Endogenous retroviruses are a source of oncogenic enhancers in acute myeloid leukemia

Özgen Deniz[1,2]*, Mamataz Ahmed[1,2], Christopher D. Todd[1,2,5], Ana Rio-Machin[2,3], Mark A. Dawson[4], Miguel R. Branco[1,2]*

[1]Blizard Institute, Barts and The London School of Medicine and Dentistry, QMUL, London E1 2AT, UK
[2]Centre for Genomic Health, Life Sciences Institute, QMUL, London E1 2AB, UK
[3]Centre for Haemato-Oncology, Barts Cancer Institute, QMUL, London EC1M 6BQ, UK
[4]Cancer Research Division, Peter MacCallum Cancer Center, Melbourne, Victoria 3002, Australia
[5]Present address: Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK
*Corresponding authors. Email: o.deniz@qmul.ac.uk, m.branco@qmul.ac.uk

Abstract

Endogenous retroviruses (ERVs) and other transposons can act as tissue-specific regulators of gene expression in *cis*, with potential to affect biological processes. In cancer, epigenetic alterations and transcription factor misregulation may uncover the regulatory potential of typically repressed ERVs, which could contribute to tumour evolution and progression. Here, we asked whether transposons help to rewire oncogenic transcriptional circuits in acute myeloid leukemia (AML). Using epigenomic data from both primary cells and cell lines, we have identified six ERV families that are frequently found in an open chromatin state in AML when compared to differentiated healthy myeloid cells. A subset of these AML-associated ERVs harbour enhancer-specific histone modifications, and are bound by hematopoiesis-associated transcription factors that play key roles in haematopoiesis and in the pathogenesis of AML. Using CRISPR-mediated genetic editing and simultaneous epigenetic silencing of multiple ERV copies, we have established causal links between ERV deregulation in AML and expression changes of adjacent genes. Finally, we show that deletion and epigenetic silencing of an ERV, through modulating expression of *APOC1* gene, leads to growth suppression by inducing apoptosis in leukemia cell lines. Our results suggest that ERV derepression provides an additional layer of gene regulation in AML that may be exploited by cancer cells to help drive oncogenic phenotypes.

Introduction

Transposable elements (TEs) act primarily as selfish genetic entities that have expanded to make up around half of the human genome (Lander et al., 2001). Although they are commonly repressed by host epigenetic mechanisms, TEs are frequently co-opted as tissue-specific gene regulatory elements, such as alternative gene promoters and distal enhancers (Chuong et al., 2017). Notably, up to 34% of transcription factor (TF) binding sites in the human genome are derived from TEs (Sundaram et al., 2014), with an overrepresentation from long terminal repeat (LTR) families, which include endogenous retroviruses (ERVs). LTRs frequently recombine, leaving the majority of ERV elements as intact solitary LTRs with potential gene regulatory activity (Belshaw et al., 2007; Thomas et al., 2018). In line with this notion, genome-wide assays have documented that numerous LTR sequences carry hallmarks of active regulatory elements (Chuong et al., 2016; 2013; Fuentes et al., 2018; Jacques et al., 2013; Kunarso et al., 2010; Lynch et al., 2011; Pontis et al., 2019; Sundaram et al., 2014). In a few instances, loss-of-function experiments have provided compelling evidence of LTR co-option for host gene regulation and cellular function in hematopoiesis (Pi et al., 2010), innate immunity (Chuong et al., 2016), pregnancy (Ferreira et al., 2016) and fertility (Flemr et al., 2013).

Various studies have documented widespread epigenetic and transcriptional deregulation of TEs in several cancer types, raising the possibility that TE-derived regulatory elements may be exploited to promote tumorigenesis (Babaian and Mager, 2016; Burns, 2017). Indeed, activation of LTR-based promoters initiates cancer-specific chimeric transcripts in Hodgkin lymphoma, melanoma and diffuse large B-cell lymphoma, among others (Babaian and Mager, 2016; Edginton-White et al., 2019; Jang et al., 2019). These so-called onco-exaptation events were recently shown to be highly prevalent in multiple cancer types (Jang et al., 2019). However, studies to date have been largely centered on LTR promoter activity. As the function of TEs as enhancers is more challenging to discern, such cases remain largely unexplored in malignancies, with one notable exception of a recent study showing that systematic perturbation of LTR5_Hs family in the human embryonic carcinoma affects host gene transcription over long genomic ranges (Fuentes et al., 2018).

Acute myeloid leukemia (AML) is a highly heterogeneous clonal malignancy, characterized by the combinations of distinct driver mutations (Cancer Genome Atlas Research Network et al., 2013). Notably, over 70% AML patients have a genetic abnormality in one epigenetic modifier (Papaemmanuil et al., 2016). Interestingly, altered chromatin landscape, including DNA methylation (Figueroa et al., 2010), histone modifications and chromatin accessibility (Assi et al., 2018; Yi et al., 2019)

establishes different AML subtypes linked to specific genetic alterations. Consequently, widespread disruption of epigenetic patterns is a characteristic feature of AML, which may stimulate TE activation as regulatory elements. In this respect, AML is an ideal model system to explore onco-exaptation of TEs and determine to what extent such events are important for maintaining the expression of genes that promote malignancy progression and evolution.

Here we have identified six ERV/LTR families with putative regulatory potential in AML, bearing enhancer-like epigenomic signatures and binding of TFs that play key roles in haematopoiesis and in the pathogenesis of AML. Deletion of selected ERV copies and epigenetic inactivation of an entire ERV family demonstrated their roles as gene regulators. Remarkably, we found that genetic and epigenetic perturbation of a single LTR copy leads to impaired cell growth by modulating expression of *APOC1* gene, suggesting that the activation of this particular LTR has a driving role in leukemia cell phenotype.

Results

Identification of putative AML-specific regulatory TEs

In order to assess whether cellular transformation in AML unmasked intrinsic regulatory activity of TEs, we identified putative regulatory TEs by focusing on repeat families that were frequently associated with open chromatin in AML. To this end, we analysed DNase-seq data from 32 AML samples generated by the Blueprint epigenome project (Yi et al., 2019), and compared them with data from differentiated myeloid cells (macrophages, monocytes) from the same consortium (Figure 1A). Additionally, we generated our own DNase-seq data from three commonly used AML cell lines with different genetic and cytogenetic backgrounds: HL-60, MOLM-13 and OCI-AML3. We overlapped DNase hypersensitive sites (DHSs) with the Repeatmasker annotation and compared the DHS frequency at each repeat family with random controls (Supplementary Table 1). We identified 12 repeat families that were enriched for DHS-associated copies in at least one of the AML cell lines and in 10% or more of the AML samples (Figure 1B). Five of these families were highly enriched across all samples, including macrophages and monocytes, as well as mobilized CD34+ cells (data from the Roadmap epigenomics project), suggesting little cell specificity. The remaining seven families displayed more variability between AML samples and, notably, tended to display little to no enrichment in differentiated myeloid cells (Figure 1B). Nearly all families were also DHS-enriched in CD34+ cells, suggesting an association with a stem cell state. The exception was the LTR2B family, whose DHS enrichment appeared to be AML-specific. Analysis of an independent

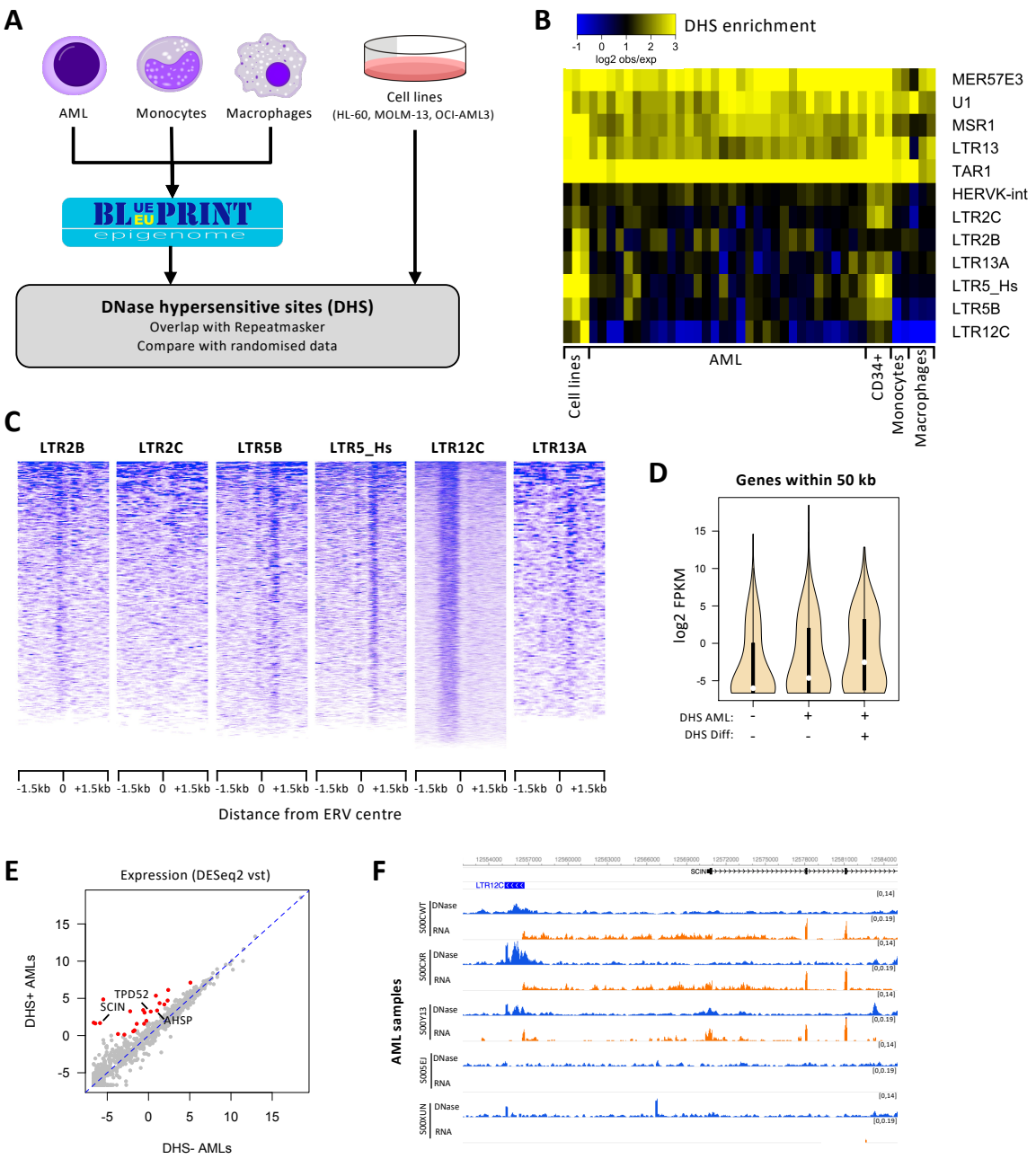

Figure 1. ERVs with regulatory potential are activated in AML and associated with gene expression differences (hematopoietic cells credit: A. Rad CC-BY-SA-3.0). **A.** Schematic of the strategy to detect repeat families associated with open chromatin in AML. **B.** Heatmap of the observed/expected enrichment for DHSs in selected repeat families. Cell lines are presented in the following order: HL-60, MOLM-13, OCI-AML3. **C.** DNase-seq profile across all elements of each AML DHS-associated repeat (A-DAR) families in OCI-AML3. **D.** Gene expression violin plots for genes within 50 kb of A-DARs, with or without overlapping DHS in AML and/or differentiated cells. **E.** For each gene lying near an A-DAR element, we compared its expression in AML samples where the respective ERV has a DHS, versus AML samples where the DHS is absent. Expression values were normalized using the variance stabilizing transformation (vst; log2 scale) in DESeq2. Highlighted are genes with >4-fold difference and vst>0. **F.** Example of a gene (*SCIN*) that displays a strict correlation between its expression and the presence of a DHS peak at a nearby LTR12C element in different AML samples.

dataset of 32 AML samples from the Bonifer lab (Assi et al., 2018) confirmed the DHS enrichment at all of the above families, and identified additional weaker associations, including with several *Alu* subfamilies (Supplementary Figure 1A). For stringency, we focused on families that were DHS-enriched in both datasets, all of which are LTRs from ERVs: LTR2B, LTR2C, LTR5B, LTR5_Hs, LTR12C, LTR13A. We excluded the internal portion of HERVK (HERVK-int) because its enrichment was largely due to the fragmented nature of the Repeatmasker annotation, and because its LTRs (LTR5B, LTR5_Hs) displayed much higher DNase-seq signal than the internal portion (Supplementary Figure 1B). We will collectively refer to the six selected ERV families as 'AML DHS-associated repeats' (A-DARs). The oldest A-DARs (LTR5B, LTR13A) date back to the common ancestor between hominoids and old world monkeys, whereas the youngest (LTR5_Hs) are human-specific (Hubley et al., 2016).

The DNase-seq profiles across each ERV displayed a consistent pattern for elements of the same family in AML cell lines (less evident for LTR2C), suggestive of TF binding events within these ERVs (Figure 1C displays OCI-AML3 profiles). This pattern was also notable in primary AML cells, albeit variable between samples (Supplementary Figure 1C), reflecting the heterogeneity of this disease. Out of a total of 4,811 A-DAR elements, 80-661 (median 263) overlapped a DHS in AML samples from the Blueprint dataset. These values were higher in the Assi et al dataset (223-1,349; median 508) possibly due to technical differences. As heterogeneity in AML is partly driven by genetics, we hypothesized that variation in DHS frequency at A-DARs could reflect distinct mutational profiles. To test this, we measured inter-sample correlations in the DHS patterns of A-DARs, which revealed distinct clusters associated to the mutational profile in AML patient samples (Supplementary Figure 2A). Although there was no strict association with particular AML subtypes, we found that samples with *NPM1* mutations were better inter-correlated than those without (Supplementary Figure 2B). The same was true for samples with *FLT3*-ITD and *DNMT3A* mutations, which frequently co-occur with *NPM1* mutations, as well as those with *CEBPA* mutations (Supplementary Figure 2B). Specific mutations may therefore help to establish particular patterns of ERV activation in AML, although other characteristics of the malignancy are also likely to affect them.

A-DAR chromatin status correlates with nearby gene expression

To test whether A-DARs were associated with gene activation, we analysed matching DNase-seq and RNA-seq data from the Blueprint consortium. ERVs can affect the expression of proximal genes, but also act at a distance via long-range interactions in 3D space (Fuentes et al., 2018; Raviram et al., 2018). However, long-range

interactions display substantial cell specificity, namely within the hematopoietic system (Javierre et al., 2016). Given the heterogeneity between AML samples and the lack of matching Hi-C data, we focused our analysis on genes within 50 kb of an ERV from the selected families. Genes lying close to A-DAR elements with DHS in 2 or more AML samples displayed higher expression levels than those lying close to A-DAR elements without DHS (Figure 1D). This was more pronounced for ERVs with DHS also present in differentiated cells. Such bulk correlations are only suggestive of a regulatory role of ERVs and need to be formally tested for a causal link, as we have previously shown (Todd et al., 2019). Nonetheless, we found ERVs with strong supporting evidence for their regulatory activity, as the expression levels of their vicinity genes were >4-fold higher in AML samples with DHS at a given ERV, versus those without (Figure 1E). This included a strict correlation between chromatin accessibility at a LTR12C element and the expression of the *SCIN* gene (Figure 1F), which has a mild predictive value for AML prognosis (Z.-H. Zhang et al., 2018). *TPD52* and *AHSP* expression, which were correlated with the presence of DHS at a LTR12C and LTR5B element respectively, have also been linked to AML prognosis (Ha et al., 2019; Zhu et al., 2017). These data suggest that at least some A-DAR elements gain gene regulatory activity in AML, which correlates with disease outcomes.

A-DARs bear signatures of enhancer elements

DNase hypersensitivity is associated with both active gene promoters and distal enhancers. LTR12C elements, for example, were previously shown to frequently act as alternative gene promoters in different cell types, including hepatocellular carcinoma (Hashimoto et al., 2015) and cell lines treated with DNMT and HDAC inhibitors (Brocks et al., 2017). In contrast, LTR5_Hs (HERV-K) elements appear to mainly act as distal enhancer elements in embryonic carcinoma cells and embryonic stem cells (Fuentes et al., 2018; Pontis et al., 2019). We therefore aimed to establish whether A-DARs could act as promoters and/or enhancers in AML.

To test for gene promoter activity, we performed *de novo* transcriptome assembly in AML samples and differentiated myeloid cells, and calculated the number of spliced transcripts for which the transcriptional start site (TSS) overlapped an A-DAR element. AML samples displayed 31-53 such transcripts, whereas differentiated cells had 20-28, most of which emanated from LTR12C elements (Figure 2A). We identified 82 spliced transcripts that were present in two or more AML samples but were absent in differentiated cells (Supplementary Table 2). Most of these were short transcripts and only 26 had evidence of splicing into exons of annotated genes. One example involved a LTR2C element active in a subset of AMLs, which acted as a non-reference promoter

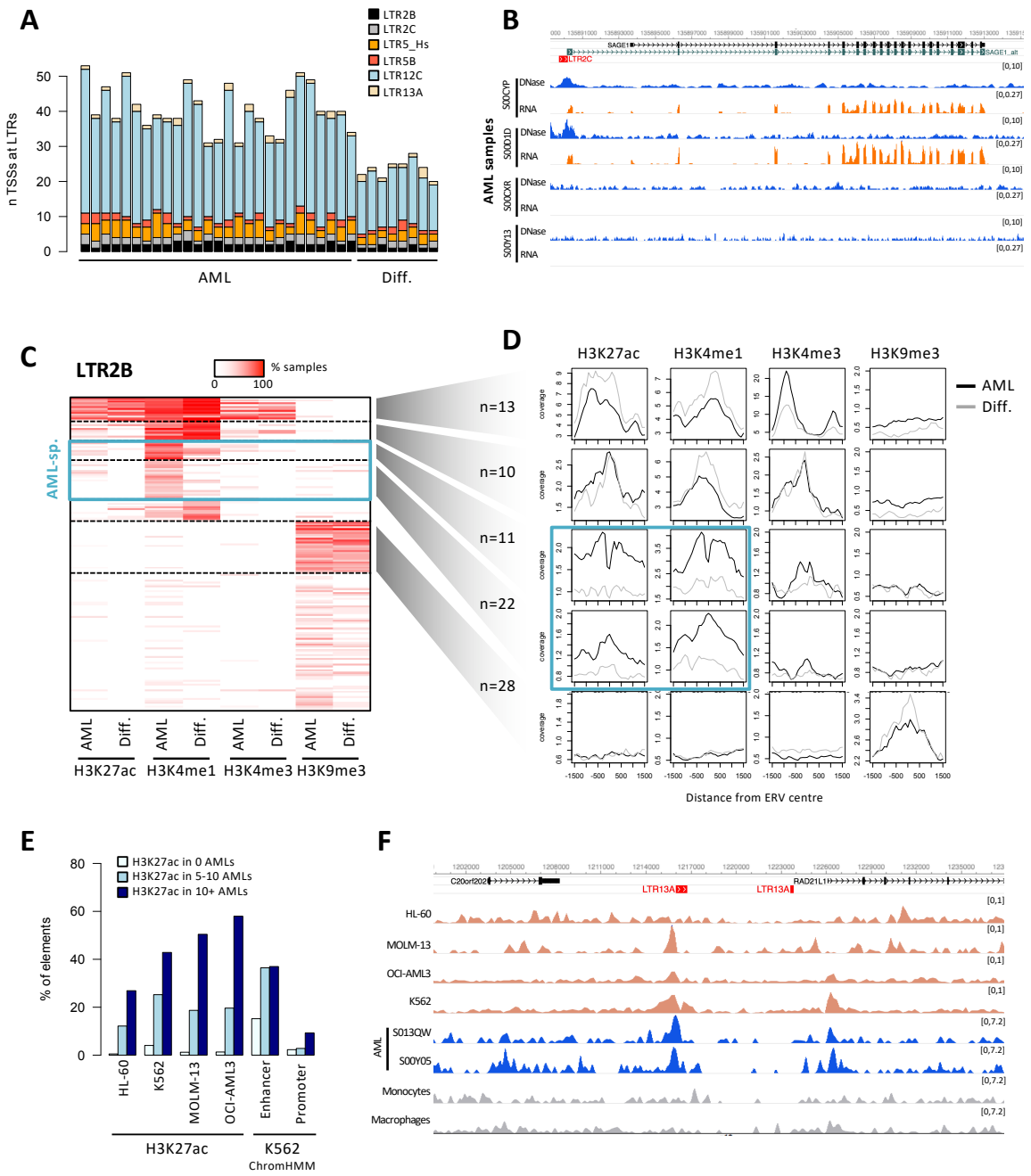

Figure 2. A-DARs bear signatures of enhancer elements. **A.** Number of transcriptional start sites of spliced transcripts that overlap with A-DAR elements in AML or differentiated myeloid cells. **B.** Example of a LTR12C element that generates an alternative promoter that drives the expression of *SAGE1* in AML samples where this element is active. **C.** Heatmap of overlap between LTR2B elements and histone modification peaks. Colour intensity represents the percentage of AML or differentiated cell samples where overlap is observed. Dashed lines segregate clusters identified by k-means clustering. **D.** Average ChIP-seq profiles for LTR2B elements within specific clusters define in C. Blue boxes highlight two clusters where H3K4me1 and H3K27ac levels are higher in AML compared with differentiated cells. **E.** Percentage of A-DAR elements that overlap H3K27ac peaks in different cell lines, or that are classified as enhancer or promoters in ChromHMM data from K562 cells. A-DAR elements were subdivided according to the number of AML samples displaying overlap with H3K27ac. **F.** Example of a LTR13A element where three cell lines reproduce the AML-specific H3K27ac marking observed in AML samples.

for *SAGE1* (Figure 2B), a known cancer/testis antigen (Chen et al., 2014; Piotti et al., 2013).

We then asked whether A-DARs are frequently marked by histone modifications typically associated with promoters or enhancers. Using ChIP-seq data from the Blueprint consortium, we first plotted the percentage of elements from each ERV family that were marked by H3K27ac, H3K4me1, H3K4me3 or H3K9me3 in AML and differentiated myeloid cells (Supplementary Figure 3A). Notably, in AML samples an average 5.7-15.2% of elements from each family overlapped H3K4me1 peaks, a mark predominantly associated with poised and active enhancers. This was substantially higher than the fraction overlapping with the active promoter mark H3K4me3 (1.3-3.4%). Indeed, a more detailed analysis of histone modification patterns at A-DAR elements showed that H3K4me1 is either found in conjunction with H3K27ac (active enhancers), or on its own (primed enhancers) but is more rarely found together with H3K4me3 (Figure 2C, Supplementary Figure 3B). Clustering analysis of these patterns demonstrated that while some elements within a family bear active marks in both AML and differentiated cells, a substantial portion (10-37%, depending on the family; median 20%) display enhancer-like profiles only in AML samples (Figure 2C, Supplementary Figure 3B). Profiles of the ChIP-seq signals confirmed that these AML-specific elements had elevated H3K4me1 and H3K27ac in AML when compared to differentiated cells (Figure 2D). A total of 1,122 and 411 A-DAR elements were marked by H3K4me1 and H3K27ac respectively, in at least two AML samples. A-DARs are therefore frequently associated with enhancer-like profiles in AML.

To test whether myeloid leukemia cell lines could be used to dissect the putative enhancer roles of A-DARs, we performed H3K27ac ChIP-seq on HL-60, MOLM-13, OCI-AML3 and K562 cells, and compared patterns with those seen in AML samples. A-DAR elements that overlap H3K27ac peaks in AML samples were also frequently associated with this mark in cell lines (Figure 2E). A ChromHMM annotation for K562 cells from ENCODE further supported that these elements often bear enhancer signatures (Figure 2E). It is worth noting that there is substantial variation in H3K27ac enrichment of A-DARs among cell lines, much like primary AML samples. Nonetheless, example loci show that H3K27ac deposition at A-DAR elements in cell lines can recapitulate primary AML data (Figure 2F), opening up the opportunity to functionally test for enhancer activity of these loci by genetic and epigenetic editing approaches.

A-DARs bind AML-related TFs

Two large-scale data mining efforts had previously identified several TFs that are associated with ERV families identified here, either through ChIP-seq data or motif analyses (Jacques et al., 2013; Sundaram et al., 2014). These included hematopoiesis- and AML-related TFs such as TAL1, SPI1, GATA2 and ARNT, amongst others. To confirm and extend these observations, we first performed our own analysis of TF ChIP-seq data from K562 cells (ENCODE consortium). Our comparison with AML data above gave us confidence that K562 cells were an adequate model to study TF binding patterns at A-DARs. We analysed all TF ChIP-seq peak data available from ENCODE, comparing their overlap with A-DARs against a random control. We selected TFs that are bound to at least 5% of the elements in a given ERV family, in a statistically significant manner, yielding a list of 217 TFs (Figure 3A; Supplementary Table 3). The vast majority of these TFs were found to be expressed in AML samples (198 had higher expression than TBP). Many of these TFs are involved in hematopoietic gene regulation and/or in the etiology of AML, including SPI1, TAL1, IKZF1 and PKNOX1 (Figure 3A). ChIP-seq profiles of individual elements revealed a localized pattern of TF binding at a subset of elements (Figure 3B). Notably, different ERV families bind different combinations of TFs.

We also performed TF motif analysis, looking for enrichment against shuffled sequence controls (Figure 3C, Supplementary Table 4). Reassuringly, the results from this analysis were largely congruent with the ChIP-seq data. Apart from confirming the presence of motifs for SPI1, PKNOX1 and other TFs, we also found enrichment for HOXA9/MEIS1 motifs in four different ERV families. High *HOXA9* and *MEIS1* expression is associated with specific AML subtypes, namely *NPM1* mutated AMLs (Mullighan et al., 2007), and are sufficient to drive leukemogenesis in mouse models (Kroon et al., 1998). It remains to be tested whether the presence of HOXA9/MEIS1 motifs is what drives *NPM1*-specific patterns of DHSs at A-DARs (Supplementary Figure 2B). In line with the high frequency of many of the identified TF motifs, we found that they were present in the consensus sequences of each ERV family (Figure 3D), suggesting that the respective retroviruses brought in these motifs within their LTRs upon invasion of the human genome. Finally, we asked whether some TF motifs were responsible for chromatin opening at individual elements. We compared motif frequency between elements with DHSs (DHS+) in at least five of the analysed AML samples and DHS-negative elements (Figure 3E). Although some TFs were preferentially associated with DHS+ elements (e.g., CEBPB in LTR2B), many were not, and none were sufficient to explain the presence or absence of DHSs. For example, even though SPI1 binding motif is present in the majority of DHS+ elements,

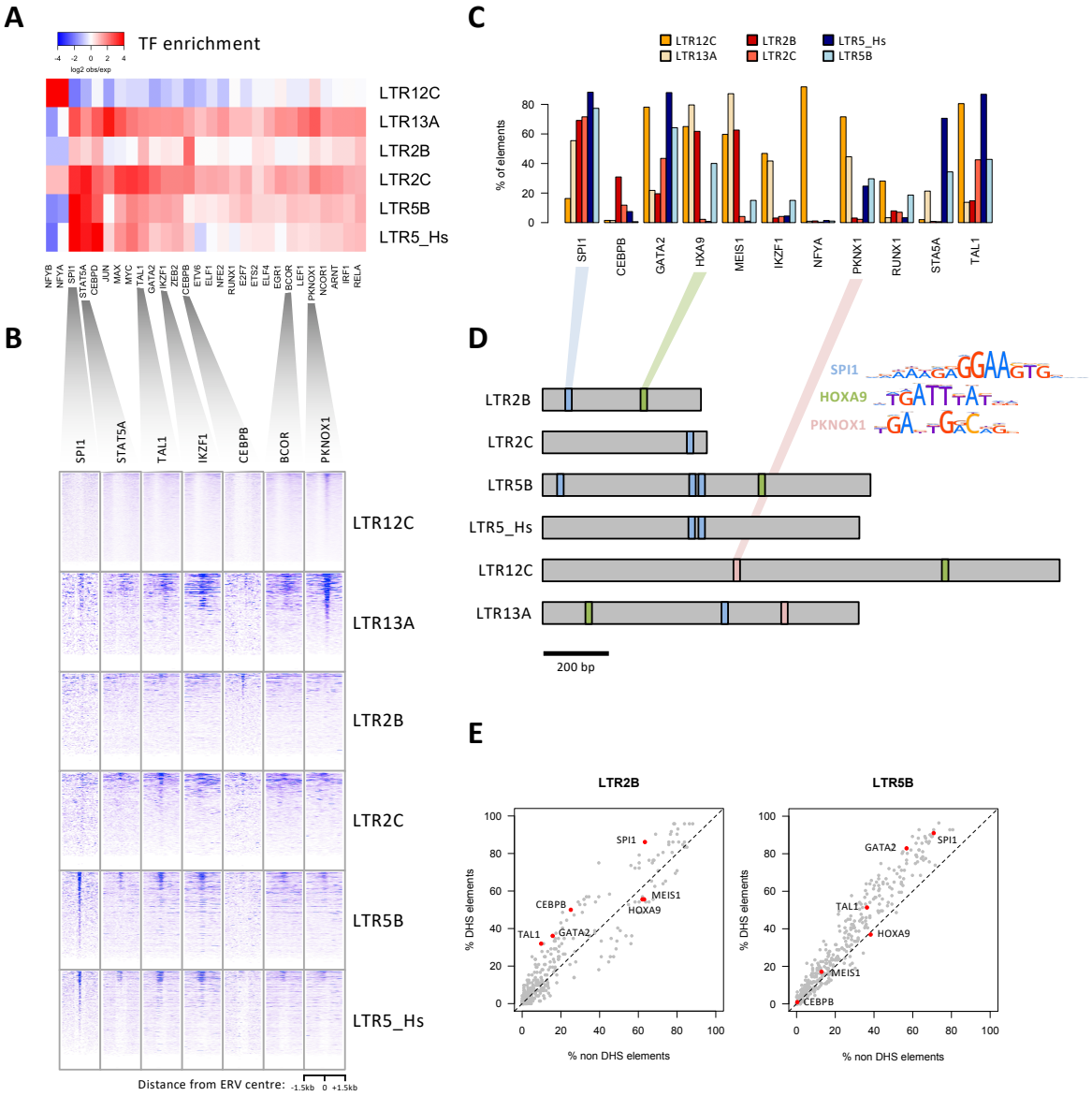

Figure 3. A-DARs bind AML-related transcription factors (TFs). **A.** Heatmap of the observed/expected enrichment for TF binding sites in K562 cells. **B.** ChIP-seq profiles of selected TFs across all elements of each A-DAR family. For each family, elements are displayed in the same order across all TF profiles. **C.** Percentage of ERVs from each family bearing a binding motif for the indicated TFs. **D.** Location of selected TF motifs at the consensus sequences of each A-DAR family. **E.** TF motif frequency at LTR2B and LTR5B elements, comparing those that overlap DHSs with those that do not.

a large portion of non-DHS elements also harbor this motif (Figure 3E). This suggests that other factors play a role in determining LTR regulatory potential, in line with our previous observations in mouse stem cells (Todd et al., 2019).

These analyses suggest that the potential regulatory activity at particular ERV families in AML is likely driven by the binding of hematopoiesis-associated TFs, which are either upregulated in AML or whose binding sites become accessible in AML through epigenetic alterations.

Genetic excision of A-DAR elements interferes with host gene expression

To test for causal roles of enhancer-like A-DAR elements in gene regulation, we used CRISPR-Cas9 to delete three candidate ERVs. The selected ERVs are enriched in H3K27ac, bound by multiple hematopoiesis-associated TFs in K562 cells (Figure 4A), and overlap DHSs in multiple AML samples, but not in monocytes or macrophages (Supplementary Figure 4). We generated clonal lines from K562 cells that had heterozygous or homozygous deletions of these A-DAR elements, and measured the expression of associated genes in multiple clones. Other leukemia cell lines (HL-60, OCI-AML3, MOLM-13) proved more refractory to genetic deletion, due to the low efficiency of Cas9 delivery and single cell expansion.

One of the deleted loci was a LTR5B element located in the first intron of *ZNF321P*, which is bound by PKNOX1, SPI1, STAT5 and TAL1 (Figure 4A, top). Deletion of this element led to a significant decrease in *ZNF321P* expression and also affected the expression of two other nearby genes, *ZNF320* and *ZNF888* (Figure 4B, left). Notably, all three genes display higher expression in AML samples when compared to monocytes and macrophages (Figure 4B, right). Interestingly, *ZNF320* is also upregulated in multiple cancer types (Machnik et al., 2019), but its relevance to cancer etiology remains unclear. ZNF320 is a member of the Krüppel-associated box (KRAB) domain zinc finger family and predominantly binds LTR14A and LTR14B elements (Imbeault et al., 2017), suggesting a potential role in ERV silencing. Heterozygous deletion of another LTR5B element, bound by BCOR, SPI1, TAL1 and RUNX1 (Figure 4A, middle), reduced the expression of Ribosomal Protein L7 Like 1 (*RPL7L1*) (Figure 4C, left), which is upregulated in AML when compared to differentiated myeloid cells (Figure 4C, right). Notably, this specific LTR5B contains a single nucleotide polymorphism (SNP) for which the minor allele (highest population frequency of 0.41) disrupts a MAFK binding motif (Supplementary Figure 4C). Using data from the GTEx project, we found that the minor allele was associated with lower *RPL7L1* expression in whole blood (Supplementary Figure 4C), suggesting that the MAFK motif is important for *RPL7L1* expression. The third deleted locus was an LTR13A element

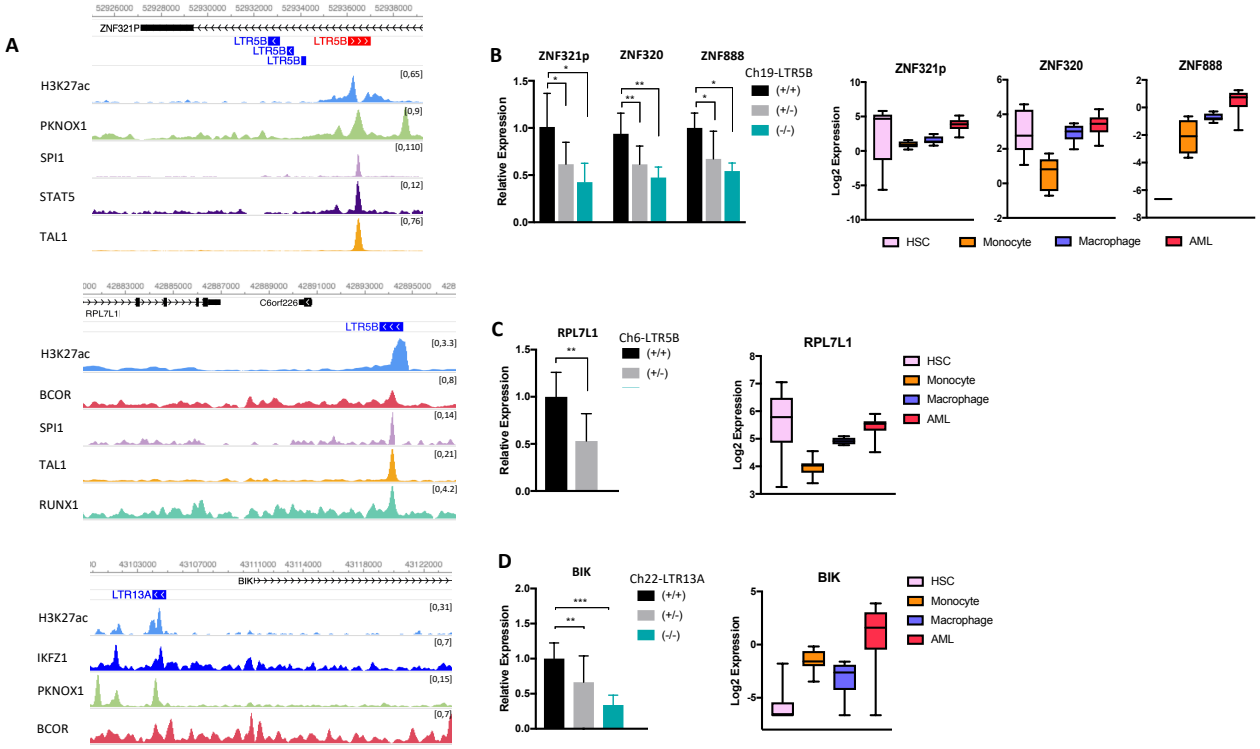

Figure 4. Regulatory ERVs modulate host gene expression. **A.** Genome browser view of three candidate ERVs, showing H3K27ac and TF ChIP-seq tracks in K562 cells. **B,C,D.** Expression of nearby genes (left) in the excision clones of the indicated ERVs (n≥4 for each independent clone: 2 +/+, 3 +/- (B,C,D) and 1 -/- (B,D); ANOVA with Tukey's multiple comparison test, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$) and (right) in HSC (n=6), monocyte (n=8), macrophage (n=6) and AML (n=27) samples (boxes indicate median ± IQR +, whiskers indicate range).

located in the vicinity of *BCL2* interacting killer (*BIK*), and is enriched for IKFZ1, PKNOX1 and BCOR binding (Figure 4A, bottom). Excision of this particular element led to around 3-fold reduction in *BIK* expression (Figure 4D, left), which is higher in AML samples when compared to other hematopoietic cell types (Figure 4D, right). This LTR13A also contains a SNP, where the minor allele (highest population frequency of 0.5) is a critical residue in a RUNX1 binding site, but that was not associated with any significant differences in *BIK* expression in whole blood (Supplementary Figure 4E).
Overall, CRISPR-mediated genetic deletion assays demonstrate a role of a subset of A-DAR elements in gene regulation in K562 cells. Moreover, DHSs within the candidate ERVs and high expression of their associated genes in AML patients provide strong evidence for their regulatory activation *in vivo*.

CRISPRi-mediated inactivation of LTR2B elements leads to growth suppression
To test the regulatory function of multiple A-DAR elements simultaneously, we next sought to epigenetically silence one ERV family by CRISPR interference (CRISPRi) using a catalytically dead Cas9 (dCas9) fused to the KRAB transcriptional repressor protein. We chose the LTR2B family mainly because this was the only one whose DHS enrichment was significant only in AML samples, and not in CD34+ cells (Figure 1B), suggesting a more cancer-specific role than other A-DARs. We designed 4 sgRNAs targeting the most conservative regions of LTR2B family, predicted to recognize around 217 copies (68%) when allowing zero mismatches complementary to the sgRNAs. Our LTR2B sgRNAs are also predicted to target copies of highly related LTR2 copies (71 copies; 8%). We first generated stable CRISPRi OCI-AML3 and K562 cell lines and then stably introduced LTR2B sgRNAs or an empty vector control using a lentiviral system. To determine dCas9 specificity on a genome-wide scale, we performed dCas9 ChIP-seq in K562 cell lines expressing LTR2B sgRNAs or empty vector. We detected 395 dCas9 peaks in cells with LTR2B sgRNAs (and none in control cells), 187 of which were associated with LTR2B elements, and 90 with LTR2 elements (Figure 5A,B). The remaining 118 peaks (Figure 5B) were included in downstream analyses to evaluate putative off-target effects. We also performed H3K27ac and H3K9me3 ChIP-seq in the same cells to assess the epigenetic changes imparted by CRISPRi. We quantified the ratio in histone modification levels at dCas9 peaks between cells expressing LTR2B sgRNAs and those with the empty vector control. As expected, upon CRISPRi in K562 cells we observed a reduction of H3K27ac signal and/or gain of H3K9me3 signal at most loci bound by dCas9, demonstrating effective epigenetic editing (Figure 5C,D). Notably, LTR2B/LTR2 target sites generally underwent more pronounced changes in H3K27ac and H3K9me3

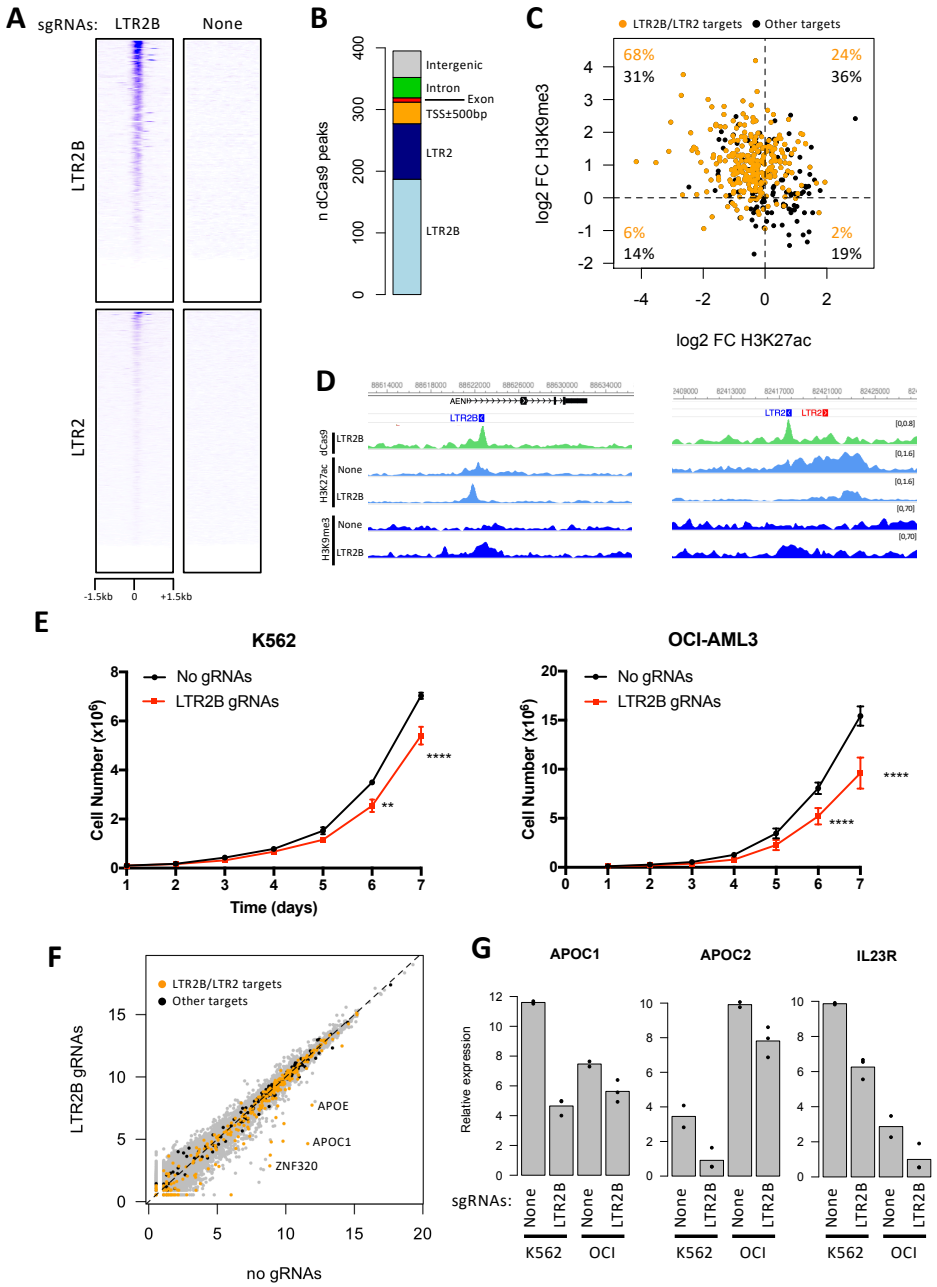

Figure 5. CRISPRi provides robust and selective targeting of LTR2B and LTR2 elements, which led to impaired cell growth. **A.** Profile of dCas9 ChIP-seq signal over LTR2B and LTR2 elements in K562 cells expressing LTR2B sgRNAs or an empty vector ('None'). **B.** Number of dCas9 peaks overlapping LTR2 and LTR2B elements, or other genomic features. **C.** Log2 ratio of the ChIP-seq signal at dCas9 peaks (1kb regions from the centre of each peak) between K562 cells expressing LTR2B sgRNAs or empty vector. Orange points highlight dCas9 peaks overlapping LTR2B or LTR2 elements. **D.** Two examples of LTR2B/LTR2 elements targeted by dCas9, showing decreased H3K27ac and increased H3K9me3. **E.** Cell proliferation assay in K562 (left) and OCI-AML3 (right) cells expressing LTR2B sgRNAs or an empty vector (n=3-5, ANOVA with Sidak's multiple comparison test, $^{**}p < 0.01$, $^{****}p < 0.0001$). **F.** Gene expression levels in K562 cells expressing LTR2B sgRNAs or empty vector. Orange points highlight genes within 50kb of a dCas9 peak targeting LTR2B/LTR2 elements; black points refer to genes within 50kb of other dCas9 peaks. **G.** Comparison of expression changes at selected genes between K562 and OCI-AML3 ('OCI') CRISPRi cells.

levels when compared to off-target sites. Changes in histone modification levels upon CRISPRi were further confirmed by ChIP-qPCR at LTR2B elements (Supplementary Figure 5A). In OCI-AML3 cells we observed a similar trend in epigenetic alterations upon CRISPRi, albeit to a lesser extent than in K562 cells (Supplementary Figure 5B,C).

Intriguingly, proliferation assays showed that epigenetic silencing of LTR2B and LTR2 elements by CRISPRi significantly suppressed cell proliferation in both K562 and OCI-AML3 cell lines (Figure 5E). To test the impact of LTR2B and LTR2 inactivation on the host transcriptome, and gain insights into the mechanism underlying impaired cell growth, we performed RNA-seq in both cell lines (Figure 5F; Supplementary Figure 5D). We identified a total of 58 and 99 differentially expressed genes in K562 and OCI-AML3 cells, respectively (Supplementary Table 5). To elucidate direct effects of CRISPRi, we focused on genes that are within 50 kb of a dCas9 peak and found 15 and 6 differentially expressed genes (in K562 and OCI-AML3 cells, respectively), all but one of which were downregulated. Only one of these genes (*BIK*), which was downregulated in OCI-AML3, was associated with an off-target dCas9 peak. The remaining genes were associated with 15 different LTR2B/LTR2 elements. In some instances, the LTR2B/LTR2 element was very close to the promoter of the affected gene, such that silencing could have resulted from H3K9me3 spreading. We therefore performed genetic deletion of one of these elements, which also led to a decrease in gene expression of the adjacent *ZNF611* gene, albeit to a lesser extent than by CRISPRi (Supplementary Figure 5E). Several genes displayed decreased expression in both cell lines (Figure 5G), although only apolipoprotein C1 (*APOC1*) reached statistical significance in both contexts. Notably, five apolipoprotein genes were downregulated in at least one of the cell lines. *APOC1*, *APOC2*, *APOC4-APOC2* and *APOE* lie within a cluster on chromosome 19, and may all be controlled by the same LTR2 element, located upstream of *APOC1*. On the other hand, *APOL1* is on chromosome 22 and close to an LTR2B insertion. Given the key roles that lipid metabolism plays in supporting cancer cell survival (Beloribi-Djefaflia et al., 2016), the coordinated downregulation of apolipoprotein genes could underpin the reduced cell growth observed upon silencing of LTR2B/LTR2 elements in leukemia cell lines.

Overall, these data show that a subset of LTR2B and LTR2 elements act as key gene regulators in leukemia cell lines, and that their epigenetic silencing impairs cell growth, providing evidence for a putative functional role in AML.

***APOC1*-associated LTR2 is required for proliferation of myeloid leukemia cells**

APOC1 has recently been shown to maintain cell survival in AML and the knockdown of *APOC1* impairs cell growth (Yang et al., 2018). Similar findings were made in pancreatic and colorectal cancer, where *APOC1* overexpression is associated with poor prognosis (Ren et al., 2019; Takano et al., 2008). We therefore asked whether ERV-mediated regulation of *APOC1* could affect cell growth. There is an LTR2 insertion upstream of the *APOC1* promoter (*APOC1*-LTR2; Figure 6A, Supplementary Figure 6A), which has been previously described to act as an alternative promoter in several tissues, but only accounts for up to 15% of total *APOC1* transcription (Medstrand et al., 2001). In K562 and OCI-AML3 RNA-seq data, we find no evidence of *APOC1*-LTR2 promoter activity (Figure 6A), which we have confirmed by RT-qPCR (Supplementary Figure 6B), suggesting that *APOC1*-LTR2 acts as an enhancer element. *APOC1*-LTR2 is enriched in STAT5 and TAL1 binding and shows an increase in H3K9me3 and decrease in H3K27ac upon CRISPRi in both K562 and OCI-AML3 (Figure 6A, Supplementary Figure 6A). To test for a direct role of *APOC1*-LTR2 in *APOC1* gene expression and cell growth, we deleted this element in K562 cells. We obtained 7 heterozygous and 8 homozygous clones from a total of 110 clones. Interestingly, none of the homozygous clones were able to grow more than 10 days in culture, suggesting that homozygous deletion may impair cell growth. To pursue the impact of *APOC1*-LTR2 on cell growth, we used lentiviral-mediated CRISPR-Cas9 delivery and performed assays in a pool of edited cells (Figure 6B). At day 6, after GFP and puromycin selection of the two flanking sgRNAs, we observed around 60% deletion of *APOC1*-LTR2 and more than 2.5-fold reduction in *APOC1* gene expression compared to an empty vector control (Figure 6C,D). Remarkably, deletion of this element was sufficient to drive a significant suppression of cell proliferation compared to control cells (Figure 6E). As there is a small fraction of unedited cells in the pool, we asked whether the unedited cells may outcompete edited cells over time. After day 20, the deletion was reduced to around 35% and only 1.2-fold difference was observed in *APOC1* expression, and consequently there was no difference in cell proliferation after day 20, indicating that *APOC1*-LTR2 provides cells with a selective growth advantage (Figure 6C,D; Supplementary Figure 6C). To further investigate how *APOC1*-LTR2 deletion leads to impaired cell growth, we analyzed cell cycle and apoptosis with flow cytometry in K562 cells at day 6. While no differences in G1, S, and G2 phases were detected, there was a significant increase in the sub-G1 population (Figure 6F) in edited cells. In agreement with this, Annexin V signal was significantly higher in edited cells compared to unedited cells at day 6 (Figure 6G), showing that the deletion of *APOC1*-LTR2 induces apoptosis, which is in line with known effects of *APOC1*

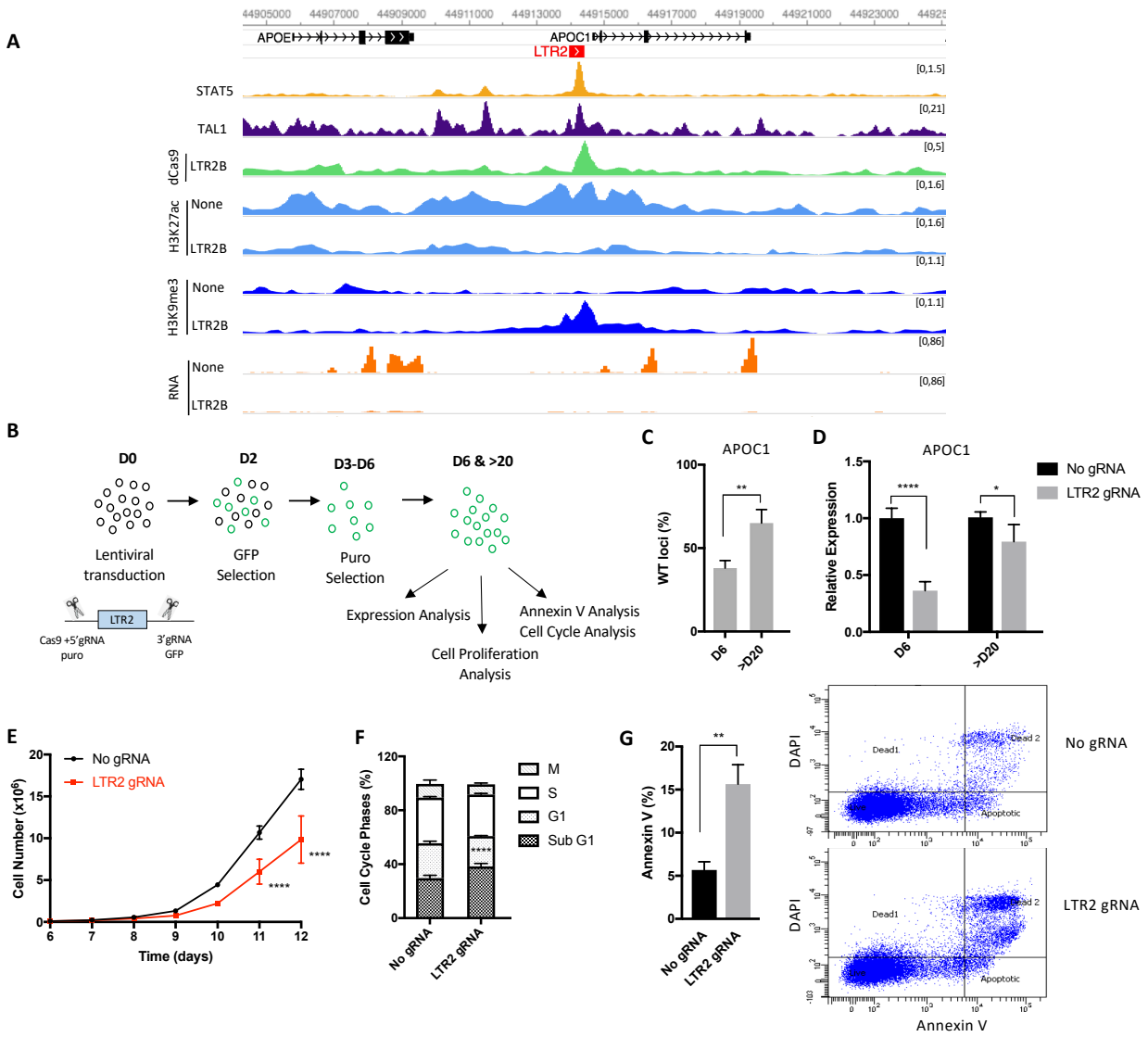

Figure 6. *APOC1*-LTR2 element regulates *APOC1* expression and is essential for cell proliferation. **A.** Genome browser snapshot for *APOC1*-LTR2 element, showing TAL1, STAT5 ChIP-seq tracks for WT; H3K27ac, H3K9me3 ChIP-seq and RNA-seq tracks for no control and CRISPRi K562 cells. **B.** Schematic of the experimental design to genetically excise *APOC1*-LTR2 element. **C.** qPCR and **D.** RT-qPCR data from cells with *APOC1*-LTR2 excision. The error bars show standard deviation (n=3-4, t-test (C) ANOVA with Tukey's multiple comparison test (D), $*p < 0.05$, $**p < 0.01$, $****p < 0.0001$). **E.** Cell proliferation assay of control and *APOC1*-LTR2 excised cells after puromycin selection (day 6) (n=4, ANOVA with Tukey's multiple comparison test, $****p < 0.0001$). **F.** Cell cycle profiles of control and *APOC1*-LTR2 excised cells (n=3, ANOVA with Tukey's multiple comparison test, $****p < 0.0001$). **G.** % of Annexin V stained cells in K562 cells upon *APOC1*-LTR2 excision (left, n=3, t-test, $**p < 0.01$). Representative flow cytometry analysis of Annexin V (right).

depletion (Ren et al., 2019; Takano et al., 2008; Yang et al., 2018). As expected, this difference is much smaller after day 20 (Supplementary Figure 6D). We also tested the effect of *APOC1*-LTR2 deletion in OCI-AML3 cells, but due to the low efficiency of Cas9 delivery and low viability of cells at day 6, we performed expression and annexin V analysis at day 10. Similar to what we observed in K562 cells, *APOC1*-LTR2 deletion in OCI-AML3 cells led to around 4-fold decrease in *APOC1* expression and increased annexin V signal, and these effects were milder at day 23 (Supplementary Figure 6E,F).

Together, our findings indicate that the *APOC1*-LTR2 element is essential for proliferation of leukemia cells by acting as an enhancer of the *APOC1* gene, which in turn controls cell survival via an anti-apoptotic mechanism.

Discussion

Here we provide multiple lines of evidence that particular ERVs are used as regulatory elements to activate gene expression in AML, which may be exploited by cancer cells to help drive disease phenotypes and cancer progression. It had been previously postulated that the epigenetically relaxed state of cancer cells provides a window of opportunity for ERV activation, triggering their intrinsic regulatory capacity (Babaian and Mager, 2016; Chuong et al., 2017; Lamprecht et al., 2010). However, to the best of our knowledge, all examples to date supporting this hypothesis have involved activation of cryptic promoters to drive expression of adjacent genes (Babaian and Mager, 2016; Jang et al., 2019). Whilst we uncovered some examples of chimeric transcripts starting from ERVs in AML (e.g., LTR2C-*SAGE1*), which are not present in differentiated myeloid cells, our analyses suggest that active A-DARs mainly harbour chromatin signatures of enhancers.

In line with previous studies, only a subset of all the elements in a given ERV family (10-37%) are marked by active histone modifications and bound by haematopoiesis-related TFs, despite a high level of motif conservation across ERV copies. Nonetheless, we identified multiple ERV elements with strong evidence supporting their role as bona fide gene regulators: 1) we found striking correlations between differential chromatin accessibility at 20 ERVs and the expression of nearby genes, some of which have been linked to AML prognosis (Figure 1E,F); 2) CRISPR-mediated genetic editing experiments revealed an additional five ERVs that act as enhancers in leukemia cells (Figure 4, Supplementary Figure 5E, Figure 6); 3) CRISPRi identified another 13 different elements whose epigenetic silencing led to the down-regulation of nearby genes (Supplementary Table 6). A more exhaustive search would likely have revealed additional regulatory elements, namely via epigenetic silencing of other ERV

families. We were also purposefully conservative in our assessment of ERV effects on gene expression, which may frequently extend beyond 50 kb, although additional evidence would then be necessary to support a direct role.

Despite the growing evidence that ERVs can act as regulatory elements in different cancers, there are limited examples for their inappropriate activation contributing to oncogenesis, a term coined as onco-exaptation (Babaian et al., 2016). The term has been frequently used to describe the gain of regulatory activity at TEs in cancer. Our view is that, similar to the term exaptation (Gould and Vrba, 1982), onco-exaptation requires that this new regulatory activity provides the cancer cell with a selective advantage. Strong demonstrations of such adaptive roles are scarce. Notably, the Wang lab recently showed that an AluJb element acts as an oncogenic promoter to drive *LIN28B* expression and tumour progression in lung cancer (Jang et al., 2019). In our study we identified an LTR2 element, the genetic and epigenetic perturbation of which suppressed cell growth and induced apoptosis of leukemia cell lines by altering lipid-related *APOC1* expression. Despite the striking cellular phenotype in cell lines, comprehensive analyses of primary AML samples are warranted to demonstrate whether these regulatory ERVs are sufficient to provide survival advantages for the cancer cells *in vivo* and contribute to prognosis of AML. Notably, DNase-seq peaks associated with *APOC1*-LTR2 in AML samples are subtler than those observed in cell lines, yet a few AML samples express relatively high levels of *APOC1* (Supplementary Figure 7A,B). Interestingly, overall survival curves based on TCGA data suggest that a small proportion of patients with high *APOC1* expression have a poorer prognosis, a pattern that is also seen in stomach adenocarcinoma (Supplementary Figure 7C,D). A considerably larger number of patients would be necessary to confirm this finding, although independent datasets have led to similar observations in colorectal and pancreatic cancer (Ren et al., 2019; Takano et al., 2008). APOC1 is also activated in monocyte-to-macrophage differentiation (Lauer et al., 1988), raising the possibility that *APOC1*-LTR2 may play other roles in haematopoiesis outside of AML.

Given their repetitive nature, one intriguing question is why particular ERVs within a family are recurrently activated in AML to drive nearby gene expression, yet the majority of them are functionally neutral. One explanation lies in the nature of inter- and intra-cellular epigenetic heterogeneity that increases during malignancy formation. This gives rise to epigenetic activation of a set of ERVs, as proposed in the epigenetic evolution model (Babaian and Mager, 2016). Accordingly, cells harbouring activated ERVs that drive oncogenes gain a selective advantage and increase in frequency during cancer evolution. Therefore, clonal expansion of these cells will enable the detection of oncogenic ERVs in a cell population. However, whether ERV activation

contributes to cancer evolution or is simply a consequence of the molecular state of cancer remains a matter of debate.

Irrespective of whether epigenetic heterogeneity at ERVs contributes to tumour evolution, distinct patterns of ERV activity are observed across different patients (Supplementary Figure 1A). These differences appear to be partly driven by the underlying mutational profiles. We also identified a SNP within an ERV that seemingly affects its regulatory activity by altering a TF binding site (Supplementary Figure 4C), suggesting that genetic variation within ERVs also contributes to inter-individual differences in ERV activity. Finally, younger ERVs such as LTR5_Hs are structurally polymorphic within the human population (Belshaw et al., 2005; Thomas et al., 2018), adding another layer of genetic variation. Regulatory ERVs may therefore foster genetic, epigenetic and transcriptional heterogeneity of the disease with potential to contribute to clinical outcomes. One significant consequence of the molecular heterogeneity of AML is the escape of resistant clones from treatment, resulting in high relapse rates. It will be therefore interesting to discover to which extent the ERV-derived heterogeneity contributes to inter-individual differences in response to AML therapies.

Altogether, our work reveals the first example of ERVs as oncogenic enhancers in AML and, more importantly, in any human malignancy. These data highlight the significance of expanding the search for oncogene drivers to the repetitive part of the genome, which may pave the way for the development of novel prognostic and therapeutic approaches.

Methods

Cell culture and cell proliferation assays

293T cells and human leukemia cell lines K562, OCI-AML3, MOLM-13 and HL-60 were routinely cultured in RPMI 1640 (and DMEM (HEK293T) supplemented with 10% fetal bovine serum, 2mM glutamax and 1% penicillin/streptomycin at 37°C in 5% carbon dioxide. Cells were maintained and split every 2-3 days.

For cell proliferation assays, exponentially growing cells were plated in 24-well plates ($1x10^5$ cells/ml). Every 2-3 days media were replaced and cells were split to $1x10^5$ cells/ml. The viable cells were counted daily for 6 days.

Cell-cycle and apoptosis assays

Cell cycle assay was performed using muse cell cycle kit by following manufacturers instructions (Millipore) and the cells were analysed by BD FACS Canto II. For

apoptosis assay, the cells were stained by an annexin V 647 (Thermofisher Scientific) and DAPI and analysed by BD FACS Canto II.

CRISPR-Cas9-mediated LTR disruption

For CRISPR/Cas9 deletion of LTRs, sgRNA oligonucleotides (Sigma-Aldrich) targeting upstream and downstream of LTRs of interest were annealed and cloned into modified eSpCas9 (1.1) vector (Addgene 71814, deposited by Feng Zhang), which expresses GFP. K562 cells were nucleofected with eSpCas9 plasmid containing gRNAs using amaxa nucleofector kit V. 2 days later, cells expressing GFP were sorted on a FACS Aria II and single cells were plated onto a 96-well plate. After 2 weeks, cells were genotyped by PCR and the gene expression of LTR-knockout cells was analyzed by RT-qPCR.

For LTR2-APOC1 deletion, 5' sgRNAs (Sigma-Aldrich) were cloned into lentiCRISPR v2 (Addgene 52961) and 3' sgRNAs were cloned into lenti_sgRNA_EFS_GFP (Addgene 65656) vector. OCI-AML3 and K562 cells were transduced with the lentiviral vectors containing sgRNAs and selected for GFP and puro. % of WT loci was determined by qPCR using APOC_R and APOC_I genotyping primers listed in Supplementary Table 7. The cells were cultured around three weeks for RNA expression and phenotypical analysis.

CRISPRi-mediated silencing of LTRs

sgRNAs (Sigma-Aldrich) targeting multiple LTR copies were cloned into lentiviral expression vector pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene 50946, deposited by K. Yusa). For LTR silencing, OCI-AML3 and K562 cells were first transduced with the lentiviral vector pHR-SFFV-KRAB-dCas9-P2A-mCherry (Addgene 60954, deposited by Jonathan Weissman), sorted for mCherry on a FACSAria II. Cells expressing mCherry were then subsequently transduced with the lentiviral sgRNA expression vector. 2 days later, the cells expressing both mCherry and BFP were sorted and cultured for transcriptional and chromatin analyses.

Lentiviral production and transduction

Lentivirus was produced in HEK 293 T cells by triple transfection with delivery vector and the packaging plasmids psPAX2 and pMD.G. The viral supernatants were collected 48 h after transfection and filtered through a 0.45 μM filter. Target cells were transduced with lentiviral supernatant supplemented with 4 μg/mL polybrene.

RNA Isolation and RT-qPCR

RNA was extracted using AllPrep DNA/RNA mini kit (Qiagen 80204) and DNAse treated with the TURBO DNA-free™ Kit (Ambion, AM1907). RNA (1 μg) was retrotranscribed using Revertaid Reverse Transcriptase (Thermo Scientific EP0441)

and the cDNA was diluted 1/10 for qPCRs using MESA BLUE MasterMix (Eurogenentec, 10-SY2X-03+NRWOUB) on a LightCycler® 480 Instrument II (Roche). A list of primers used can be found in Supplementary Table 7.

RNA-seq Library Preparation

Ribosomal RNA-depleted RNA-seq libraries were prepared from 200-500 ng of total RNA using the low input ScriptSeq Complete Gold Kit (Epicentre). Libraries were sequenced on an Illumina NextSeq 500 with single-end 75 bp reads.

Chromatin Immunoprecipitation

Approximately 10^7 cells were fixed with 1% formaldehyde for 12 minutes in PBS and quenched with glycine. Chromatin was sonicated using a Bioruptor Pico (Diagenode), to an average size of 200-700 bp. Immunoprecipitation was performed using 75 µg of chromatin and 5 µg of Cas9 antibody (Diagenode #C15200229-100) or 15 µg of chromatin and 2.5 µg of H3K27ac and H3K9me3 antibody (Active Motif #3913, Diagenode #C15410193). Final DNA purification was performed using the GeneJET PCR Purification Kit (Thermo Scientific. #K0701) and DNA was eluted in 80 µL of elution buffer. This was diluted 1/10 and analysed by qPCR, using the KAPA SYBR® FAST Roche LightCycler® 480 2X qPCR Master Mix (Kapa Biosistems, Cat. KK4611). A list of primers used can be found in Supplementary Table 7.

Library preparation and sequencing for ChIP-seq and DNase-seq

ChIP-seq and DNase-seq libraries were prepared from 1-5 ng ChIP DNA or DNase DNA samples using NEBNext Ultra II DNA library Prep Kit (Illumina). Libraries were sequenced on an Illumina NextSeq 500 with single-end or paired-end 75 bp reads.

Chromatin accessibility assay

To asses chromatin accessibility, DNase I digestion was performed as previously described (Song and Crawford, 2010). 5 million cells were resuspended in RSB buffer (10 mM NaCl, 3mM $MgCl_2$, 10mM Tris-Cl, pH 7.4). After cell lysis, the nuclei were digested with DNase I with 0, 0.1, 2, 5, 15 and 30 U for 10 min at 37ºC. Digests were inactivated by the addition of 50 mM EDTA. RNA and proteins were digestion by RNase A (0.5 mg/ml) for 15 min at at 37°C and then by proteinase K (0.5 mg/ml) for 1 h at 65 ºC. DNA was purified by phenol-chloroform extraction and ethanol precipitation. The resuspended DNA was analysed by qPCR, using the KAPA SYBR® FAST Roche LightCycler® 480 2X qPCR Master Mix (Kapa Biosistems, Cat. KK4611) and chromatin digested with 15 U was selected for library preparation and sequencing.

Primary processing of high-throughput sequencing data

Reads from high-throughput sequencing data generated here or from external datasets (Supplementary Table 8) were trimmed using first trimmed using Trim Galore. ChIP-seq and DNase-seq data were aligned to the hg38 genome assembly using

Bowtie2 v2.1.0 (Langmead and Salzberg, 2012), followed by filtering of uniquely mapped reads with a custom script. ChIP-seq peak detection was performed using MACS2 v2.1.1 (Y. Zhang et al., 2008) with -q 0.05; for histone marks the option --broad was used. DNase-seq peak detection was performed using F-seq v1.84 (Boyle et al., 2008) with options -f 0 -t 6. RNA-seq data were mapped using Hisat2 v2.0.5 (Kim et al., 2019) with option --no-softclip. Raw read counts for each gene were generated in Seqmonk with the RNA-seq quantitation pipeline, and normalised gene expression values calculated with the variance stabilizing transformation in DESeq2 (Love et al., 2014). BigWig tracks were generated using the bamCoverage function of deepTools2.0, with CPM normalisation and 200 bp bin size. Other processed data from Blueprint, ENCODE and other sources (Supplementary Table 8) were downloaded as peak annotations or expression values (e.g., FPKM).

DHS enrichment at repeat families

DHSs (i.e., DNase-seq peaks) were intersected with the Repeatmasker annotation and the number of overlapped DHSs per repeat family calculated. For comparison, 1000 random controls were generated by shuffling the DHSs in a given sample, avoiding unmappable regions of the genome. P values were calculated based on the number of random controls for which the number of DHS overlaps displayed more extreme values (at either tail of the distribution) than those seen with the real DHSs. Enrichment values were calculated by dividing the number of real DHS overlaps with the mean number of DHS overlaps in the random controls. Significantly enriched repeat families had: 1) $p<0.05$, 2) >2-fold enrichment, 3) >20 copies overlapped by DHSs. Selected families were significantly enriched for DHSs in at least one of the cell lines analysed (HL-60, OCI-AML3, MOLM-13) and in >10% of AML samples.

Mutational profile analysis

A-DAR elements overlapping DHSs in at least one sample were selected, and a correlation matrix built based on the patterns of DHS overlap between samples. These were compared with the AML mutational profiles extracted from the respective publications (Assi et al., 2018; Yi et al., 2019). Correlation coefficients between AML samples sharing a particular mutation were compared with correlation coefficients between samples without the mutation.

Identification of active A-DAR promoters

Aligned BAM files from Blueprint RNA-seq data were processed using StringTie v1.3.3b (Pertea et al., 2015) with options --rf -G to generate sample-specific guided transcriptome assemblies. Spliced transcripts initiating at A-DAR elements were then identified by intersecting the TSSs of multi-exon transcripts A-DAR annotations. A-

DAR elements with TSSs in AML samples but not in differentiated cells were selected and the associated transcripts visually inspected.

K562 TF ChIP-seq analysis

ENCODE TF ChIP-seq peak files from K562 (Supplementary Table 8) were downloaded and intersected with A-DAR annotations, as well as with a randomly shuffled version of these elements. TFs significantly enriched (corrected $p<0.05$) in at least one of the A-DAR families, covering at least 5% of the elements in that family were selected. For each TF, average enrichment values were calculated across technical and biological replicates, as well as independent ChIP-seq experiments of the same TF.

TF motif analysis

Motif analysis of A-DARs was performed using the AME and FIMO tools of the MEME SUITE v5.0.1 (Bailey et al., 2015) using the HOCOMOCO v11 human TF motif database. Motifs enriched in at least one A-DAR family were identified using AME, and motif frequency and location extracted using FIMO. Consensus sequences were downloaded from Dfam (Hubley et al., 2016).

CRISPRi ChIP-seq and RNA-seq analyses

Normalised H3K27ac and H3K9me3 ChIP-seq read counts were extracted around dCas9 peaks (±500 bp from the peak centre). Genes within 50 kb of a dCas9 peak were considered as putative direct targets of CRISPRi. Differential gene expression analysis was performed using DEseq2 (Love et al., 2014).

Data availability

Datasets are available through the Gene Expression Omnibus (GEO) under accession number GSE136764.

Acknowledgements

We thank Yasmine Benbrahim for ideas informing bioinformatic analyses, the Dawson lab for their guidance in lentiviral transduction of AML cell lines, Brian Huntly for providing OCI-AML3, MOLM-13 and HL-60 cell lines, and Diego Villar for critical reading of the manuscript. This work was supported by funding from Barts Charity (Small Project Grants – MGU0462). O.D. received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n° 608765. M.R.B. was supported by a Sir Henry Dale Fellowship (101225/Z/13/Z), jointly funded by the Wellcome Trust and the Royal Society. This study makes use of data generated by the Blueprint Consortium. A full list of the investigators who contributed to the generation

of the data is available from www.blueprint-epigenome.eu. Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 – BLUEPRINT. This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (King et al. 2017).

Author Contributions

O.D. and M.R.B. designed the study and experiments and wrote the manuscript. O.D. performed cell culture, DNase-seq, ChIP-seq, RNA-seq, CRISPR, CRISPRi and cellular phenotyping. M.A. generated the ZNF611-LTR2B KO. C.D.T. assisted in the design and execution of CRISPR experiments. A.R.M. performed overall survival analyses. M.A.D. assisted in the establishment of CRISPRi cell lines. M.R.B. performed bioinformatic analyses.

Competing Interests

The authors declare no competing interests.

Tables

Supplementary Table 1: The results of DHS permutation test for each transposon in AML cell lines, samples and other hematopoietic cell types

Supplementary Table 2: List of AML-associated spliced transcripts emanated from A-DARs

Supplementary Table 3: K562-enriched TFs that are bound to A-DARs

Supplementary Table 4: TF motif frequency within A-DARs with and without DHSs

Supplementary Table 5: Differentially expressed genes in K562 and OCI-AML3 upon CRISPRi

Supplementary Table 6: List of ERV elements that display regulatory activity in myeloid leukemia

Supplementary Table 7: Primers used in this study

Supplementary Table 8: External datasets used in this study

References

Assi, S.A., Imperato, M.R., Coleman, D.J.L., Pickin, A., Potluri, S., Ptasinska, A., Chin, P.S., Blair, H., Cauchy, P., James, S.R., Zacarias-Cabeza, J., Gilding, L.N., Beggs, A., Clokie, S., Loke, J.C., Jenkin, P., Uddin, A., Delwel, R., Richards, S.J., Raghavan, M., Griffiths, M.J., Heidenreich, O., Cockerill, P.N., Bonifer, C., 2018. Subtype-specific regulatory network rewiring in acute myeloid leukemia. Nat Genet 368, 2059. doi:10.1038/s41588-018-0270-1

Babaian, A., Mager, D.L., 2016. Endogenous retroviral promoter exaptation in human cancer. Mob DNA 7, 24. doi:10.1186/s13100-016-0080-x

Babaian, A., Romanish, M.T., Gagnier, L., Kuo, L.Y., Karimi, M.M., Steidl, C., Mager, D.L., 2016. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. Oncogene 35, 2542–2546. doi:10.1038/onc.2015.308

Bailey, T.L., Johnson, J., Grant, C.E., Noble, W.S., 2015. The MEME Suite. Nucleic Acids Res. 43, W39–49. doi:10.1093/nar/gkv416

Beloribi-Djefaflia, S., Vasseur, S., Guillaumond, F., 2016. Lipid metabolic reprogramming in cancer cells. Oncogenesis 5, e189–e189. doi:10.1038/oncsis.2015.49

Belshaw, R., Dawson, A.L.A., Woolven-Allen, J., Redding, J., Burt, A., Tristem, M., 2005. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol 79, 12507–12514. doi:10.1128/JVI.79.19.12507-12514.2005

Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., Tristem, M., 2007. Rate of recombinational deletion among human endogenous retroviruses. J Virol 81, 9437–9442. doi:10.1128/JVI.02216-06

Boyle, A.P., Guinney, J., Crawford, G.E., Furey, T.S., 2008. F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics 24, 2537–2538. doi:10.1093/bioinformatics/btn480

Brocks, D., Schmidt, C.R., Daskalakis, M., Jang, H.S., Shah, N.M., Li, D., Li, J., Zhang, B., Hou, Y., Laudato, S., Lipka, D.B., Schott, J., Bierhoff, H., Assenov, Y., Helf, M., Ressnerova, A., Islam, M.S., Lindroth, A.M., Haas, S., Essers, M., Imbusch, C.D., Brors, B., Oehme, I., Witt, O., Lübbert, M., Mallm, J.-P., Rippe, K., Will, R., Weichenhan, D., Stoecklin, G., Gerhäuser, C., Oakes, C.C., Wang, T., Plass, C., 2017. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet 49, 1052–1060. doi:10.1038/ng.3889

Burns, K.H., 2017. Transposable elements in cancer. Nat Rev Genet 17, 415–424. doi:10.1038/nrc.2017.35

Cancer Genome Atlas Research Network, Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A.G., Hoadley, K., Triche, T.J., Laird, P.W., Baty, J.D., Fulton, L.L., Fulton, R., Heath, S.E., Kalicki-Veizer, J., Kandoth, C., Klco, J.M., Koboldt, D.C., Kanchi, K.-L., Kulkarni, S., Lamprecht, T.L., Larson, D.E., Lin, L., Lu, C., McLellan, M.D., McMichael, J.F., Payton, J., Schmidt, H., Spencer, D.H., Tomasson, M.H., Wallis, J.W., Wartman, L.D., Watson, M.A., Welch, J., Wendl, M.C., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y., Chiu, R., Chu, A., Chuah, E., Chun, H.-J., Corbett, R., Dhalla, N., Guin, R., He, A., Hirst, C., Hirst, M., Holt, R.A., Jones, S., Karsan, A., Lee, D., Li, H.I., Marra, M.A., Mayo, M., Moore, R.A., Mungall, K., Parker, J., Pleasance, E., Plettner, P., Schein, J., Stoll, D., Swanson, L., Tam, A., Thiessen, N., Varhol, R., Wye, N., Zhao, Y., Gabriel, S., Getz, G., Sougnez, C., Zou, L., Leiserson, M.D.M., Vandin, F., Wu, H.-T., Applebaum, F., Baylin, S.B., Akbani, R., Broom, B.M., Chen, K., Motter, T.C., Nguyen, K., Weinstein, J.N., Zhang, N., Ferguson, M.L., Adams, C., Black, A., Bowen, J., Gastier-Foster, J., Grossman, T., Lichtenberg, T., Wise, L., Davidsen, T., Demchok, J.A., Shaw, K.R.M., Sheth, M., Sofia, H.J., Yang, L., Downing, J.R., Eley, G., 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N. Engl. J. Med. 368, 2059–2074. doi:10.1056/NEJMoa1301689

Chen, Y.-T., Panarelli, N.C., Piotti, K.C., Yantiss, R.K., 2014. Cancer-testis antigen expression in digestive tract carcinomas: frequent expression in esophageal squamous cell carcinoma and its precursor lesions. Cancer Immunol Res 2, 480–486. doi:10.1158/2326-6066.CIR-13-0124

Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18, 71–86. doi:10.1038/nrg.2016.139

Chuong, E.B., Elde, N.C., Feschotte, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 351, 1083–1087. doi:10.1126/science.aad5497

Chuong, E.B., Rumi, M.A.K., Soares, M.J., Baker, J.C., 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet 45, 325–329. doi:10.1038/ng.2553

Edginton-White, B., Cauchy, P., Assi, S.A., Hartmann, S., Riggs, A.G., Mathas, S., Cockerill, P.N., Bonifer, C., 2019. Global long terminal repeat activation participates in establishing the unique gene expression programme of classical

Hodgkin lymphoma. Leukemia 33, 1463–1474. doi:10.1038/s41375-018-0311-x
Ferreira, L.M.R., Meissner, T.B., Mikkelsen, T.S., Mallard, W., O'Donnell, C.W., Tilburgs, T., Gomes, H.A.B., Camahort, R., Sherwood, R.I., Gifford, D.K., Rinn, J.L., Cowan, C.A., Strominger, J.L., 2016. A distant trophoblast-specific enhancer controls HLA-G expression at the maternal-fetal interface. Proc Natl Acad Sci USA 113, 5364–5369. doi:10.1073/pnas.1602886113
Figueroa, M.E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P.J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., Campagne, F., Mazumdar, M., Greally, J.M., Valk, P.J.M., Löwenberg, B., Delwel, R., Melnick, A., 2010. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell 17, 13–27. doi:10.1016/j.ccr.2009.11.020
Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., Svoboda, P., 2013. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell 155, 807–816. doi:10.1016/j.cell.2013.10.001
Fuentes, D.R., Swigut, T., Wysocka, J., 2018. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife 7, 861. doi:10.7554/eLife.35989
Gould, S.J., Vrba, E.S., 1982. Exaptation—a Missing Term in the Science of Form. Paleobiology 8, 4–15. doi:10.1017/S0094837300004310
Ha, M., Han, M.-E., Kim, J.-Y., Jeong, D.C., Oh, S.-O., Kim, Y.H., 2019. Prognostic role of TPD52 in acute myeloid leukemia: A retrospective multicohort analysis. J. Cell. Biochem. 120, 3672–3678. doi:10.1002/jcb.27645
Hashimoto, K., Suzuki, A.M., Santos, Dos, A., Desterke, C., Collino, A., Ghisletti, S., Braun, E., Bonetti, A., Fort, A., Qin, X.-Y., Radaelli, E., Kaczkowski, B., Forrest, A.R.R., Kojima, S., Samuel, D., Natoli, G., Buendia, M.A., Faivre, J., Carninci, P., 2015. CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors. Genome Research 25, 1812–1824. doi:10.1101/gr.191031.115
Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., Wheeler, T.J., 2016. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44, D81–9. doi:10.1093/nar/gkv1272
Imbeault, M., Helleboid, P.-Y., Trono, D., 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543, 550–554. doi:10.1038/nature21683
Jacques, P.-É., Jeyakani, J., Bourque, G., 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genetics 9,

e1003504. doi:10.1371/journal.pgen.1003504

Jang, H.S., Shah, N.M., Du, A.Y., Dailey, Z.Z., Pehrsson, E.C., Godoy, P.M., Zhang, D., Li, D., Xing, X., Kim, S., O'Donnell, D., Gordon, J.I., Wang, T., 2019. Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genet 51, 1–14. doi:10.1038/s41588-019-0373-3

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., Burden, F., Farrow, S., Cutler, A.J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., BLUEPRINT Consortium, Stunnenberg, H.G., Todd, J.A., Zerbino, D.R., Stegle, O., Ouwehand, W.H., Frontini, M., Wallace, C., Spivakov, M., Fraser, P., 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell 167, 1369–1384.e19. doi:10.1016/j.cell.2016.09.037

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915. doi:10.1038/s41587-019-0201-4

King T, Butcher S, & Zalewski L. 2017. Apocrita - High Performance Computing Cluster for Queen Mary University of London. *Zenodo*. http://doi.org/10.5281/zenodo.438045

Kroon, E., Krosl, J., Thorsteinsdottir, U., Baban, S., Buchberg, A.M., Sauvageau, G., 1998. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. EMBO J 17, 3714–3725. doi:10.1093/emboj/17.13.3714

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., Bourque, G., 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42, 631–634. doi:10.1038/ng.600

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M.A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K.D., Callen, D.F., Harte, M.F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I., Janz, M., Cockerill, P.N., Siebert, R., Dörken, B., Bonifer, C., Mathas, S., 2010. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat. Med. 16, 571–9– 1p following 579. doi:10.1038/nm.2129

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., McKernan, K.,

Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., la Bastide, de, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins,

F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. doi:10.1038/35057062

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi:10.1038/nmeth.1923

Lauer, S.J., Walker, D., Elshourbagy, N.A., Reardon, C.A., Levy-Wilson, B., Taylor, J.M., 1988. Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. Journal of Biological Chemistry 263, 7277–7286.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550–21. doi:10.1186/s13059-014-0550-8

Lynch, V.J., Leclerc, R.D., May, G., Wagner, G.P., 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet 43, 1154–1159. doi:10.1038/ng.917

Machnik, M., Cylwa, R., Kiełczewski, K., Biecek, P., Liloglou, T., Mackiewicz, A., Oleksiewicz, U., 2019. The expression signature of cancer-associated KRAB-ZNF factors identified in TCGA pan-cancer transcriptomic data. Mol Oncol 13, 701–724. doi:10.1002/1878-0261.12407

Medstrand, P., Landry, J.R., Mager, D.L., 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. Journal of Biological Chemistry 276, 1896–1903. doi:10.1074/jbc.M006557200

Mullighan, C.G., Kennedy, A., Zhou, X., Radtke, I., Phillips, L.A., Shurtleff, S.A., Downing, J.R., 2007. Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. Leukemia 21, 2000–2009. doi:10.1038/sj.leu.2404808

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., Gundem, G., Van Loo, P., Martincorena, I., Ganly, P., Mudie, L., McLaren, S., O'Meara, S., Raine, K., Jones, D.R., Teague, J.W., Butler, A.P., Greaves, M.F., Ganser, A., Döhner, K., Schlenk, R.F., Döhner, H., Campbell, P.J., 2016. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. 374, 2209–2221. doi:10.1056/NEJMoa1516192

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., Salzberg,

S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295. doi:10.1038/nbt.3122

Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J., Tuan, D., 2010. Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci USA 107, 12992–12997. doi:10.1073/pnas.1004139107

Piotti, K.C., Scognamiglio, T., Chiu, R., Chen, Y.-T., 2013. Expression of cancer/testis (CT) antigens in squamous cell carcinoma of the head and neck: evaluation as markers of squamous dysplasia. Pathol. Res. Pract. 209, 721–726. doi:10.1016/j.prp.2013.08.004

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T.W., Jaenisch, R., Trono, D., 2019. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell 24, 724–735.e5. doi:10.1016/j.stem.2019.03.012

Raviram, R., Rocha, P.P., Luo, V.M., Swanzey, E., Miraldi, E.R., Chuong, E.B., Feschotte, C., Bonneau, R., Skok, J.A., 2018. Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate. Genome Biol. 19, 1–19. doi:10.1186/s13059-018-1598-7

Ren, H., Chen, Z., Yang, L., Xiong, W., Yang, H., Xu, K., Zhai, E., Ding, L., He, Y., Song, X., 2019. Apolipoprotein C1 (APOC1) promotes tumor progression via MAPK signaling pathways in colorectal cancer. Cancer Manag Res 11, 4917–4930. doi:10.2147/CMAR.S192529

Song, L., Crawford, G.E., 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc 2010, pdb.prot5384–pdb.prot5384. doi:10.1101/pdb.prot5384

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., Wang, T., 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Research 24, 1963–1976. doi:10.1101/gr.168872.113

Takano, S., Yoshitomi, H., Togawa, A., Sogawa, K., Shida, T., Kimura, F., Shimizu, H., Tomonaga, T., Nomura, F., Miyazaki, M., 2008. Apolipoprotein C-1 maintains cell survival by preventing from apoptosis in pancreatic cancer cells. Oncogene 27, 2810–2822. doi:10.1038/sj.onc.1210951

Thomas, J., Perron, H., Feschotte, C., 2018. Variation in proviral content among human genomes mediated by LTR recombination. Mob DNA 9, 36–15. doi:10.1186/s13100-018-0142-3

Todd, C.D., Deniz, O., Taylor, D., Branco, M.R., 2019. Functional evaluation of

transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife 8, 1473. doi:10.7554/eLife.44344

Yang, X., Lu, B., Sun, X., Han, C., Fu, C., Xu, K., Wang, M., Li, D., Chen, Z., Opal, P., Wen, Q., Crispino, J.D., Wang, Q.-F., Huang, Z., 2018. ANP32A regulates histone H3 acetylation and promotes leukemogenesis. Leukemia 32, 1587–1597. doi:10.1038/s41375-018-0010-7

Yi, G., Wierenga, A.T.J., Petraglia, F., Narang, P., Janssen-Megens, E.M., Mandoli, A., Merkel, A., Berentsen, K., Kim, B., Matarese, F., Singh, A.A., Habibi, E., Prange, K.H.M., Mulder, A.B., Jansen, J.H., Clarke, L., Heath, S., van der Reijden, B.A., Flicek, P., Yaspo, M.-L., Gut, I., Bock, C., Schuringa, J.J., Altucci, L., Vellenga, E., Stunnenberg, H.G., Martens, J.H.A., 2019. Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes. CellReports 26, 1059–1069.e6. doi:10.1016/j.celrep.2018.12.098

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. doi:10.1186/gb-2008-9-9-r137

Zhang, Z.-H., Zhang, W., Zhou, J.-D., Zhang, T.-J., Ma, J.-C., Xu, Z.-J., Lian, X.-Y., Wu, D.-H., Wen, X.-M., Deng, Z.-Q., Lin, J., Qian, J., 2018. Decreased SCIN expression, associated with promoter methylation, is a valuable predictor for prognosis in acute myeloid leukemia. Mol. Carcinog. 57, 735–744. doi:10.1002/mc.22794

Zhu, G.-Z., Yang, Y.-L., Zhang, Y.-J., Liu, W., Li, M.-P., Zeng, W.-J., Zhao, X.-L., Chen, X.-P., 2017. High Expression of AHSP, EPB42, GYPC and HEMGN Predicts Favorable Prognosis in FLT3-ITD-Negative Acute Myeloid Leukemia. Cell. Physiol. Biochem. 42, 1973–1984. doi:10.1159/000479837