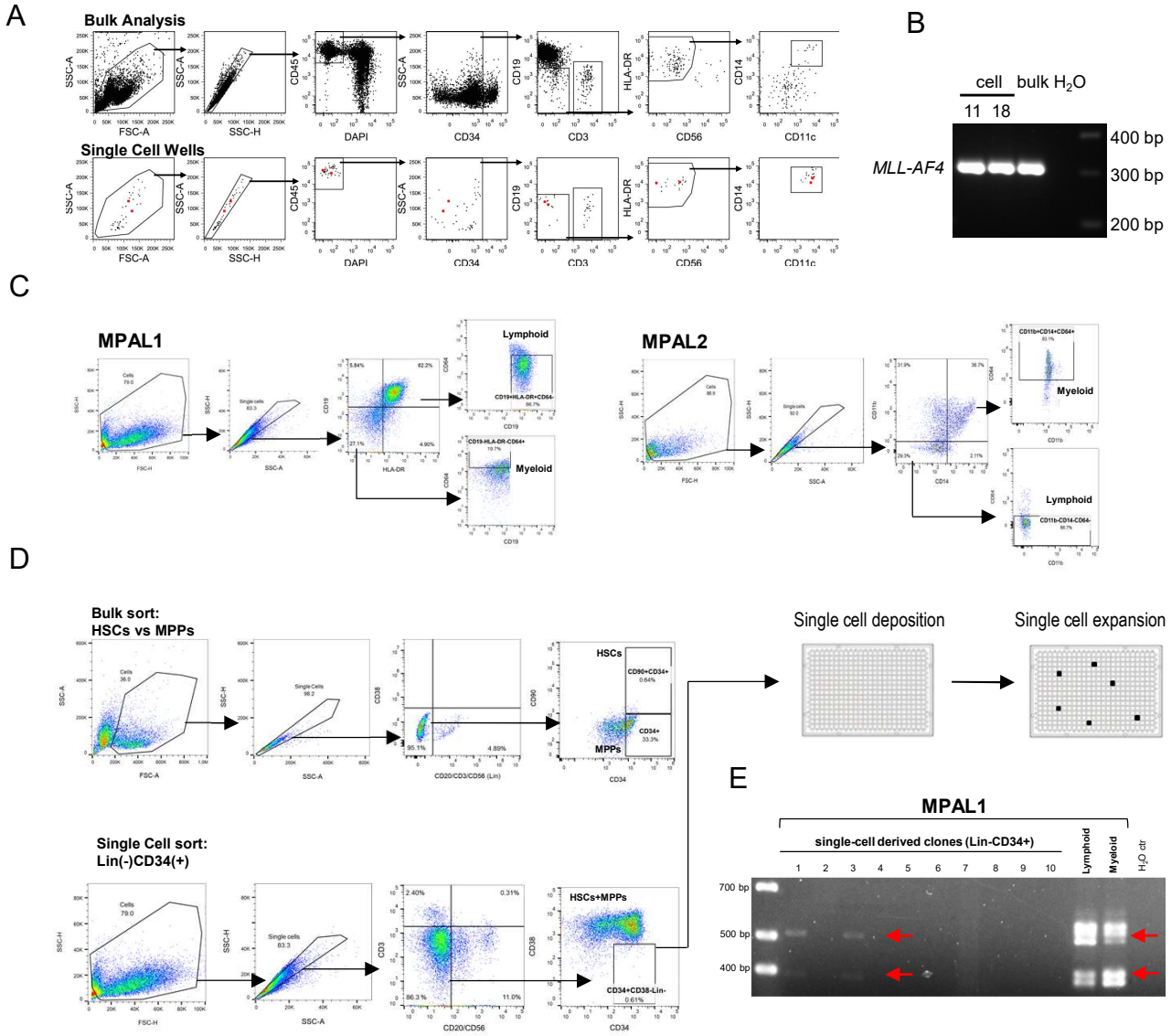

A
Bulk Analysis
Single Cell Wells
FSC-A
SSC-A
SSC-H
DAPI
CD45
CD34
CD19
CD3
HLA-DR
CD56
CD14
CD11c
B
cell
bulk H2O
11 18
MLL-AF4
400 bp
300 bp
200 bp
C
MPAL1
MPAL2
Lymphoid
Myeloid
D
Bulk sort:
HSCs vs MPPs
HSCs
MPPs
Single Cell sort:
Lin(-)CD34(+)
HSCs+MPPs
Single cell deposition
Single cell expansion
E
MPAL1
single-cell derived clones (Lin-CD34+)
1 2 3 4 5 6 7 8 9 10
Lymphoid
Myeloid
H2O ctr
700 bp
500 bp
400 bp

Figure S4

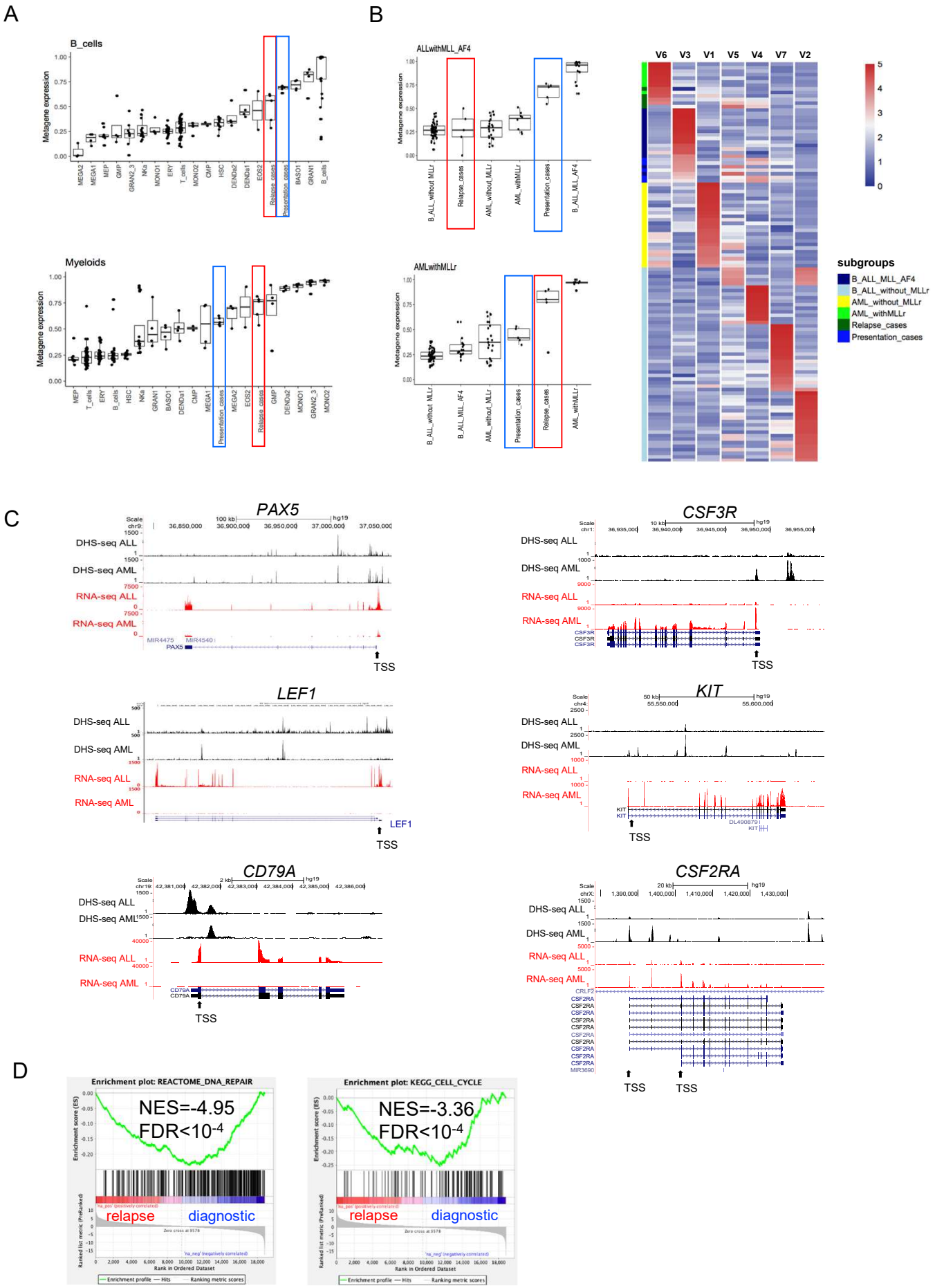
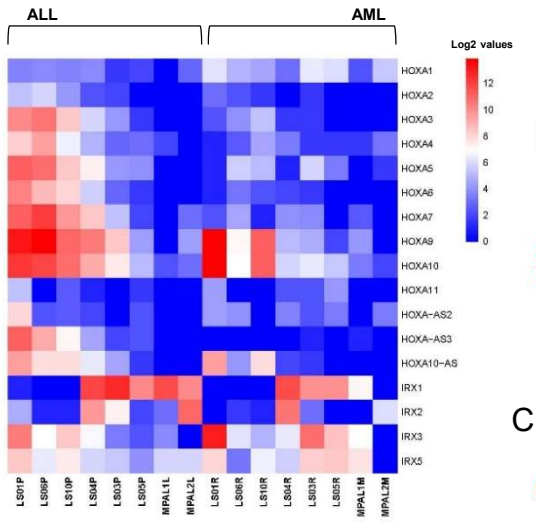
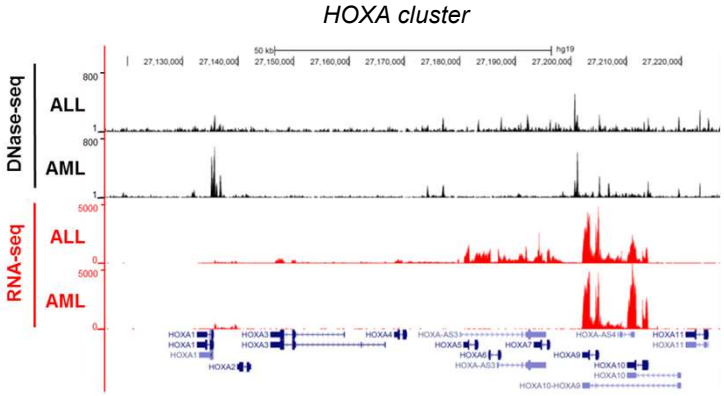

A
B_cells
Metagene expression
Relapse_cases
Presentation_cases
Myeloids
B
ALLwithMLL_AF4
AMLwithMLLr
V6 V3 V1 V5 V4 V7 V2
subgroups
B_ALL_MLL_AF4
B_ALL_without_MLLr
AML_without_MLLr
AML_withMLLr
Relapse_cases
Presentation_cases
C
PAX5
DHS-seq ALL
DHS-seq AML
RNA-seq ALL
RNA-seq AML
TSS
CSF3R
LEF1
KIT
CD79A
CSF2RA
D
Enrichment plot: REACTOME_DNA_REPAIR
NES=-4.95
FDR<10-4
relapse
diagnostic
Enrichment plot: KEGG_CELL_CYCLE
NES=-3.36
FDR<10-4

Figure S5

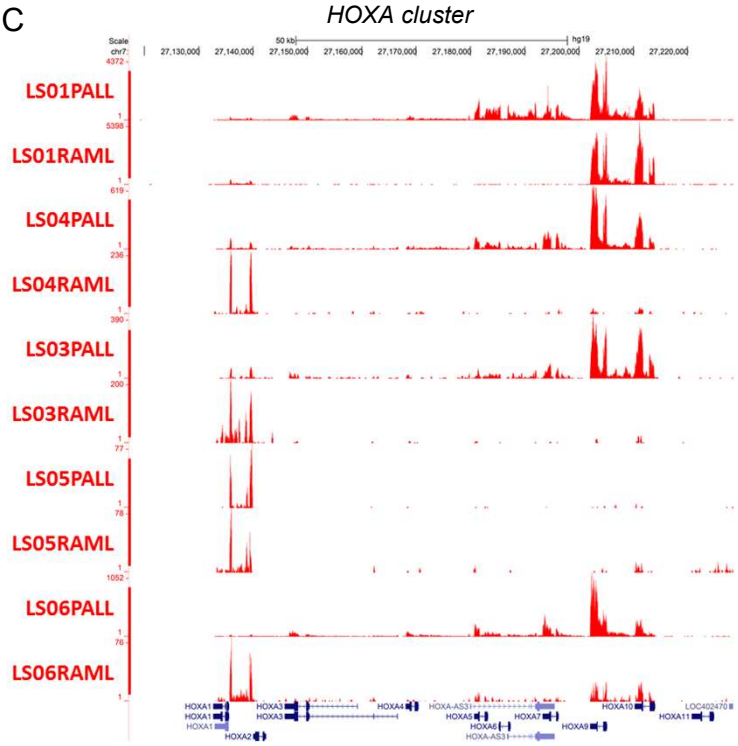
A



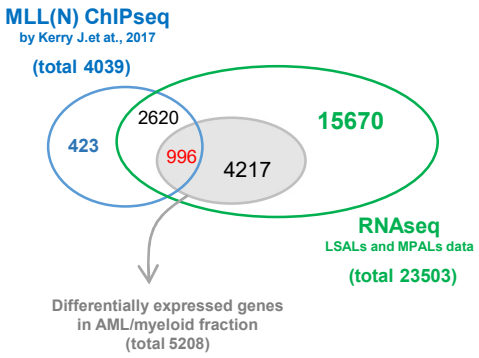
B

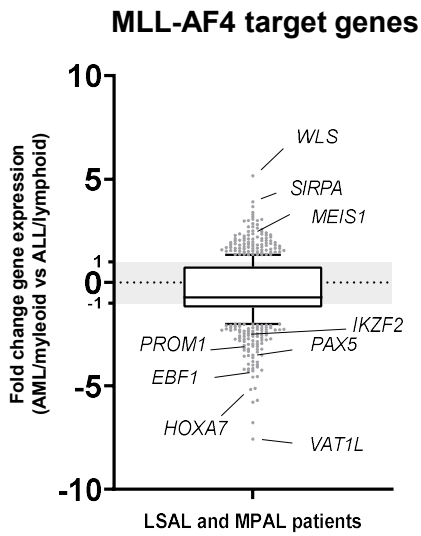


C



D





E

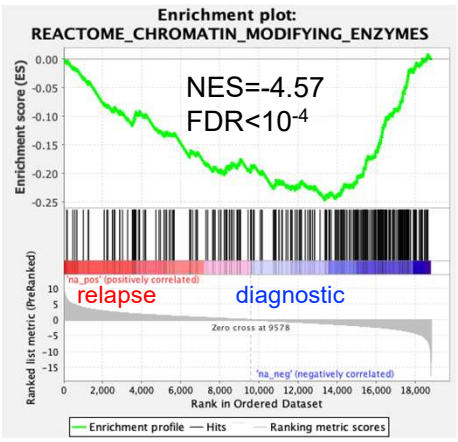
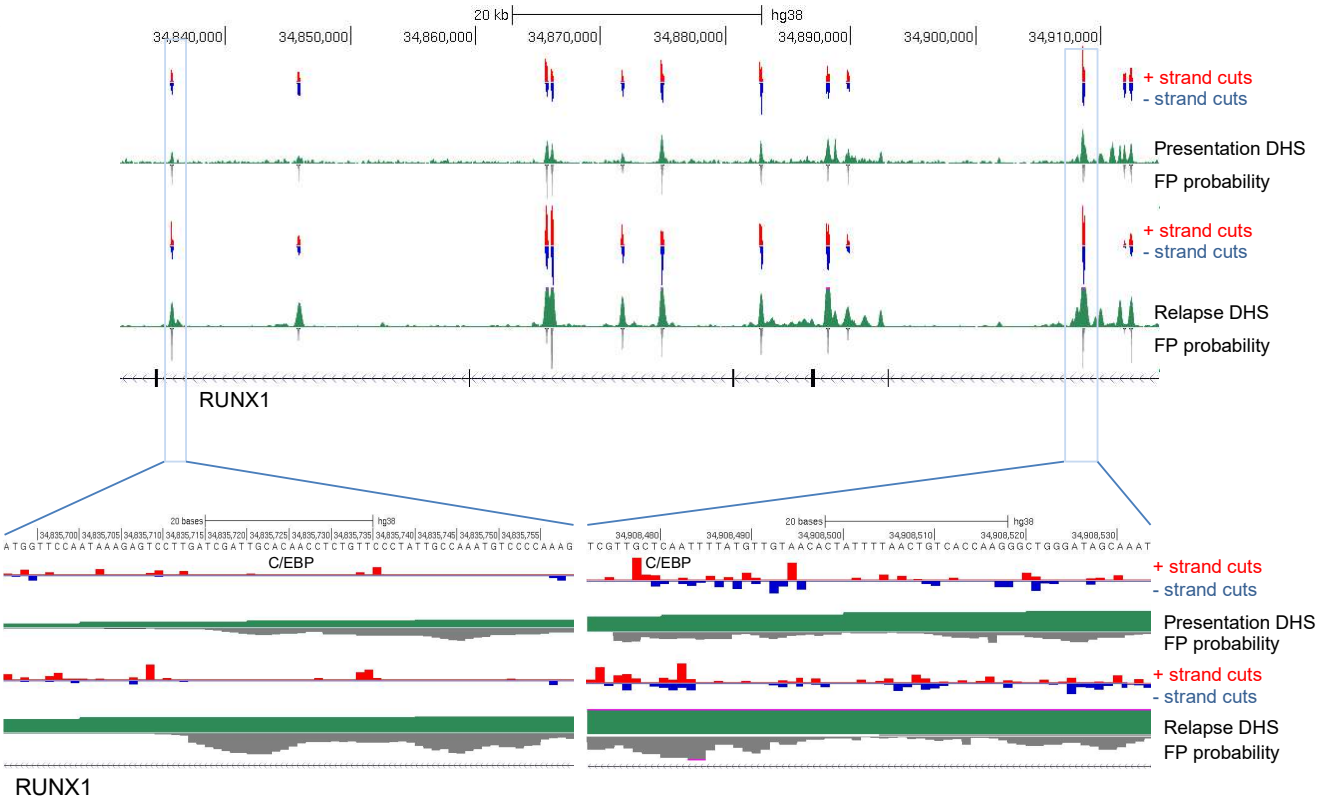
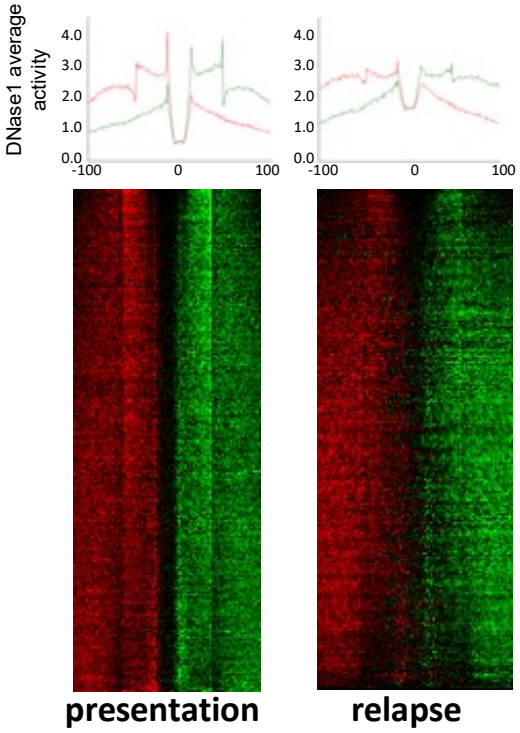


Figure S6

A



B



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

Figure S7

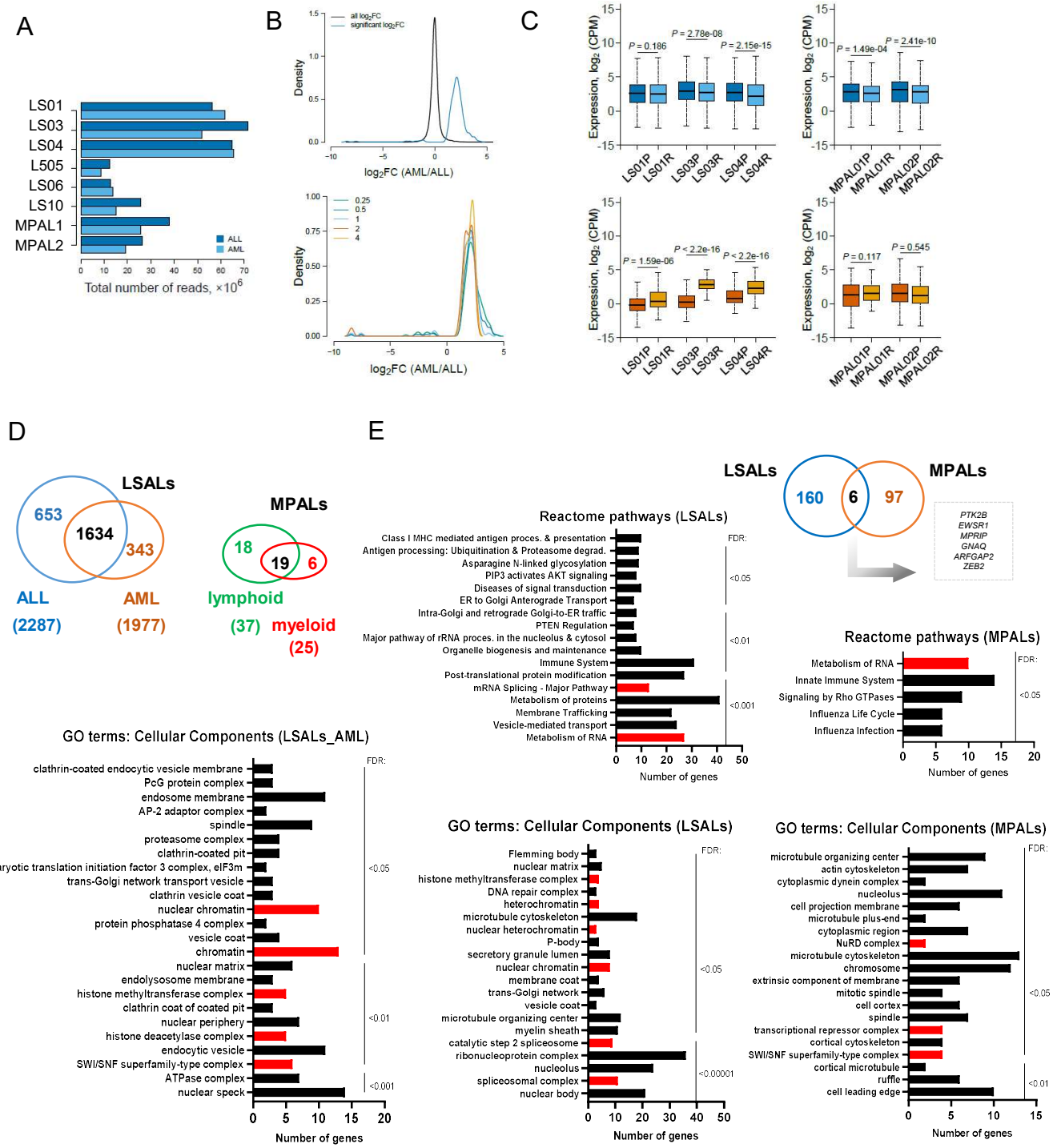
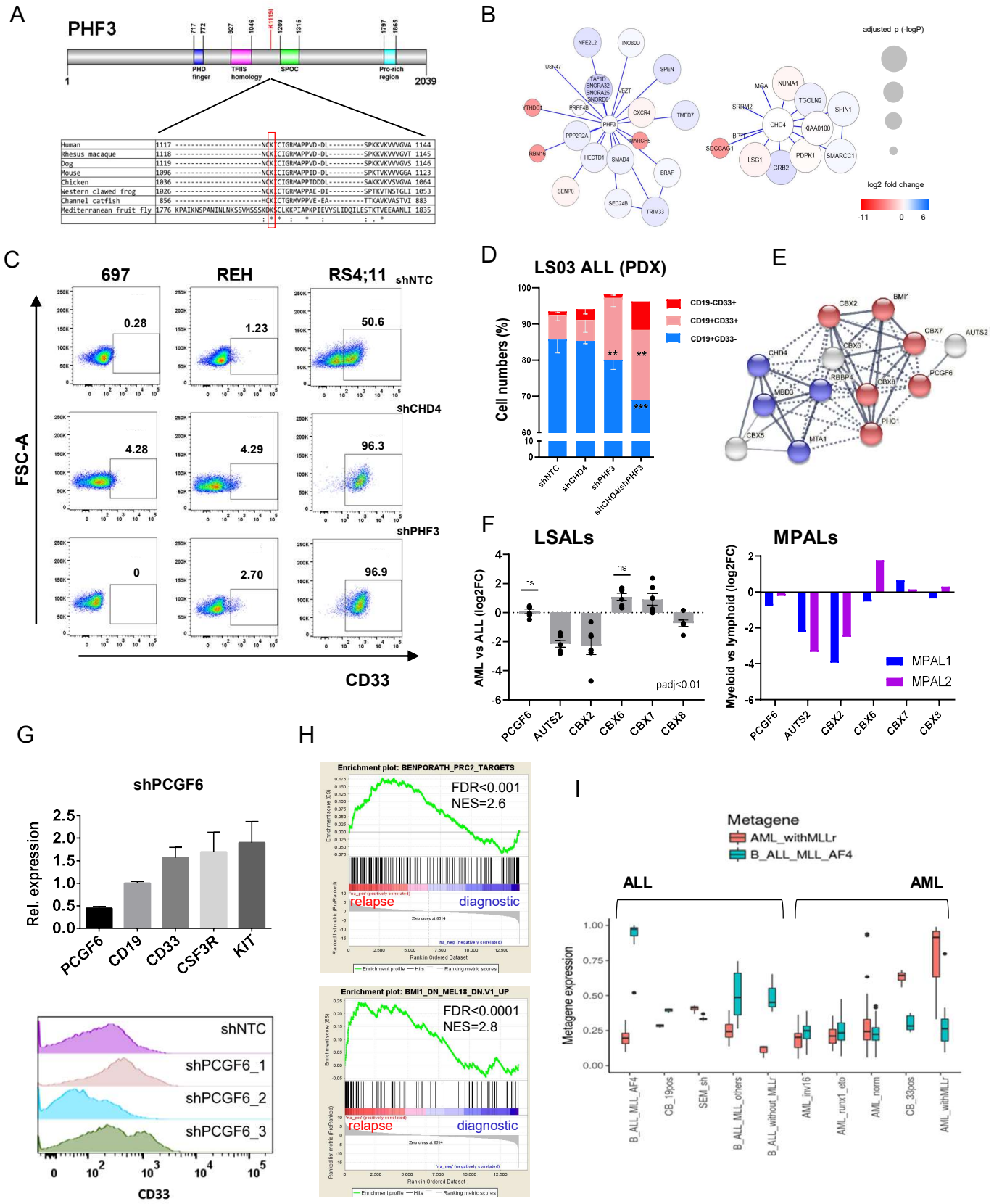

A
LS01
LS03
LS04
L505
LS06
LS10
MPAL1
MPAL2
ALL
AML
Total number of reads, ×10^6
B
all log2FC
significant log2FC
Density
log2FC (AML/ALL)
0.25
0.5
1
2
4
C
Expression, log2 (CPM)
P = 0.186
P = 2.78e-08
P = 2.15e-15
P = 1.49e-04
P = 2.41e-10
P = 1.59e-06
P < 2.2e-16
P < 2.2e-16
P = 0.117
P = 0.545
LS01P
LS01R
LS03P
LS03R
LS04P
LS04R
MPAL01P
MPAL01R
MPAL02P
MPAL02R
D
LSALs
653
1634
343
ALL
(2287)
AML
(1977)
MPALs
18
19
6
lymphoid
(37)
myeloid
(25)
GO terms: Cellular Components (LSALs_AML)
FDR:
clathrin-coated endocytic vesicle membrane
PcG protein complex
endosome membrane
AP-2 adaptor complex
spindle
proteasome complex
clathrin-coated pit
eukaryotic translation initiation factor 3 complex, eIF3m
trans-Golgi network transport vesicle
clathrin vesicle coat
nuclear chromatin
protein phosphatase 4 complex
vesicle coat
chromatin
nuclear matrix
endolysosome membrane
histone methyltransferase complex
clathrin coat of coated pit
nuclear periphery
histone deacetylase complex
endocytic vesicle
SWI/SNF superfamily-type complex
ATPase complex
nuclear speck
<0.05
<0.01
<0.001
Number of genes
E
Reactome pathways (LSALs)
Class I MHC mediated antigen proces. & presentation
Antigen processing: Ubiquitination & Proteasome degrad.
Asparagine N-linked glycosylation
PIP3 activates AKT signaling
Diseases of signal transduction
ER to Golgi Anterograde Transport
Intra-Golgi and retrograde Golgi-to-ER traffic
PTEN Regulation
Major pathway of rRNA proces. in the nucleolus & cytosol
Organelle biogenesis and maintenance
Immune System
Post-translational protein modification
mRNA Splicing - Major Pathway
Metabolism of proteins
Membrane Trafficking
Vesicle-mediated transport
Metabolism of RNA
<0.05
<0.01
<0.001
Number of genes
LSALs
MPALs
160
6
97
PTK2B
EWSR1
MPRIP
GNAQ
ARFGAP2
ZEB2
Reactome pathways (MPALs)
Metabolism of RNA
Innate Immune System
Signaling by Rho GTPases
Influenza Life Cycle
Influenza Infection
<0.05
Number of genes
GO terms: Cellular Components (LSALs)
Flemming body
nuclear matrix
histone methyltransferase complex
DNA repair complex
heterochromatin
microtubule cytoskeleton
nuclear heterochromatin
P-body
secretory granule lumen
nuclear chromatin
membrane coat
trans-Golgi network
vesicle coat
microtubule organizing center
myelin sheath
catalytic step 2 spliceosome
ribonucleoprotein complex
nucleolus
spliceosomal complex
nuclear body
<0.05
<0.00001
Number of genes
GO terms: Cellular Components (MPALs)
microtubule organizing center
actin cytoskeleton
cytoplasmic dynein complex
nucleolus
cell projection membrane
microtubule plus-end
cytoplasmic region
NuRD complex
microtubule cytoskeleton
chromosome
extrinsic component of membrane
mitotic spindle
cell cortex
spindle
transcriptional repressor complex
cortical cytoskeleton
SWI/SNF superfamily-type complex
cortical microtubule
ruffle
cell leading edge
<0.05
<0.01
Number of genes

Figure S8


A
PHF3
PHD finger
TFIIS homology
SPOC
Pro-rich region
1
2039
Human
Rhesus macaque
Dog
Mouse
Chicken
Western clawed frog
Channel catfish
Mediterranean fruit fly
B
adjusted p (-logP)
log2 fold change
-11
0
6
C
697
REH
RS4;11
shNTC
shCHD4
shPHF3
FSC-A
CD33
0.28
1.23
50.6
4.28
4.29
96.3
0
2.70
96.9
D
LS03 ALL (PDX)
Cell numbers (%)
CD19-CD33+
CD19+CD33+
CD19+CD33-
shNTC
shCHD4
shPHF3
shCHD4/shPHF3
E
F
LSALs
MPALs
AML vs ALL (log2FC)
Myeloid vs lymphoid (log2FC)
ns
padj<0.01
MPAL1
MPAL2
PCGF6
AUTS2
CBX2
CBX6
CBX7
CBX8
G
shPCGF6
Rel. expression
PCGF6
CD19
CD33
CSF3R
KIT
shNTC
shPCGF6_1
shPCGF6_2
shPCGF6_3
CD33
H
Enrichment plot: BENPORATH_PRC2_TARGETS
FDR<0.001
NES=2.6
relapse
diagnostic
Enrichment plot: BMI1_DN_MEL18_DN.V1_UP
FDR<0.0001
NES=2.8
relapse
diagnostic
I
Metagene
AML_withMLLr
B_ALL_MLL_AF4
ALL
AML
Metagene expression
B_ALL_MLL_AF4
CB_19pos
SEM_sh
B_ALL_MLL_others
B_ALL_without_MLLr
AML_inv16
AML_runx1_eto
AML_norm
CB_33pos
AML_withMLLr