ENHANCER HIJACKING BY *EVI1*
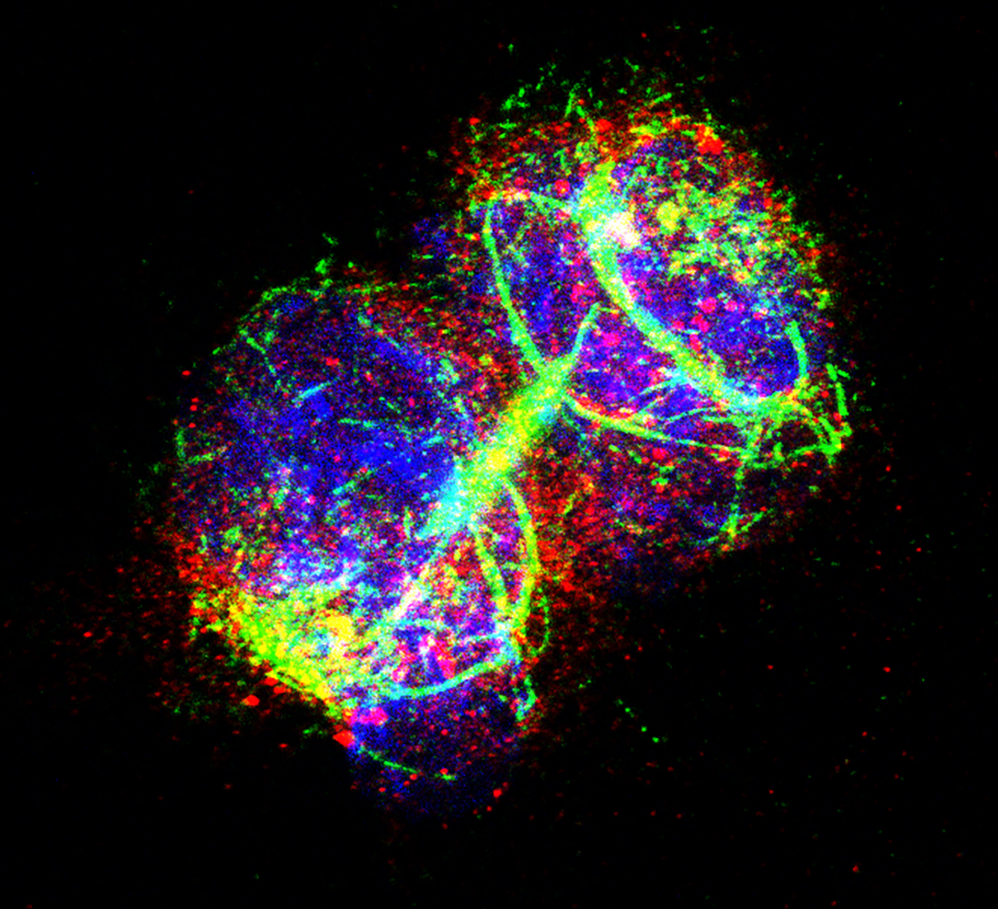
IN ACUTE MYELOID LEUKEMIA

Sophie Ottema

ENHANCER HIJACKING BY *EVI1* IN ACUTE MYELOID LEUKEMIA

SOPHIE OTTEMA

Enhancer Hijacking by *EVI1*

in Acute Myeloid Leukemia

Enhancer Overname door *EVI1*
in Acute Myeloide Leukemie

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. F.A. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Tuesday 21th of September 2021 at 10:30 am
by

Sophie Ottema
born in Spijkenisse, The Netherlands

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotors: Prof. dr. H.R Delwel
              Prof. dr. P. Sonneveld

Other members: Prof. dr. I. Touw
                  Prof. dr. P. Verrijzer
                  Dr. M. von Lindern

Co-promoter: Dr. L. Smeenk

CONTENTS

1

General introduction

GENERAL INTRODUCTION

Chromatin structure and functional organisation of the genome

The human genome consists of 23 chromosome pairs coding for the complete set of nucleic acid sequences for humans, apart from mitochondrial DNA. This set of chromosomal nucleic acid sequences contains coding and non-coding DNA. The coding DNA, referred to as genes, mostly encode for proteins. Some DNA is transcribed into non-coding functional RNA, like ribosomal RNA essential for translation of mRNA into proteins. Although the majority of the genome is non-coding, many of these non-coding genomic regions have a regulatory function.

To fit all chromosomal DNA into the cell's nucleus, the genome is assembled into nucleoprotein complexes known as nucleosomes. Nucleosomes consist of a combination of DNA and proteins; the most abundant proteins in this complex are the histones. Each 147 base pairs (bps) of DNA is wrapped around one histone complex, an octamer comprising of two copies of H2A, H2B, H3 and H4 histone subunits respectively [1, 2]. The genomic DNA packaged by histone complexes is together called chromatin. Within these histone-subunits we discriminate two common regions, the 'histone-fold' and a 'histone-tail'. The histone-fold is a structurally conserved motif responsible for the formation of histone heterodimers. The histone amino terminal tails, and the way they may be post-translationally modified, affect the structure and stability of the nucleosome cores [3]. These modifications play a role in higher order chromatin structure. For instance, they may define whether chromatin is open or closed. Accessibility of the DNA allows binding of transcriptional regulators to specific nucleotide motifs within regulatory elements in order to control transcription of genes. Site-specific modifications of histone-tails, such as acetylation or methylation of lysines (K), correlate with the biological function of the chromatin at that location [1, 2]. Histone-marks are widely used to predict the presence of 'open' active gene regulatory regions such as H3K9- or H3K27acetylation (H3K9ac or H3K27ac). Inactive or 'closed' chromatin is particularly defined by tri-methylation of H3K9 or H3K27 (H3K9me3 or H3K27me3) [2]. To identify active regulatory regions in the chromatin, so-called ChIP-seq experiments (see insert) have been applied throughout the studies presented in this thesis.

ChIP-seq (Chromatin Immunoprecipitation followed by next generation sequencing) is applied to determine protein-DNA interactions, typically of transcription factors and chromatin-associated proteins, genome wide. In short, the chromatin is cross-linked and sheared into <500bp fragments, followed by chromatin immune precipitation using an antibody directed against the DNA binding protein of interest. These antibodies may either identify a specific modification of chromatin, e.g. H3K27ac or H3K27me3, or they may be specific for transcription factors, such as MYB, GATA2, etc. After the pulldown of the protein-DNA complex, DNA is separated from the proteins, oligonucleotide adapters are added by ligation and the fragments are amplified. This process is followed by high throughput sequencing of the fragments. The sequences acquired are aligned to the genome. DNA fragments that are highly present in such an experiment are bound by the specific transcription factor, or present in regions with specific histone modification, to which the antibody used was directed against. These regions can be identified as peaks in the presented ChIP-seq tracks [4].

Gene promoters and enhancers

Histone modifications determine accessibility and consequently the activation state of a gene, and in particular of its regulatory elements: promoter and enhancer(s). Active promoters recruit transcription components, e.g. RNA polymerase needed to initiate the transcription of the corresponding gene. In order to increase the activity but also to tightly regulate the expression of a gene in the proper cell, gene promoters frequently interact with one or multiple enhancers [5]. Enhancers are located either up- or down-stream of the transcriptional start site (TSS) and are sometimes located relatively far from the promoter. For instance, an enhancer of the *MYC* gene which is particularly studied in this thesis, is located 1.7 Megabase (Mb) away from the *MYC* promoter. An active enhancer is typically bound by transcription factors that specifically recognize certain nucleotide motifs. At enhancers, multiple transcription factor motifs are present usually in a relatively small region of several hundred nucleotides; this allows multiple factors to bind enhancers as a complex. Transcription factors in turn recruit transcriptional coactivators like CBP (**CREB-Binding-Protein**)-p300 and subunits of the mediator complex. These protein complexes, together with the gene's promoter, activate and enhance transcription of the gene [6]. Binding of transcription factors and co-factors to enhancers and promoters can be identified by ChIP-seq as explained in the insert above. To enhance transcription, an enhancer is able to physically interact with the genes promoter even when this enhancer is located very distal of this gene [7]. Such 3D-genome interactions can be determined by techniques such as 4C-seq (see insert). This technology has been applied in this thesis to identify and study specific enhancer-promoter interactions.

4C-seq (Circularized Chromatin Conformation Capture by next generation sequencing) is a technology used to study the three-dimensional organization of the genome. In this thesis this technology is applied to identify chromatin enhancer-promoter interactions and to study which genomic regions or DNA binding proteins are involved in this interaction. In short, genomic regions that are close to each other (by protein-protein and/or DNA-protein interactions) can be crosslinked by fixation. When DNA is fragmented using restriction enzymes, small complexes of DNA and proteins are generated with sometimes two pieces of DNA coming from a promoter and an enhancer fragment in the same cross-linked complex. When a ligation step is introduced in the mixture, the DNA fragments in close proximity of such a cross-linked complex will be ligated to each other. When the crosslinks are removed hybrid DNA-molecules are left. These molecules are trimmed again using a secondary restriction enzyme and circularized once more in a second ligation step. These ligated DNA fragments are amplified using primers that recognize a so-called viewpoint (this may be a promotor or a predicted enhancer sequence) and analysed using next generation sequencing [8]. Alignment of the nucleotide sequences acquired to the human genome will then predict which genomic regions form a complex with the viewpoint.

Tissue specific transcriptional regulation

Frequently genes contain multiple enhancers. Such a feature would promote developmental robustness and the ability to survive environmental changes [9]. When multiple enhancers are present near a gene, some appear stronger than others; these enhancers usually have more defined transcription factor binding and can increase the quantity of transcriptional bursts [10]. A combined stretch of multiple enhancers acting as one big organised enhancer is often referred to as a super-enhancer. Super-enhancers are thought to be particularly important in tissue specific gene regulation, while they often control genes that code for key transcriptional regulators or cell identity genes. Super-enhancers can be defined by a large stretch of H3K27ac, often with high levels of mediator interaction and the binding of master transcription factors. Typically, genes associated with super-enhancers show higher expression levels compared to genes under the control of normal enhancers in a tissue specific manner [11]. Furthermore, super-enhancers can be affected or dynamically remodelled by environmental changes such as hormones of injury response signalling [12]. Altogether, it is not surprising that super-enhancers are frequently involved in the activation of oncogenes in cancer [13].

As mentioned, super-enhancers consist of a set close to each other located individual enhancers, all with their own unique function. It has been shown that super-enhancers can be hierarchically organised with one particular individual enhancer module acting as an epicentric-enhancer affecting gene expression more strongly than the other enhancer

modules within the super-enhancer [12, 14]. In fact, it is possible that in one organ or cell type a certain enhancer module drives gene expression, whereas in another cell type another module may have that role. An example of a gene that has multiple enhancers is *CEBPA* [15, 16]. *CEBPA* is expressed in many different cell types in a human body. Avellino and colleagues demonstrated that out of 14 enhancers located near *CEBPA*, only one appeared essential for expression of the gene in hematopoietic progenitors. Most likely the other enhancers are active in different organs and cell types. To investigate which enhancer is responsible for expression of a gene in a certain cell type, genome editing using CRISPR-Cas9 technology is a very powerful tool. CRISPR-Cas9 technology (see insert) has been used to study the biological function of the individual enhancer modules within super-enhancers in several studies and also in this thesis we make grateful use of this technique [14, 15, 17, 18].

CRISPR-Cas9 (clustered regulatory interspaced short palindromic repeats – Cas9 nuclease) technology to facilitate genome engineering. The RNA-guided Cas9 nuclease can simply be directed to cut the target DNA of interest in a living cell by defining a 20-nt sequence within a single-guide RNA (sgRNA). Cas9 induces double strand DNA breaks at the locus of interest. The cell can repair these breaks in two ways: by making use of the non-homologous end-joining (NHEJ) pathway, or by homology directed repair (HDR). NHEJ is very error-prone and can result in random insertion/deletion (Indel) mutations at the cut-site. This feature is used in our advantage to create mutations in important regulatory or coding regions in the genome, leading to defective gene expression or gene control. Similarly, we can use this method to make bigger deletions by simply directing Cas9 by two sgRNAs to delete a desired genomic region. To edit the genome more precise, the HDR pathway can be leveraged by supplying the cell with a repair template in the form of a plasmid. This allows high fidelity and precise genome editing [19].

Hematopoiesis: how to discriminate between distinct hematopoietic cell types?

The formation of blood cells in the bone marrow, known as hematopoiesis, is a highly active and regenerative process required to provide the human body with oxygen, a functional immune and hemostatic-system [20]. Proliferation and differentiation of hematopoietic cells is thought to occur in a hierarchical manner starting with a pool of hematopoietic stem cells (HSC), which are located in the bone marrow [21]. These stem cells have the capacity to self-renew in order to maintain the regenerative capacity of the hematopoietic system and to differentiate into committed myeloid and lymphoid progenitor cells, which subsequently can give rise to all the distinct differentiated blood cells [22]. The different hematopoietic stem and progenitor cells (HSPCs) can (in part) be discriminated from each other using different monoclonal antibodies directed to various cell surface markers [23]. The cell surface marker CD34 was described at first to be present on HSCs [24]. This marker is still used clinically for the enrichment of HSCs for bone marrow transplants [25]. However, CD34 is not only present on HSCs but marks all HSPCs cells starting from HSC with long-term repopulation potential throughout the oligopotent and unipotent progenitors [25]. Using a variation of (surface) markers, HSCs, progenitors and all mature cells of the different hematopoietic lineages can be distinguished by flow cytometry analysis [23]. Flow cytometry (see insert), including the analysis of cell surface markers, is used in this thesis to demonstrate the origin and differentiation potential of leukemic cells.

Flow cytometry is a powerful technique that measures the physical properties of cells within a high-velocity fluid stream passing through a laser beam. Flow cytometry can provide us with information about the properties of cells like morphology, cell cycle, differentiation or lineage stage. It has the potential to analyse the properties of single cells or subgroups of cells within a population [26]. For example, surface-markers are widely used to characterize HSPCs (e.g. CD34+) and more differentiated blood cells (e.g. granulocytes: CD15+). Also genetically modified cells that possess a fluorescent protein as a marker, are often used to analyse corresponding gene expression patterns in cell populations (e.g. eGFP as read out for EVI1 levels in this thesis). FACS (Fluorescent activated cell sorting) is a powerful addition to flow cytometry that allows us to sort cells based on their properties defined by the markers used.

Hematopoiesis: transcriptional regulation

The enormous demand of the body for mature blood cells challenges the balance between proliferation and differentiation of HSPCs. In order to keep a healthy balance throughout the lifetime of a human, these processes are highly regulated. Cross talk between the HSPCs, the bone marrow niche and regulatory factors released by different cell types regulate these processes [27]. In response to extracellular signals, HSPCs undergo lineage specific epigenetic changes that direct gene expression programs to drive cells to differentiate into mature blood cells [28, 29]. The binding of specific transcription factors and transcriptional regulators to promoter and enhancer regions of key genes involved in hematopoietic development are essential for gene expression regulation [30, 31]. Gene expression programs active in specific cell types require a unique set of transcription factors. For HSPCs a set of transcription factors has been described to be active and to occupy regulatory loci essential for genes driving myeloid development. This set of transcription factors is often referred to as the heptad transcription factors and is composed of the factors: RUNX1, GATA2, ERG, LMO1, LYL1, FLY and SCL [32-34]. Key transcription factors like these are of

particular interest in this thesis, since the leukemic cells central in our studies appear from HSPC origin (marked by CD34) and may therefor also depend on these heptad factors. We particularly investigated how regulatory elements, in particular the ones driving the expression of the oncogene EVI1, are bound (ChIP-seq; see first insert) and activated by those transcription factors.

Acute myeloid leukemia

When irregularities occur in HSPC development, severe diseased conditions may arise. For example, when progenitors fail to fully differentiate into functional mature blood cells and these committed cells do not lose their ability to proliferate, leukemic transformation may be initiated [21]. These abnormalities frequently occur upon mutations in genes required for normal development. The most common form of acute leukemia in adults is acute myeloid leukemia (AML). AML patients commonly present with a combination of leucocytosis and signs of bone marrow failure. Mutations or larger chromosomal aberrations involving myeloid developmental genes or their regulatory loci, cause abnormal gene regulation resulting in defects in the normal developmental processes. The constant production cells that are unable to differentiate leads to an accumulation of poorly differentiated cells in marrow and blood. AML patients are diagnosed by the presence of at least 20% poorly differentiated malignant cells (blasts) of a myeloid (progenitor) origin in their bone marrow and frequently also in the peripheral blood. If left untreated, infections and bleedings may occur, followed by death typically within a few months [35].

AML is a highly heterogeneous disease, meaning that many different subtypes can be recognized. Each subtype contains unique cytogenetic or subtler genetic defects often involving regulatory genes. Genetic abnormalities are used for risk stratification of AML and for treatment choice [36]. Many groups and sub-groups have been defined by the World health Organisation (WHO), ranging from a favourable prognosis, to patients with cytogenetic aberrations predictive for a high risk of treatment failure [36]. The subgroup of AML with 3q26-rearrangements resulting in aberrant expression of the EVI1 gene, are among the patients that are very difficult to treat. The majority of 3q26-rearranged EVI1+ AMLs are refractory; upon chemotherapy the patients generally do not reach remission [37-39]. Not reaching remission means in practice that successful treatment options are very limited for these patients, resulting in a disease that is almost universally fatal to date [40].

EVI1 and MDS1-EVI1 in leukemia

The typical refractory nature of 3q26-rearranged AMLs is thought to be (partially) the result of aberrant EVI1 expression [37-39, 41]. The EVI1 (ecotropic virus integration site 1) gene encodes a transcription factor located at chromosomal band 3q26.2 in the MECOM locus (MDS1 and EVI1 complex locus). This zinc finger DNA-binding protein is involved in normal hematopoietic development and it is an oncogene involved in leukemic transformation

when expression levels are not controlled correctly. Two major transcripts, each with a distinct transcriptional start sites can be expressed resulting in variability in the EVI1 mRNA. One transcript variant, MDS1-EVI1, results in a longer protein with a function that is not completely clear [42]. The MDS1-EVI1 protein contains a so-called PR-domain at its N-teminus with unclear function, which is absent in EVI1. Otherwise MDS1-EVI1 and EVI1 are identical [43]. It has been proposed that MDS1-EVI1 has opposite function compared to EVI1 and might therefore be considered to be a tumor-suppressor [44]. This idea is supported by studies in which a prognostic significance for EVI1 overexpression but not MDS1-EVI1 overexpression was found in a large group of de novo AML patients without 3q26-rearrangements [41]. Moreover, EVI1 is particularly overexpressed in AMLs with a 3q26-rearrangement, whereas MDS1-EVI1 is specifically not activated upon translocation.

In healthy individuals and in the bone marrow of mice, EVI1 is expressed in dormant hematopoietic stem cells [45]. In fact, EVI1 appears to be particularly expressed in dormant hematopoietic stem cells (HSCs) and Evi1 negative murine hematopoietic stem cells complete lost their capability to repopulate a bone marrow upon transplantation [45]. The importance of EVI1 in human hematopoietic development is evident as well. Several examples of new-borns have been reported with severe bone marrow failure associated with mutations or even complete loss of the MECOM locus [46-49]. EVI1 expression is lost during normal hematopoietic development [50], demonstrating that the role of the gene is really limited to stem cell maintenance. In agreement with this is the EVI1 overexpression observed in AML. Adding to that, when EVI1 expression in these AML cells is depleted (partial) differentiation is observed [17]. The mechanism by which EVI1 is of aberrantly expressed in 3q26-rearranged AMLs and how we can possibly intervene with this is the theme of this thesis.

Chromosomal rearrangements and EVI1 overexpression in leukemia

AML patients with chromosomal aberrations in the 3q26 locus are often found to overexpress EVI1 but not MDS1-EVI1. Patients with 3q26 cytogenetic aberrations, including an inversion inv(3)(q21;q26) or translocation t(3;3)(q21;q26), have a high risk of treatment failure [36]. These patients overexpress the gene EVI1 located at 3q26 due to the rearrangement of a GATA2 enhancer located at 3q21 [17, 51]. Studies investigating this particular enhancer rearrangement resulting in oncogenic EVI1 expression are described in Chapter 3 and 4 of this thesis. Besides inv(3)/t(3;3), many other EVI1+ AML cases with 3q26-rearrangements have been reported, for example translocations t(2;3)(p21;q26), t(3;7)(q26;q24), t(3;6)(q26;q11) and t(3;8)(q26;q24) [37, 44, 52-58]. We hypothesized that in these AMLs, enhancers of other myeloid genes had been translocated driving EVI1 expression. This hypothesis was particularly tested in the studies reported in Chapter 5. Besides the distinct reciprocal 3q26-rearranged AMLs, other more complex chromosomal aberrations are found among EVI1 overexpressing AML patients, a unique cohort with atypical 3q26-rearrangements is studied

in Chapter 2 of this thesis [52]. We hypothesized that these leukemias should be considered as being 3q26-rearranged AMLs as well. This is of particular importance, as it would mean that each of these leukemias could be treated similarly as inv(3)/t(3;3) if better treatment options would be available in the future.

To identify specific breakpoints of all these different rearrangements involving the 3q26 locus, a technique that we will refer to as 3q-seq (see insert) is applied throughout all studies in this thesis [17]. This approach is essential for the identification of a putative enhancer being translocated to *EVI1* in each of these AMLs. The genome modification experiments, combined with flow cytometry, ChIP-seq and 4C-seq have been essential to answer these questions.

3q-seq (Targeted sequencing of chromosomal region 3q21.1-3q26.2) is a specific next generation sequencing technique used to identify alterations near the *EVI1* locus in 3q26-rearranged AMLs. In short, genomic DNA of AML patients harbouring 3q26-rearrangements was fragmented and sample libraries were constructed. Target sequences of the chromosomal region of interest 3q21.1-3q26.2 were captured using custom in-solution oligonucleotide baits. The amplified captured libraries were paired-end sequenced. Using this approach chromosomal aberrations, translocation or inversions involving the 3q21.1 or 3q26.2 locus were identified [17]. A similar approach was used to identify the breakpoints in the other 3q26-rearranged AMLs in the studies presented in this thesis and determine the loci that translocated to *EVI1*

Scope of the thesis

The main objective of this thesis is to further unravel how a translocated enhancer can hyper-activate oncogenic expression of *EVI1* and what the mechanism is of *EVI1* activation in AMLs which do not carry a classical inv(3)/t(3;3) but show other (complex) 3q26-rearrangements. We hypothesized that enhancer activated *EVI1* expression in AML is, although caused by different chromosomal rearrangements, mechanistically highly similar.

In **Chapter 2**, we investigated a unique cohort of 3q26-rearranged primary AML samples with a variety of aberrations in the *MECOM* locus. We asked the question to what extend these leukemia's resembled *EVI1*+ inv(3)/t(3;3)(q21;q26) AML by identifying translocation partner locus's, analysing copy number variants and gene expression patterns.

In **Chapter 3**, the mechanism of *EVI1* deregulation by the *GATA2* enhancer was further dissected by a CRISPR-Cas9 based enhancer scan in an inv(3)(q21;q26) model system. We addressed the question which motifs in the enhancer and consequently which transcription factors drive *EVI1* expression. The data not only provides insight into which transcription factor binding sites in the enhancer drive expression, but also some light is shed on the question why the enhancer is a much stronger activator of *EVI1* at the translocated allele then of *GATA2* on the normal allele. The study presented in this chapter also provided a suggestion for *EVI1* transcriptional interference, by targeting a specific DNA-binding motif within the translocated enhancer.

In **Chapter 4**, a functional genomic analyses on a cohort of inv(3)/t(3;3) patient and cell line samples, followed by *in vitro* studies using CRISPR-Cas9 were performed to investigate a potential role of breakpoint associated retro-elements in *EVI1* deregulation.

In **Chapter 5**, we studied the mechanism of *EVI1* overexpression in AMLs with other recurrent translocation involving the 3q26 locus, with specific a focus on patients with a t(3;8)(q26;q24). We defined in a cohort of t(3;8) AMLs the breakpoints at 3q26 and 8q24 and generated based on this patient data an *in vitro* t(3;8) model. This system provided us with a tool to study the mechanism of *EVI1* deregulation using CRISPR-Cas9 technology and next generation sequencing (NGS) methods, pointing us to a common mechanism for enhancer hijack by *EVI1* in 3q26-rearranged leukemia's.

In the last chapter our findings are summarized and discussed in light of the current literature. This discussion is followed by thoughts on how to further investigate mechanisms of oncogene deregulation and how to possibly interfere in these mechanisms of aberrant gene activation, in order to treat malignancies like AML more efficiently and specifically in the future.

REFERENCES

1. Munshi, A., et al., *Histone modifications dictate specific biological readouts.* J Genet Genomics, 2009. **36**(2): p. 75-88.

2. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

3. Iwasaki, W., et al., *Corrigendum to: Contribution of histone N-terminal tails to the structure and stability of nucleosomes.* FEBS Open Bio, 2018. **8**(9): p. 1567.

4. Abcam, C.g.e.a.b. *ChIP guide: epigenetic applications by Abcam*. Retrieved 22 december 2020 [cited 2020 22 december]; Available from: https://www.abcam.com/epigenetics/studying-epigenetics-using-chip.

5. Andersson, R. and A. Sandelin, *Determinants of enhancer and promoter activities of regulatory elements.* Nature Reviews Genetics, 2020. **21**(2): p. 71-87.

6. Pennacchio, L.A., et al., *Enhancers: five essential questions.* Nature reviews Genetics, 2013. **14**(4): p. 288-295.

7. Latchman, D.S., *Transcription factors: an overview.* International journal of experimental pathology, 1993. **74**(5): p. 417-422.

8. Krijger, P.H.L., et al., *4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis.* Methods, 2020. **170**: p. 17-32.

9. Perry, M.W., et al., *Shadow enhancers foster robustness of Drosophila gastrulation.* Current biology : CB, 2010. **20**(17): p. 1562-1567.

10. Long, H.K., S.L. Prescott, and J. Wysocka, *Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.* Cell, 2016. **167**(5): p. 1170-1187.

11. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.

12. Ko, J.Y., S. Oh, and K.H. Yoo, *Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development.* Molecules and cells, 2017. **40**(3): p. 169-177.

13. Sengupta, S. and R.E. George, *Super-Enhancer-Driven Transcriptional Dependencies in Cancer.* Trends Cancer, 2017. **3**(4): p. 269-281.

14. Huang, J., et al., *Dissecting super-enhancer hierarchy based on chromatin interactions.* Nat Commun, 2018. **9**(1): p. 943.

15. Avellino, R., et al., *An autonomous CEBPA enhancer specific for myeloid-lineage priming and neutrophilic differentiation.* Blood, 2016. **127**(24): p. 2991-3003.

16. Avellino, R. and R. Delwel, *Expression and regulation of C/EBPα in normal myelopoiesis and in malignant transformation.* Blood, 2017. **129**(15): p. 2083-2091.

17. Groschel, S., et al., *A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia.* Cell, 2014. **157**(2): p. 369-381.

18. Shin, H.Y., et al., *Hierarchy within the mammary STAT5-driven Wap super-enhancer.* Nat Genet, 2016. **48**(8): p. 904-911.

19. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11): p. 2281-2308.

20. Doulatov, S., et al., *Hematopoiesis: a human perspective.* Cell Stem Cell, 2012. **10**(2): p. 120-36.

21. Seita, J. and I.L. Weissman, *Hematopoietic stem cell: self-renewal versus differentiation.* Wiley Interdiscip Rev Syst Biol Med, 2010. **2**(6): p. 640-53.

22. Jordan, C.T. and I.R. Lemischka, *Clonal and systemic analysis of long-term hematopoiesis in the mouse.* Genes Dev, 1990. **4**(2): p. 220-32.

23. Ali, M.A.E., et al., *Functional dissection of hematopoietic stem cell populations with a stemness-monitoring system based on NS-GFP transgene expression.* Sci Rep, 2017. **7**(1): p. 11442.

24. Civin, C.I., et al., *Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells.* J Immunol, 1984. **133**(1): p. 157-65.

25. Hughes, M.R., et al., *A sticky wicket: Defining molecular functions for CD34 in hematopoietic cells.* Exp Hematol, 2020. **86**: p. 1-14.

26. Morteza Jalali, F.Y.L.S.a.M.J., *Basic Science Methods for Clinical researchers*. 2017.

27. Morrison, S.J. and D.T. Scadden, *The bone marrow niche for haematopoietic stem cells.* Nature, 2014. **505**(7483): p. 327-34.

28. Arzate-Mejía, R.G., D. Valle-García, and F. Recillas-Targa, *Signaling epigenetics: novel insights on cell signaling and epigenetic regulation.* IUBMB Life, 2011. **63**(10): p. 881-95.

29. Atlasi, Y. and H.G. Stunnenberg, *The interplay of epigenetic marks during stem cell differentiation and development.* Nat Rev Genet, 2017. **18**(11): p. 643-658.

30. Heinz, S. and C.K. Glass, *Roles of lineage-determining transcription factors in establishing open chromatin: lessons from high-throughput studies.* Curr Top Microbiol Immunol, 2012. **356**: p. 1-15.

31. Bottardi, S., et al., *Lineage-specific transcription factors in multipotent hematopoietic progenitors: a little bit goes a long way.* Cell Cycle, 2007. **6**(9): p. 1035-9.

32. Beck, D., et al., *Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes.* Blood, 2013. **122**(14): p. e12-22.

33. Diffner, E., et al., *Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia.* Blood, 2013. **121**(12): p. 2289-300.

34. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.* Cell Stem Cell, 2010. **7**(4): p. 532-44.

35. De Kouchkovsky, I. and M. Abdul-Hay, *'Acute myeloid leukemia: a comprehensive review and 2016 update'.* Blood Cancer J, 2016. **6**(7): p. e441.

36. Estey, E.H., *Acute myeloid leukemia: 2019 update on risk-stratification and management.* Am J Hematol, 2018. **93**(10): p. 1267-1291.

37. Lugthart, S., et al., *Clinical, Molecular, and Prognostic Significance of WHO Type inv(3)(q21q26.2)/t(3;3) (q21;q26.2) and Various Other 3q Abnormalities in Acute Myeloid Leukemia.* Journal of Clinical Oncology, 2010. **28**(24): p. 3890-3898.

38. Lugthart, S., et al., *High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated.* Blood, 2008. **111**(8): p. 4329-4337.

39. Gröschel, S., et al., *High EVI1 Expression Predicts Outcome in Younger Adult Patients With Acute Myeloid Leukemia and Is Associated With Distinct Cytogenetic Abnormalities.* Journal of Clinical Oncology, 2010. **28**(12): p. 2101-2107.

40. Döhner, H., D.J. Weisdorf, and C.D. Bloomfield, *Acute Myeloid Leukemia.* New England Journal of Medicine, 2015. **373**(12): p. 1136-1152.

41. Barjesteh van Waalwijk van Doorn-Khosrovani, S., et al., *High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients.* Blood, 2003. **101**(3): p. 837-845.

42. Wieser, R., *The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions.* Gene, 2007. **396**(2): p. 346-57.

43. Fears, S., et al., *Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family.* Proc Natl Acad Sci U S A, 1996. **93**(4): p. 1642-7.

44. Nucifora, G., L. Laricchia-Robbio, and V. Senyuk, *EVI1 and hematopoietic disorders: History and perspectives.* Gene, 2006. **368**: p. 1-11.

45. Kataoka, K., et al., *Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity.* J Exp Med, 2011. **208**(12): p. 2403-16.

46. Bouman, A., et al., *Congenital thrombocytopenia in a neonate with an interstitial microdeletion of 3q26.2q26.31.* Am J Med Genet A, 2016. **170A**(2): p. 504-509.

47. van der Veken, L.T., et al., *Lethal neonatal bone marrow failure syndrome with multiple congenital abnormalities, including limb defects, due to a constitutional deletion of 3' MECOM.* Haematologica, 2018. **103**(4): p. e173-e176.

48. Weizmann, D., et al., *New MECOM variant in a child with severe neonatal cytopenias spontaneously resolving.* Pediatr Blood Cancer, 2020. **67**(5): p. e28215.

49. Germeshausen, M., et al., *MECOM-associated syndrome: a heterogeneous inherited bone marrow failure syndrome with amegakaryocytic thrombocytopenia.* Blood Adv, 2018. **2**(6): p. 586-596.

50. Bindels, E.M., et al., *EVI1 is critical for the pathogenesis of a subset of MLL-AF9-rearranged AMLs.* Blood, 2012. **119**(24): p. 5838-49.

51. Yamazaki, H., et al., *A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression.* Cancer Cell, 2014. **25**(4): p. 415-27.

52. Ottema, S., et al., *Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia.* Blood, 2020. **136**(2): p. 224-234.

53. Lin, P., et al., *Translocation (3;8)(q26;q24): a recurrent chromosomal abnormality in myelodysplastic syndrome and acute myeloid leukemia.* Cancer Genetics and Cytogenetics, 2006. **166**(1): p. 82-85.

54. Lennon, P.A., et al., *Aberrant EVI1 expression in acute myeloid leukemias associated with the t(3;8)(q26;q24).* Cancer Genetics and Cytogenetics, 2007. **177**(1): p. 37-42.

55. De Braekeleer, M., et al., *Breakpoint heterogeneity in (2;3)(p15–23;q26) translocations involving EVI1 in myeloid hemopathies.* Blood Cells, Molecules, and Diseases, 2015. **54**(2): p. 160-163.

56. Trubia, M., et al., *Characterization of a recurrent translocation t(2;3)(p15–22;q26) occurring in acute myeloid leukaemia.* Leukemia, 2006. **20**(1): p. 48-54.

57. Storlazzi, C.T., et al., *A novel chromosomal translocation t(3;7)(q26;q21) in myeloid leukemia resulting in overexpression of EVI1.* Annals of Hematology, 2004. **83**(2): p. 78-83.

58. Tang, G., et al., *t(3;8)(q26.2;q24) Often Leads to MECOM/MYC Rearrangement and Is Commonly Associated with Therapy-Related Myeloid Neoplasms and/or Disease Progression.* J Mol Diagn, 2019. **21**(2): p. 343-351.

2

Atypical 3q26/*MECOM* rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia

Sophie Ottema[1,2,*], Roger Mulet-Lazaro[1,2,*], H. Berna Beverloo[3], Claudia Erpelinck[1,2], Stanley van Herk[1,2], Robert van der Helm[3], Marije Havermans[1,2], Tim Grob[1], Peter J. M. Valk[1], Eric Bindels[1], Torsten Haferlach[4], Claudia Haferlach[4], Leonie Smeenk[1,2], Ruud Delwel[1,2]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands, [2] Oncode Institute, Erasmus University Medical Center, Rotterdam, The Netherlands, [3] Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands, [4] MLL Munich Leukemia Laboratory, Munich, Germany.

*These authors contributed equally to this work

Running Title: oncogenic *EVI1* upregulation in 3q26-rearranged AML

Key points:
- *EVI1* overexpression, super-enhancer hijacking, lack of *MDS1-EVI1* and frequent *GATA2* deficiency define 3q26/MECOM rearranged AML
- 3q26/*MECOM* rearranged AML is a single entity, including but not limited to inv(3)/t(3;3), and requires specialized diagnostic assays

ABSTRACT

Acute myeloid leukemia (AML) with inv(3)/t(3;3)(q21q26) is a distinct WHO recognized entity, characterized by its aggressive course and poor prognosis. In this subtype of AML, the translocation of a *GATA2* enhancer (3q21) to *MECOM* (3q26) results in overexpression of the *MECOM* isoform *EVI1* and monoallelic expression of *GATA2* from the unaffected allele. The full-length *MECOM* transcript, *MDS1-EVI1*, is not expressed as the result of the 3q26 rearrangement. Besides the classical inv(3)/t(3;3), a number of other 3q26/*MECOM* rearrangements with poor treatment response have been reported in AML. Here we demonstrate, in a group of 33 AML patients with atypical 3q26 rearrangements, *MECOM* involvement with *EVI1* overexpression, but no or low *MDS1-EVI1* levels. Moreover, the 3q26 translocations in these AML patients often involve super-enhancers of genes active in myeloid development (e.g. *CD164*, *PROM1*, *CDK6 or MYC*). In more than 50% of these cases allele specific *GATA2* expression was observed, either by copy number loss or by an unexplained allelic imbalance. Altogether, atypical 3q26 recapitulate the main leukemic mechanism of inv(3)/t(3;3) AML, namely *EVI1* overexpression driven by enhancer hijacking, absent *MDS1-EVI1* expression and potential *GATA2* involvement. Therefore, we conclude that both atypical 3q26/*MECOM* and inv(3)/t(3;3) can be classified as a single entity of 3q26-rearranged AMLs. Routine analyses determining *MECOM* rearrangements, *EVI1* and *MDS1-EVI1* expression are required to recognize 3q-rearranged AML cases.
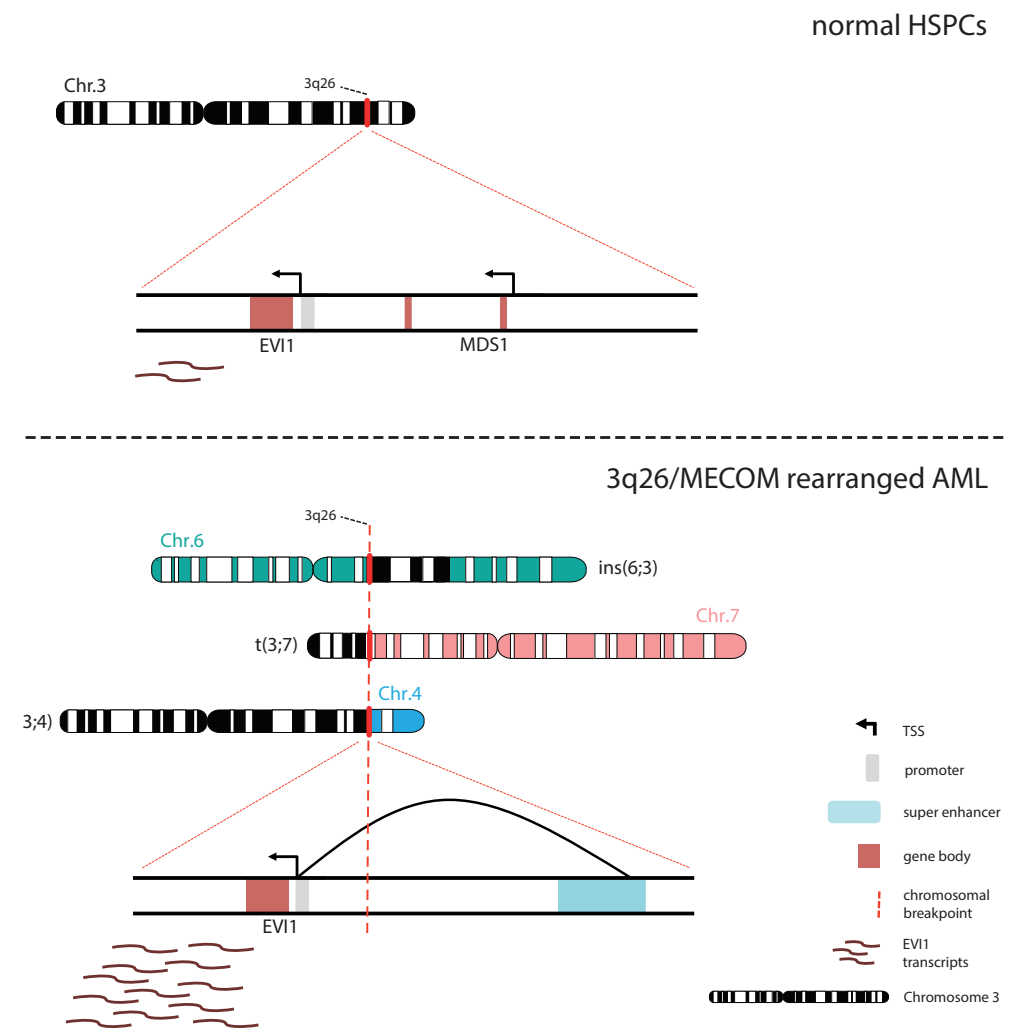
INTRODUCTION

Risk classification of patients with acute myeloid leukemia (AML) is based on the various genetic and epigenetic abnormalities previously identified and determines choice of treatment[1-5]. Understanding the biological consequences of these abnormalities is essential to develop new treatments for AML, especially for chemotherapy resistant subtypes. AML with inv(3)(q21q26) or t(3;3)(q21;q26)[6-9], henceforth referred to as inv(3)/t(3;3), is one of such subgroups with very poor response to therapy and a very aggressive course.

Recurrent translocations and inversions in AML most frequently generate oncogenic fusion genes[10-12]. However, in the case of an inv(3) or t(3;3), both rearrangements cause the translocation of an enhancer of the *GATA2* gene, located at 3q21, to the *MECOM* locus at chromosome 3q26[13,14]. *MECOM* encodes the transcript isoforms *MDS1-EVI1* and *EVI1*, which can be transcribed from two distinct promoters. In inv(3)/t(3;3) AML, the translocated *GATA2* enhancer causes overexpression *of EVI1*, but not of *MDS1-EVI1*. Translocation of the *GATA2* oncogenic enhancer in AML with an inv(3)/t(3;3) leads to *EVI1* upregulation and simultaneously abolishes *GATA2* expression from the rearranged allele[13,14]. Notably, germline haploinsufficiency or loss-of-function mutations in *GATA2* are the underlying causes of a wide spectrum of disorders, including MonoMAC and Emberger syndrome[15-18]. Those patients have a severely increased chance to develop AML compared to healthy individuals. Together with the fact that *GATA2* encodes a transcription factor essential for normal hematopoietic development[19], this suggests that loss of one *GATA2* allele increases the transforming ability of *EVI1* in chromosome 3q26-rearranged AMLs.

In a previous study of newly diagnosed 6515 AML patients, a group of leukemias with undefined 3q abnormalities was reported[9]. Although these patients did not present with a classical inv(3)/t(3;3), they also exhibited frequent *EVI1* overexpression and a very poor survival[9]. Here we addressed the question whether patients within this group harboring rearrangements at 3q26, resemble inv(3)/t(3;3) AML. Our study identifies critical similarities in the pathophysiology of both atypical 3q26 and inv(3)/t(3;3) AMLs: myeloid enhancer-driven *EVI1* overexpression, accompanied by low or no *MDS1-EVI1* transcription and, in approximately 50% of the cases *GATA2* deficiency. Given their clinical and biological similarities, we conclude that atypical 3q26-rearranged AML and inv(3)/t(3;3) constitute a single entity.

RESULTS

Frequent *MECOM* rearrangements in atypical 3q26 AML

To study *MECOM* involvement we performed Fluorescent In Situ Hybridization (*MECOM*-FISH, Figure S1A) in 33 AML patient samples whose karyotypes do not harbor a classical inv(3)/t(3;3) but had rearrangements at 3q26. These cases were classified as atypical 3q26 rearranged AML (Table 1, Table S1). A rearranged FISH pattern was found in 25 cases, i.e. a part of the *MECOM* signal was found translocated from chromosome band 3q26 to another locus in the genome (Table 1, Figure S1B). SNP-array hybridizations revealed losses or gains on 3q26 or and/or partner loci in 7 of these 25 cases (Table 1, Table S1). In 12 of these 25 *MECOM* rearranged cases, no copy number gains (CNG) or losses (CNL) were found, which is in agreement with the existence of balanced translocations (Table 1). In the remaining 6 it was unclear whether rearrangements were balanced or not. In 4 of the total cohort of 33 cases (#HF-13, 14, 15, 16), FISH analysis suggested amplification of the 3q26/*MECOM* locus (Table 1), which was confirmed by SNP-array (Table 1). In 2/33 atypical 3q26 samples (#TG-04, TG-06) no clear *MECOM* rearrangements could be detected. Together, these results point to common *MECOM* involvement in AML with atypical 3q26 rearrangements.

High *EVI1* mRNA levels transcribed from one allele in atypical 3q26 AML

Routine diagnostic RT-PCR[8] (Table 1) showed *EVI1* overexpression in 30 out of 33 atypical 3q26 cases. RNA sequencing (n=26) revealed that on average, *EVI1* transcript levels were over 9 fold higher (p=3.00e09) in atypical 3q26 AML than in control non-3q26 AML (Figure 1A). To discriminate between the two *MECOM* alleles, we assessed single nucleotide variants (SNVs) in RNA-seq and 3q-capture data. We could identify informative heterozygous SNPs in the DNA of 15 patients out of 33 patients and demonstrated equal distribution of the two *EVI1* alleles (Figure 1B, left bar in red and blue). RNA-seq data demonstrated monoallelic *EVI1* mRNA expression in those 15 leukemia samples (Figure 1B, right bar in red), strongly suggesting that *EVI1* is only transcribed from the rearranged *MECOM* allele in atypical 3q26 AML.

Low *MDS1-EVI1* expression is a common feature of atypical 3q26 AML

Although two messenger RNAs can be transcribed from the *MECOM* locus, i.e. *MDS1-EVI1 (ME)* and *EVI1* (Figure S1D)[20,21], inv(3)/t(3;3) AMLs are *EVI1+/ME-*. Similarly, in 29 out of 33 atypical 3q26 AML samples *MDS1-EVI1* transcripts were absent or expressed at very low levels as reported for inv(3)/t(3;3) leukemias (Table 1 and Figure 1C).

Frequent disruption of *MDS1* in atypical 3q26 AML underlies its low expression

In 23 out of 33 cases, we were able to exactly define the breakpoints within the *MECOM* locus (Figure 1D). Breakpoints occurred either upstream (N=17) or downstream (N=6) of the

EVI1 gene. In 15 out of the 17 cases with an upstream *EVI1* rearrangements, the breakpoints occurred between the MDS1 and EVI1 promoter (Figure 1D), as was reported in AML with a translocation t(3;3)(q21;q26)[15]. In those AMLs, the *MDS1* promoter has been dislocated due to the translocation, which avoids transcription of the long-form *MDS1-EVI1* (Figure S1D and 1C). In the 2 other AMLs (#-SO-23, HF-21) with a 5'-*EVI1* breakpoint, the rearrangements occurred upstream of the *MDS1* promoter. Accordingly, one of those patients (#SO-23) showed *EVI1+/ME+* expression. In the other case (#HF-21), neither *EVI1* nor *MDS1-EVI1* was detectable. The 6 cases with breakpoints 3' of *EVI1* showed an *EVI1+/ME-* expression pattern. Why 3q26 rearrangements with downstream breakpoints, as in AML with inv(3), show no or low *MDS1-EVI1* levels remains unresolved. CNV analysis of the 3q-capture DNA-seq and the SNP-array hybridizations revealed deletions within the *MDS1* region in 6 atypical 3q26 AML patients: #HF-15, HF-16, HF-20, HF21, TG-05, and SO-11 (Figure 2 and S3A, Table 1 and S1). Notably, these deletions underlie the loss of *MDS1* expression in #HF-16 and HF-21, where this cannot be explained by a translocation. *EVI1* exons were never deleted in those samples, and in fact were amplified in 3 of them (#HF-15, HF-16, TG-05). Altogether, the data strongly support the hypothesis that *EVI1* and not *MDS1-EVI1* expression is essential in transformation of 3q26-rearranged AMLs.

Unique rearrangements between *MECOM* and myeloid genes in atypical 3q26 AMLs
In 20/33 atypical 3q26 cases, the translocated partner locus of *MECOM*/3q26 could be identified by 3q-capture DNA-seq (Table 1). In two cases (#TG-03 and #SO-45) a cryptic inv(3)/t(3;3) *GATA2/MECOM* rearrangements was found. In 7 other cases, previously reported recurrent 3q26 translocations were identified, i.e. t(2;3)(p21;q26) (N=3), t(3;7) (q26;q21) (N=2), t(3;8)(q26;q24) (N=1) and t(3;6)(q26;q25) (N=1). The genes thought to be involved in those translocations are *THADA*, *CDK6*, MYC, and *ARID1B* respectively[22-27] (Table 1). These abnormalities were most probably missed at diagnosis due to the complex genetic nature of these cases. In the other 11 atypical 3q26 AMLs, novel and unique MECOM/3q26 rearranged partner loci were found (Table 1). We hypothesize that regulatory elements of these genes were hijacked by *EVI1*, resulting in loss of expression of the gene at the rearranged allele. Combined DNA-seq/RNA-seq SNP analysis applied to these AMLs revealed monoallelic or skewed expression of some of these genes in the translocated locus. As an example in AML with ins(6;3)(q21;q21q26) (#HF-23, Figure 3A), t(3;4)(q26;p15) (#HF-19, Figure 3B) or a t(3;7)(q26;p22) (#SO-20, Figure 3C) skewed expression of *CD164*, *PROM1* (*CD133*) or *FSCN1/EIF2AK1* were found respectively (Figure 3D). Whether the repressed allele was rearranged could not be assessed due to lack of patient material. These genes are all expressed in CD34+ cells and myeloid progenitor cells[28], and both *CD164*[29-31] and *PROM1*[32] are known to play a prominent role in hematopoiesis.

MECOM hijacks myeloid-specific enhancers that may activate *EVI1* transcription
As chromatin of patient cells were not available, we studied the chromatin state at *CD164, PROM1* (*CD133*) and *FSCN1/EIF2AK1* in normal bone marrow CD34+ cells as well as in the inv(3) myeloid cell line MOLM-1[28,33]. As depicted in Figure 3A, 3B and 3C, binding of p300, presence of H3K27ac and lack of H3K4me3, were indicative of active enhancers within the regions that were translocated to *MECOM* in cases #HF-19, #HF-23 and #SO-20 respectively. In fact, the size of the H3K27 acetylated regions (>10kb) suggested the presence of a "super-enhancers"[34] in those loci (Figure 3E). Strong binding of key myeloid transcription factors like FLI1, GATA2 and RUNX1 (Figure 3A, 3B and 3C) in CD34+ bone marrow cells[28], further supports the notion that active myeloid "super-enhancers" translocate to *MECOM* in atypical 3q26 rearrangements to activate *EVI1* expression. ChIP-seq analysis of normal CD34+ and MOLM1 cells also showed the presence of "super-enhancers" in the regions near *THADA*, *MYC* and *CDK6*, that translocate to *MECOM* in AMLs with translocations t(2;3), t(3;8) and (t3;7) respectively (Table 1, Figure 3E, Figure S2A-E). The loss of these enhancers in one allele should lead to a reduction in total gene expression, but given that most of these translocations are unique to one patient, it is not possible to conduct a statistical analysis. Instead, for every gene that putatively loses its enhancer, we compared its average expression in the whole cohort to the expression in the individuals with the translocation. In line with our hypothesis, all genes except *MYC* exhibited reduced expression (Figure S3C). Together the data point to a mechanism of *EVI1* overexpression driven by hijacked myeloid "super-enhancers" in atypical 3q26 rearranged AML.

Atypical 3q26 AMLs exhibit *GATA2* deficiency in half of the cases
In inv(3)/t(3;3) AML, the dislocation of the *GATA2* enhancer causes loss of expression of *GATA2* from the rearranged allele[13,14]. We addressed the question whether *GATA2* expression was reduced in atypical 3q26 AML without 3q21/*GATA2* rearrangements. RNA-seq data demonstrated comparable *GATA2* expression levels for the atypical 3q26 AMLs as for the inv(3)/t(3;3) AMLs (Figure 4A), which was slightly lower than in non-3q26 rearranged AMLs, although not statistically significant. Analysis of SNP-array data (performed for 27 atypical 3q26 AMLs) revealed copy number loss of parts of chromosome 3 including *GATA2* and/or its enhancer in 5 atypical 3q26 AML patients (#TG-08, TG-10, HF-15, HF-16 and HF-21, Figure 4C). In 2 of these cases loss of one chromosome 3 was also noted cytogenetically (Table 1). CNV analysis of the 3q-captured data of all 33 cases was used to verify copy number changes detected by SNP-array: 5 cases with *GATA2* or *GATA2*-enhancer loss were identified (Table S1) of which two are shown in Figure S3B. In 16 AMLs of our cohort, we could discriminate between two *GATA2* alleles based on SNP differences, identified by combined RNA- and DNA-seq data analysis. In 4 of those 16 cases *GATA2* expression was monoallelic or significantly skewed to one allele (p<0.05, marked by * in Figure 4B). As methylation of the *GATA2* promoters could explain allele specific expression, bisulfite-sequencing experiments

were performed. However, we did not obtain any evidence for *GATA2* promoter methylation in these patients. Thus, the mechanism by which these cases showed unbalanced allelic *GATA2* expression remains unclear. Overall, we observed *GATA2* loss or skewed expression in 12 of the 22 (>50%) cases that we could analyze in full. No mutations in *GATA2* were found in any of the 33 atypical 3q26 AMLs. We conclude that in a subset of atypical 3q26 rearranged AML *EVI1* overexpression was accompanied by loss or diminished *GATA2* transcription from one allele, which resembles inv(3)/t(3;3) AML[13].

Table 1: Cytogenetic and *MECOM* associated alterations in atypical 3q26 AML

PT#	Karyotype Chr.3[1,2]	FISH *EVI1*[3]	SNP Chr.3[4]	*EVI1*[5]	*MDS1-EVI1*[5]	Breakpoint	Gene partner[6]
SO-03	add(3)(q26)	rearranged	Chr.3q26 balanced	+	-	breakpoint not found	
SO-06	?der(3)(q27)	rearranged	Chr.3q26 CNL 5' MECOM	+	-	inv(3;3)(p23q26), complex	TGFBR2
SO-11	der(3)add(3)(p17?)add(3)(q27?6)	rearranged	Chr.3q26 CNL 3' and 5' MECOM	+	-	t(3;7)(q26;q11.23/q21.12), complex	DMTF1
SO-20	add(3)(q26)	rearranged	Chr.3q26 balanced	+	-	t(3;7)(q26;p22.2), complex	TNRC18/FBXL18
SO-23	add(3)(q27?5)	rearranged	Chr.3q26 balanced	+	+	t(3;6)(q26;q25)	ARID1B
SO-45	del(3)(q27?3q27?6)	rearranged	Chr.3q26 balanced	+	-	t(3;3)(q21;q26)+t(3;16)(q26;q22.1), complex	GATA2
SO-47	add(3)(q26)	rearranged	Chr.3q26 balanced	+	-	t(2;3)(q21;q26)	THADA
BB-01	no 3q aberrations	rearranged	Chr.3q26 CNL MDS1	+	+	breakpoint not found	
TG-01	t(3;11)(q26;q27?4)	rearranged	not done	+	+	t(3;11)(q26;q24)	HSPA8-MECOM[7]
TG-02	t(3;18)(q26;q17?)	rearranged	not done	+	-	t(3;18)(q26;q21)	MECOM-TCF4[7]
TG-03	no 3q aberrations	rearranged	Chr.3q26 balanced	+	-	inv(3)(q21q26)	GATA2
TG-04	ins(3;3)(q26;q21q26)	unclear	Chr.3q26 balanced	+	+	breakpoint not found	IL12A-AS1
TG-05	?add(3)(q25)	Loss	Chr.3q26 CNL MDS, CNG EVI1	+	-	del(3)(q25.3-q26.2)	
TG-06		normal	Chr.3q26 balanced	+	-	breakpoint not found	
TG-08	add(3)(q26)	Loss	Chr.3q21 CNL GATA2	+	+	breakpoint not found	
TG-10	-3	rearranged	Chr.3q21 CNL GATA2	+	-	t(3;6)(q26;p22)	TDP2/JARID2
HF-01	der(7)t(3;7)(q26;q11.2)	rearranged	not done	+	-	t(3;7)(q26;q21)	CDK6
HF-02	der(7)t(3;7)(q26;q22)	rearranged	not done	+	-	breakpoint not found	CDK6
HF-03	der(7)t(3;7)(q26;q21)	rearranged	not done	+	-	t(3;7)(q26;q21)	
HF-04	der(7)t(3;7)(q26;p11)t(3;7)(q26;q21), -3	rearranged	not done	+	+	breakpoint not found	
HF-13	der(3)t(3;14)(q21;q?)	Amplified	Chr.3q26 CNG, MECOM balanced	+	-	breakpoint not found	
HF-14	der(3)(::3p12->3q13::3q26->3q26::)	Amplified	Chr.3q26 CNG MECOM	+	-	breakpoint not found	TRA2B-MECOM[7]
HF-15	r(3)(p11q26)del(3)(q14q26)	Amplified	Chr.3q26 CNG EVI1/CNL MDS1, Chr.3q21 CNL GATA2	+	-	inv(3)(q13.33q26.2)	GTF2E1/STXBP5L
HF-16	der(2)ins(2;3)(q31;q22q26)	Amplified	Chr.q26 CNG 5' and 3' EVI1/CNL MDS1, Chr.3q21 CNL GATA2	+	-	breakpoint not found	
HF-17	t(5;8)(p13;p21)	rearranged	Chr.3q26 balanced	+	-	t(3;8)(q26;q24.1)	MYC
HF-18	t(2;3;6)(p15;q26;q26)	rearranged	Chr.3q26 balanced	+	-	t(2;3)(p21;q26)+t(3;5)(q26;q34)+t(3;6)(q26;q27)	THADA
HF-19	t(3;4)(q26;p15)	rearranged	Chr.3q26 balanced	+	-	t(3;4)(q26;p15)	PROM1, CD38
HF-20	t(3;8)(q26;p23)	rearranged	Chr.3q26 CNL MDS1	+	-	t(3;8)(q26;p23)	TNKS/MSRA
HF-21	der(8)t(3;8)(q26;p23)	rearranged	Chr.3q26 CNL MDS1, Chr.3q21 CNL GATA2	-	-	t(3;8)(q26;p24), complex	FAM135B
HF-22	der(3)t(2;3)(p14;q26)	rearranged	Chr.3q26 balanced	+	-	t(2;3)(p21;q26)	THADA
HF-23	ins(6;3)(q21;q21q26)	rearranged	Chr.3q26 balanced	+	-	ins(6;3)(q21;q21q26)	CD164
HF-24	der(3)del(3)(p12p26)inv(3)(p26q26)	rearranged	Chr.3q26 balanced	-	-	breakpoint not found	
HF-25	t(3;10)(q26;q21)	rearranged	Chr.3q26 balanced	-	-	t(3;10)(q26;q21)	ARID5B

1. Cytogenetic aberrations with a specific focus on 3q26. Complete karyotype is provided in Table S1. **2.** Patient numbers BB-01, TG-03, TG-10 and HF-17 (#) did not show a 3q26 rearrangement by karyotyping, but were identified as rearranged by routine *MECOM FISH*. **3.** FISH was carried out as outlined in materials and methods section and scored as: normal, loss, amplified or rearranged. In sample TG-04 the FISH results were unclear. **4.** CNL: Copy Number Loss, CNG: Copy Number Gain. **5.** *EVI1+* and *MDS1-EVI1+* were determined as previously reported.[2-4] **6.** Partner gene: the gene(s), expressed in CD34+ cells, located in closest vicinity to the breakpoint is indicated. **7.** Fusion transcript.
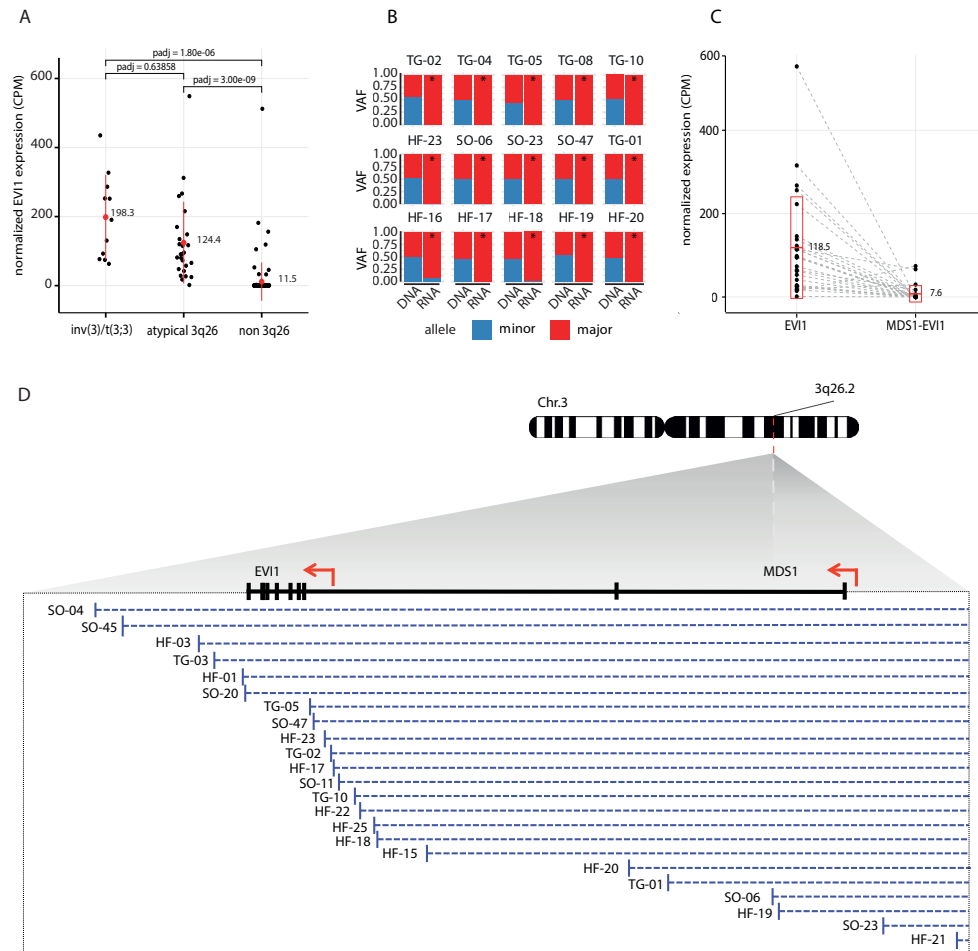
Figure 1. *MECOM* rearrangements, *EVI1* overexpression and absence of *MDS1-EVI1* expression in atypical 3q26 rearranged AML

(A) Normalized *EVI1* expression (counts per million (CPM) from RNA-seq data) determined in inv(3)/t(3;3) (N=11), atypical 3q26 (N=26) compared to non-3q26 AML (N=111). **(B)** Allele specific expression analysis using DNA-seq and RNA-seq data. The major allele is the allele of which the most SNPs were measured; the minor allele represents the allele that was underrepresented in the measurements. In order to perform this analysis, SNPs needed to be present in the sample. In 15/33 cases this analysis could be carried out. * indicates significant differential expression between alleles (p<0.05, χ² test). **(C)** Relative *EVI1* and *MDS1-EVI1* expression (CPM, RNA-seq) in atypical 3q26 AMLs (N=26). The red crossbar represents the mean and red box the standard deviation. **(D)** Schematic depiction of the breakpoints within the *MECOM* locus (3q26) determined by 3q-capture. The breakpoints could be determined in 23 AML cases. In 6 cases the breakpoint was 3'of *EVI1*, in 15 cases 5' of the *EVI1* promoter but 3' of the *MDS1-EVI1* promoter and in 2 AMLs 5'of the *MDS1-EVI1* promoter.
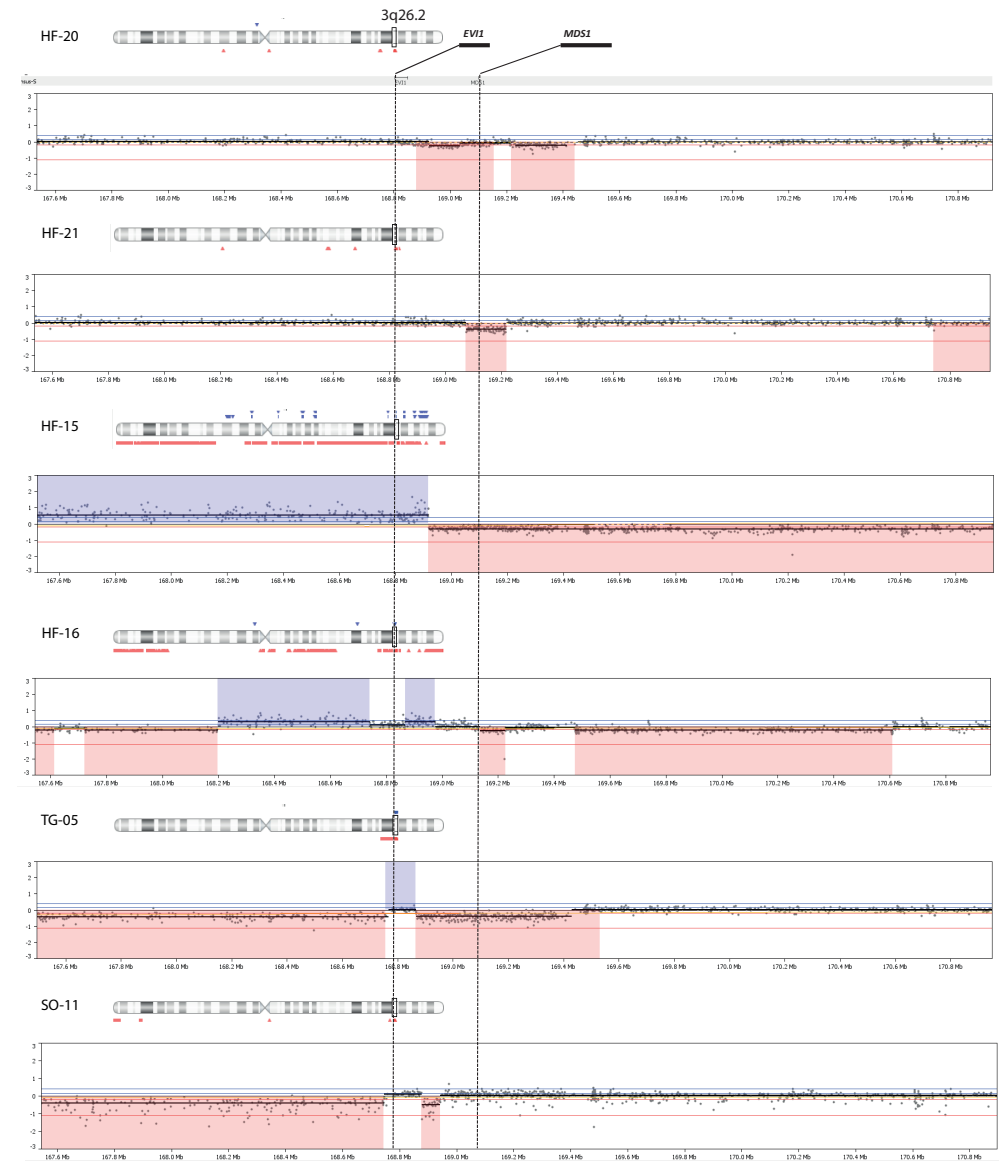
Figure 2. Copy number changes in the *MECOM* locus in atypical 3q26 AML

SNP array showing copy number losses (CNL) in red and copy number gains (CNG) in blue at chromosome band 3q26. *EVI1* and *MDS1-EVI1* are marked. Only the samples for which copy number changes were found in this locus are illustrated (N=6).
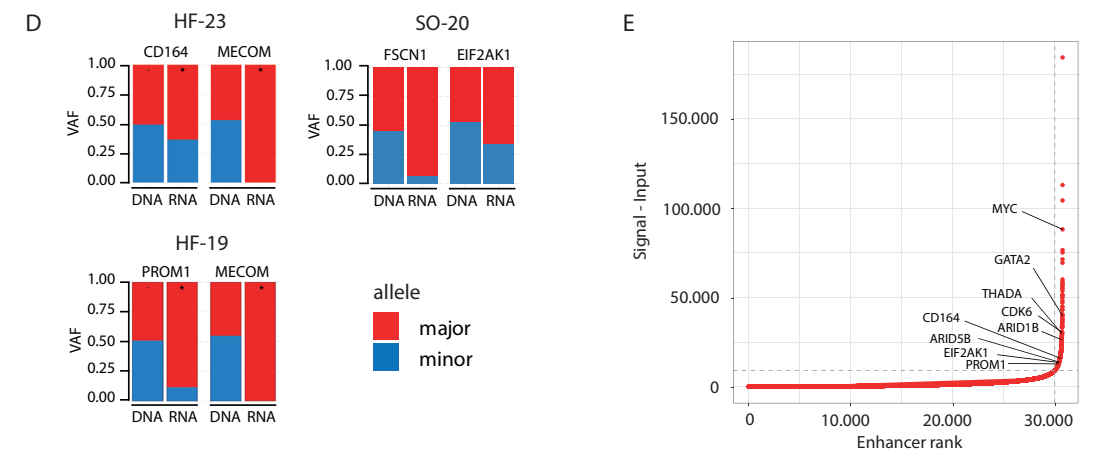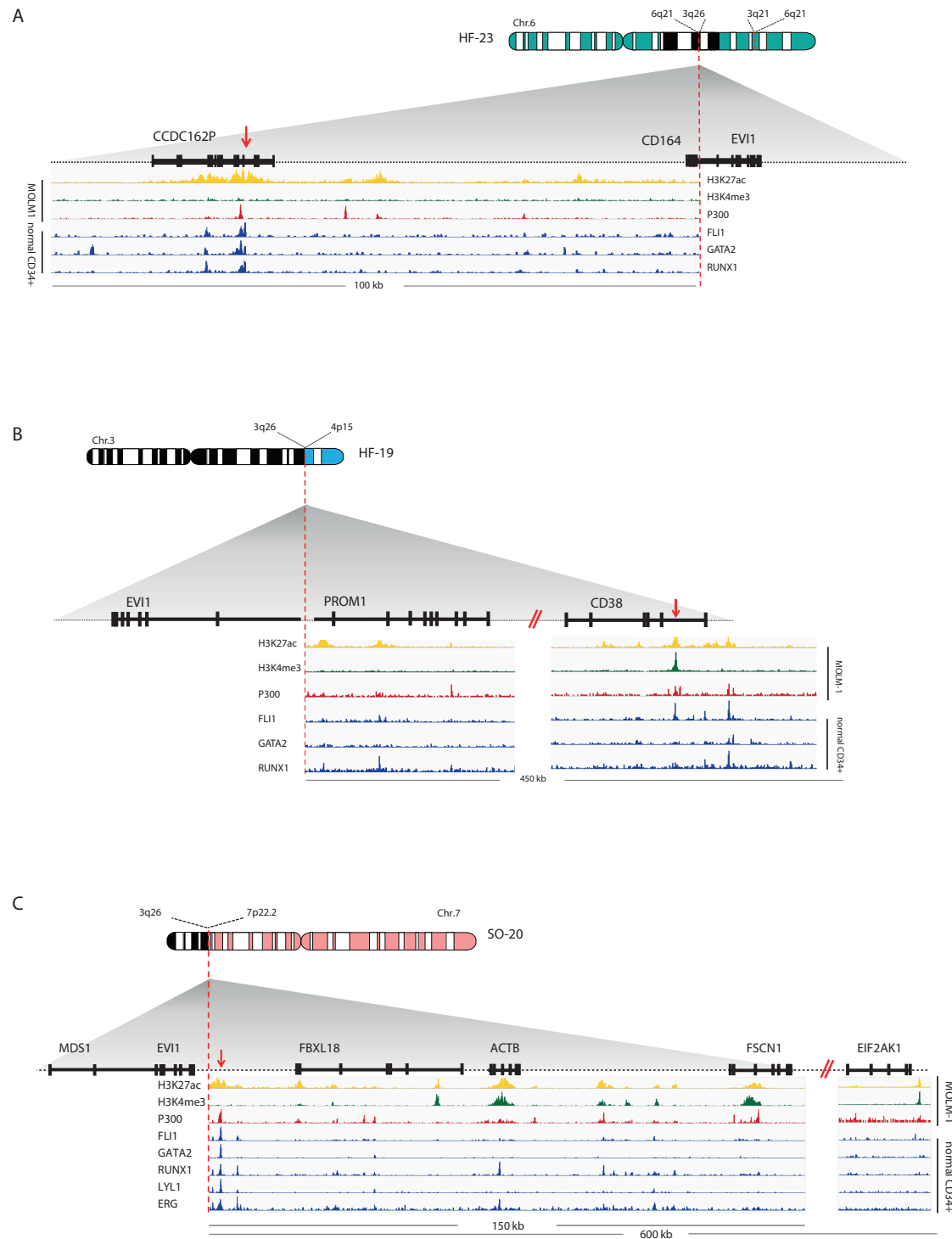
Figure 3. Rearrangements involving *3q26/EVI1* and newly identified partner loci

(**A**), (**B**) and (**C**) Schematic depictions of chromosomal rearrangements of three unique atypical 3q26 patient samples, i.e. ins(6;3)(q26;q21q26) in patient #HF-23, t(3;4)(q26;p15) in patient #HF-19 and t(3;7)(q26;p22) in patient #SO-20 respectively. Figures show the loci and genes that have been rearranged and brought into the vicinity of *MECOM*: loci with *CD164* and *CCDC162P* (6q21) in A, *PROM1* and *CD38* (4p15) in B and *FBXL18, ACTB, FSCN1* and *EIF2AK1* (7p22) in C respectively. ChIP-seq tracks indicative for active enhancer elements, i.e. H3K27ac (yellow), H3K4me3 absence (green) and P300 (red), have been obtained from the MOLM-1 myeloid cell line[1]. Previously published ChIP-seq tracks of myeloid transcription factors FLI1, GATA2, RUNX1, LYL1 and ERG using normal CD34+ cells are shown[2] (blue). Enhancers possibly involved in *EVI1* activation are indicated with a red arrow. (**D**) Bar plots showing skewed expression of genes that putatively donated their enhancer. The bar plots show the genes with skewed expression, *CD164* (#HF-23), *PROM1* (#HF-19), *FSCN1* and *EIF2AK1* (#SO-20). In 2 out of 3 samples, monoallelic *EVI1* expression was found (#HF-23, #HF-19). Allele specific *EVI1* expression could not be determined in for #SO-20, since no SNPs could be detected. * Indicates significant differential expression between alleles (p<0.05, χ² test). (E) Hockey stick plot showing the classification of these long stretches of H3K27ac (**A, B and C**) found in the partner loci as super-enhancers (based on MOLM-1 H3K27ac ChIP-seq data using the ROSE algorithm).
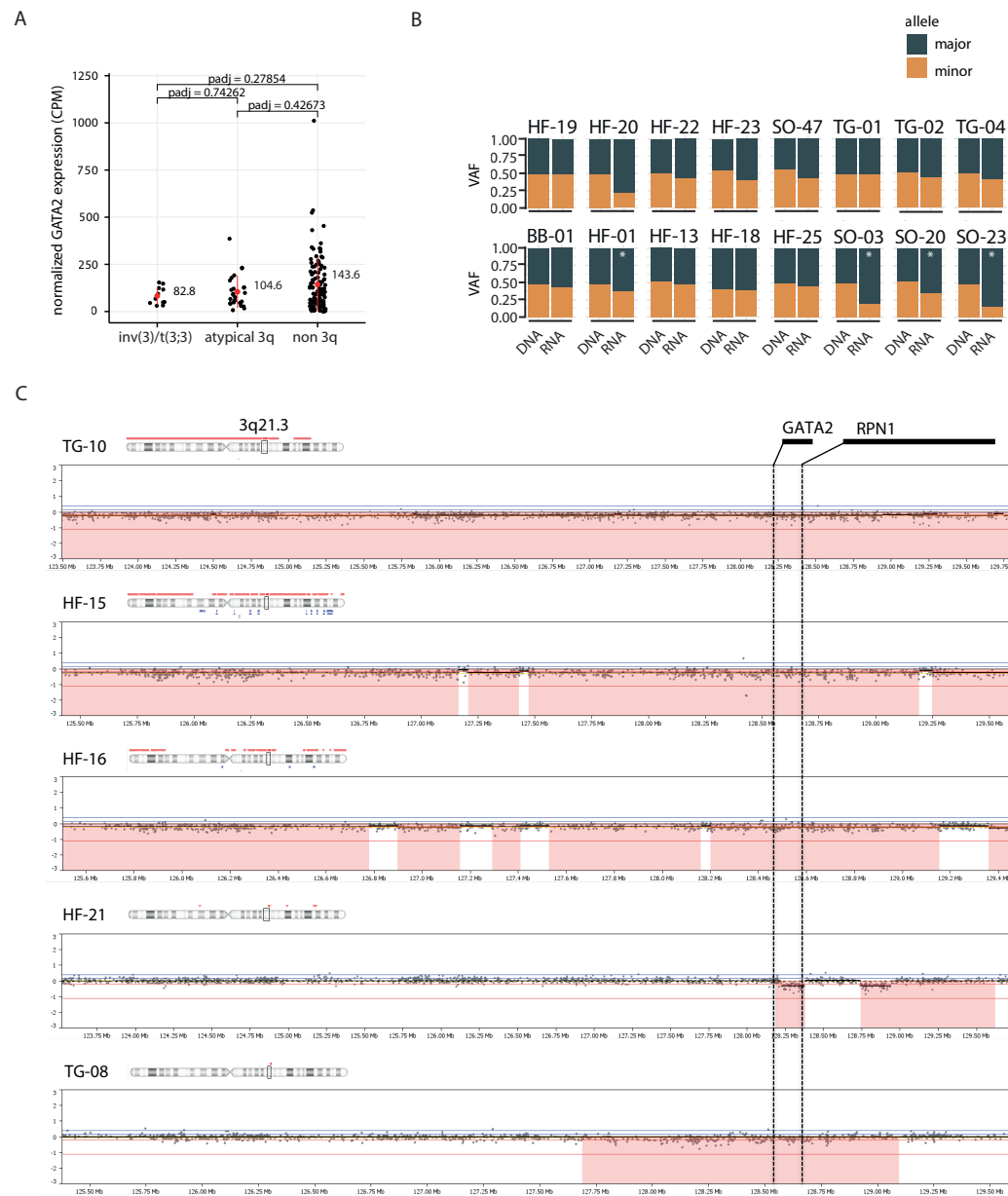
Figure 4. Copy number loss of *GATA2* or imbalanced *GATA2* expression in atypical 3q26 AML
(**A**) *GATA2* expression (CMP, RNA-seq) determined in inv(3)/t(3;3) (N=11), in atypical 3q26 (N=26) and non-3q26 rearranged AML (N=111). Differences were not statisticavly significant (padj<0.05). Red dot represents the mean and the red bar the standard deviation. (**B**) Allele specific analysis using DNA-seq and RNA-seq data showed significant skewed expression of *GATA2* to one allele in 5 cases. In #HF-20 read depth was too low for a significance call. * indicates significant differential expression between alleles (p<0.05, χ² test). (**C**) SNP array data presented at chromosomal locus 3q21.3 using, showing CNLs in the *GATA2* locus, resulting in loss (red) of the *GATA2* gene or its enhancer (located in between *GATA2* and *RPN1*).

DISCUSSION

Atypical 3q26-rearranged AML represents a group of very poor risk leukemias with various undefined 3q26 rearrangements whose role in leukemogenesis is unclear[9]. Using a multipronged approach, we here demonstrate that in atypical 3q26-rearranged AML, *MECOM* is relocated, leading to *EVI1* overexpression in the absence of *MDS1-EVI1* transcription. We found potential myeloid super-enhancers to be translocated to *MECOM*. In approximately 50% of the study cohort *GATA2* skewed expression or copy number loss was found, despite lack of *GATA2* involvement in the rearrangement. We conclude that atypical 3q26 AML genocopy inv(3)/t(3;3) leukemias[13,14] and these two groups should be classified and treated as single entity.

In atypical 3q26 AMLs, chromosomal rearrangements bring *MECOM* into the vicinity of regulatory elements of genes active in myeloid cells, such as *THADA, CDK6, MYC, ARID1B, CD164, PROM1 (CD133) or FSCN1/EIF2AK1*[22-27]. We hypothesize that a mechanism of super-enhancer hijacking causes *EVI1* overexpression in variant 3q26-AMLs, as has been reported for the -77 kb *GATA2* enhancer in inv(3)/t(3;3) leukemias. ChIP-seq data from normal CD34+ bone marrow cells and myeloid cell lines revealed that transcription factors (TFs) that bind to the *GATA2* distal enhancer, including RUNX1, LYL1, SCL, FLI1, ERG, LMO2, and GATA2 itself[28], also interact with the loci translocated in atypical 3q26 AMLs. It will be challenging to model these translocations and study *EVI1* promoter interaction and regulation by these distinct super-enhancers. As super-enhancers have been reported to be hypersensitive to bromodomain-inhibitors[35,36], it will be interesting to study responses of the distinct 3q26-rearranged AMLs to those compounds.

It is well established that *EVI1* is an oncogenic driver of AML, but the role of *MDS1-EVI1* in leukemic transformation has not been thoroughly studied. *Evi1* was first identified as the ecotropic viral insertion site-1 in mouse leukemias, in which *Evi1* but not *Mds1-Evi1* was overexpressed due to retroviral insertions[37]. Patients with X-linked chronic granulomatous disease who received gene therapy to correct *GP91* (*PHOX*) mutations in hematopoietic progenitor cells, similarly developed AML due to retroviral insertions driving *EVI1* and not *MDS1-EVI1* overexpression[38]. Here we demonstrate that in atypical 3q26 AML, as reported in AML with inv(3)/t(3;3), overexpression of *EVI1* was accompanied by absence or low expression of *MDS1-EVI1*. We hypothesize that the translocated enhancers in 3q26-rearranged AMLs are able to contact and co-activate the promoter of *EVI1*, but not the promoter of *MDS1-EVI1*.

Monoallelic expression of *GATA2* is another hallmark of inv(3)/t(3;3), caused by loss of the *GATA2* enhancer at the rearranged allele. Does monoallelic *GATA2* play a role in leukemic transformation in inv(3)/t(3;3)? In over 50% of the atypical 3q26 AMLs analyzed, skewed or monoallelic expression of *GATA2* was evident, due to cryptic *GATA2/MECOM* translocation, deletion of *GATA2* or a regulatory element or by currently unknown mechanisms. *EVI1*

overexpressing mice develop myeloid leukemias with a shorter latency when they are *GATA2* heterozygous[39]. Moreover, individuals with inherited *GATA2* mutations or loss of expression of one allele have a high chance to develop AML[15-19]. Altogether, loss of one *GATA2* allele appears to have an effect on leukemia development. A larger patient cohort is required to investigate whether *GATA2* monoallelic expression has an impact on prognosis of 3q26-rearranged AML.

Atypical 3q26 AMLs are difficult to define, as they are cytogenetically complex and heterogeneous. This underscores the importance of routine molecular diagnostic assays to recognize this subgroup of AML patients. We propose to identify 3q26/MECOM rearrangements by using *MECOM* FISH (Figure S1), which is applied routinely in AML diagnostics. Quantitative *EVI1* and *MDS1-EVI1* mRNA expression analysis can be indicative for *EVI1* deregulation by enhancer hijacking. Together, this combined analysis can be used to classify this subgroup of AML patients.
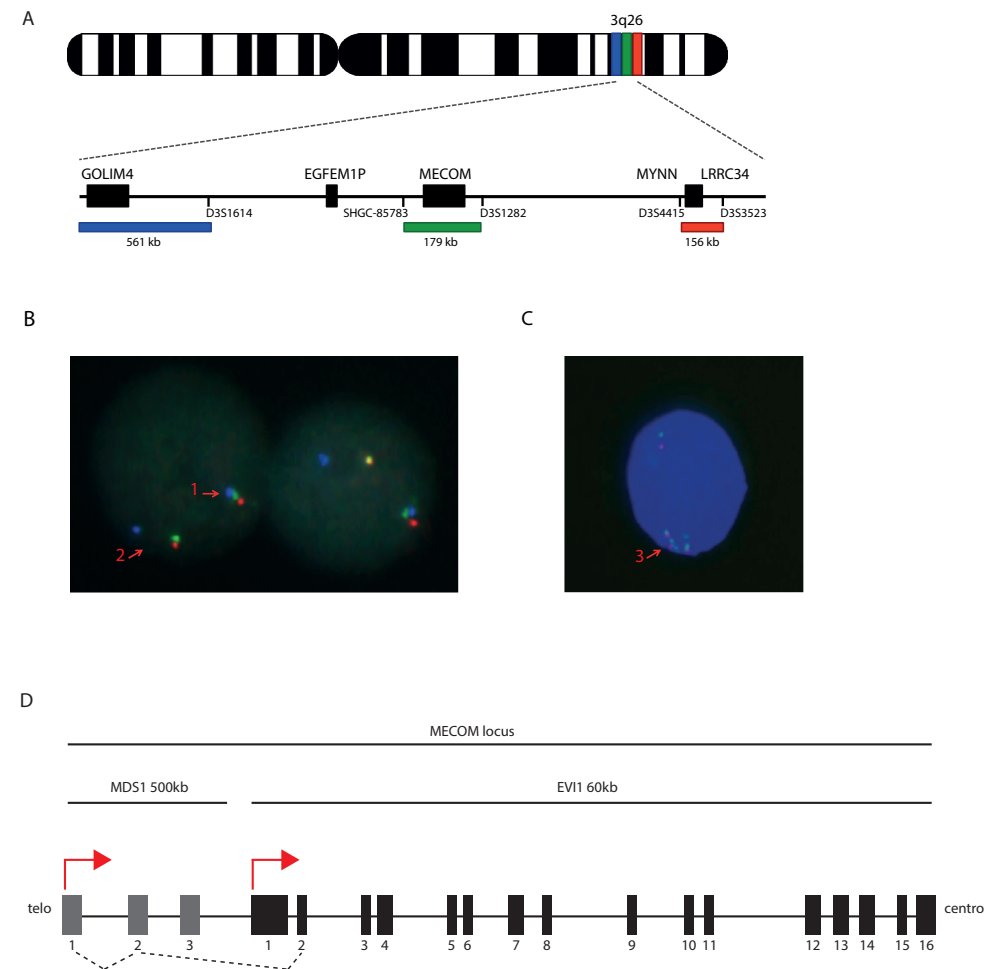

Figure S1. Schematic overview of the *MECOM* locus and *MECOM* FISH
(**A**) Schematic overview of *MECOM* breakapart FISH (Cytocell, figure modified from manufacturer website). (**B**) Example of *MECOM* FISH of AML cells with an inv(3)(q21q26). Arrow 1 indicates a normal allele with three probes. Separation of the blue probe from the green/red probes (arrows 2) recognizes the rearranged allele. (**C**) Example of a FISH experiment showing *MECOM* amplification (arrow 3). (**D**) Schematic overview of the *MECOM* locus, showing the exons of the long form *MDS1-EVI1* and the short form *EVI1* (adapted from Odera et al., 2017[3]).
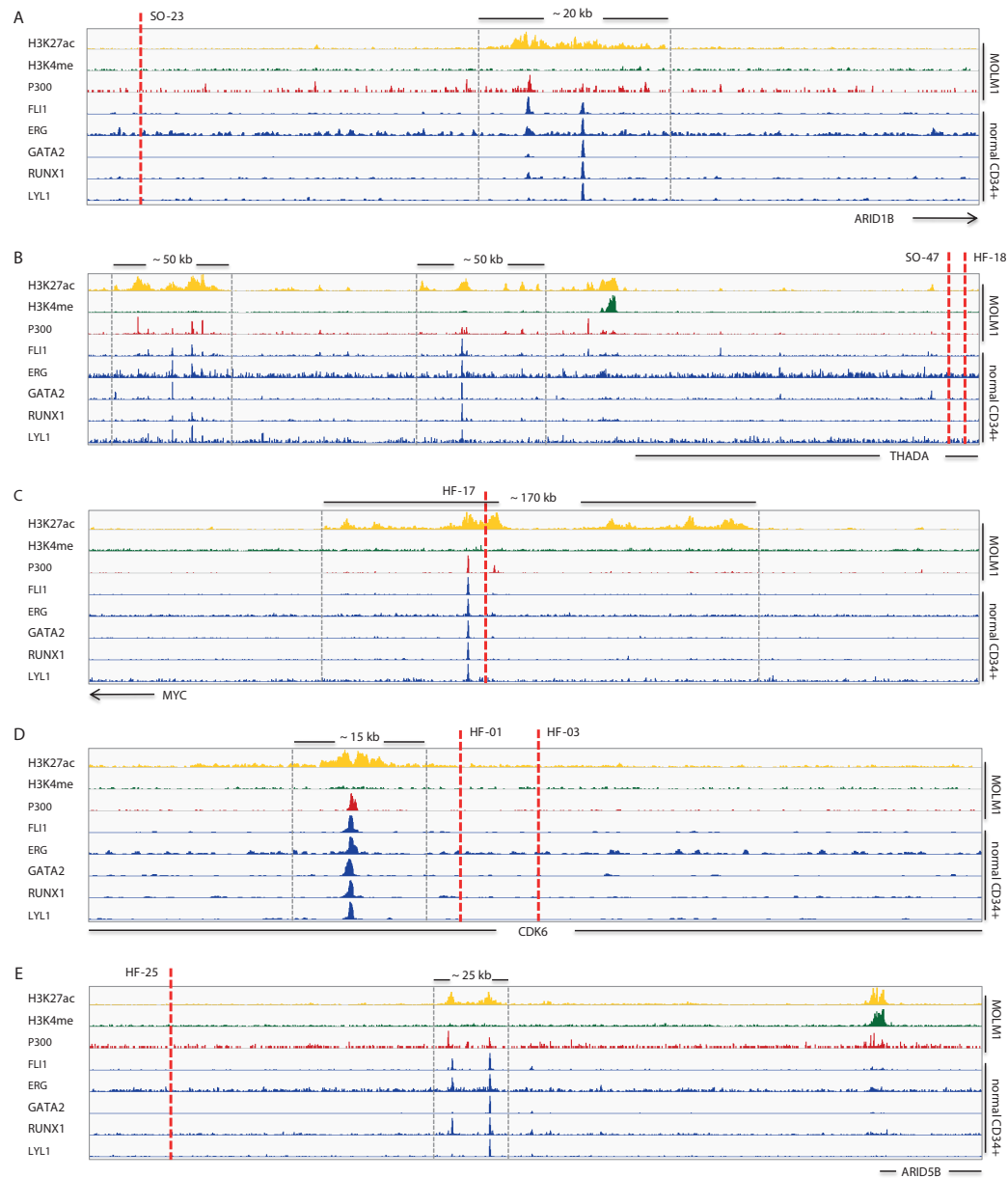
Figure S2. Super-enhancers translocated to *EVI1* in atypical 3q26

Potential super-enhancer regions (based on Figure 3G) in translocated loci, indicated by the presence of a large region of H3K27ac (yellow track), P300 (red track) and the absence of H3K4me1 (green track) (ChIP-seq data MOLM-1) and strong binding of early myeloid transcription factors, FLI1, ERG, GATA2, RUNX1 and LYL1 (blue track, ChIP-seq data CD34+ cells[2]). In each panel the translocated region that is brought in to close proximity of *EVI1* and the gene thought to be involved are depicted, the dashed red line indicates the chromosomal breakpoint. (**A**) #SO-23 t(3;6)(q26;q25), *ARID1B*. (**B**) #SO-47 and #HF-18, t(2;3)(p21;q26), *THADA*. (**C**) #HF-17, t(3;8)(q26;q24), *MYC*. (**D**) #HF-01, #HF-03, t(3;7)(q26;q11), *CDK6*. (**E**) #HF-25, t(3;10)(q26;q21), *ARID5B*.
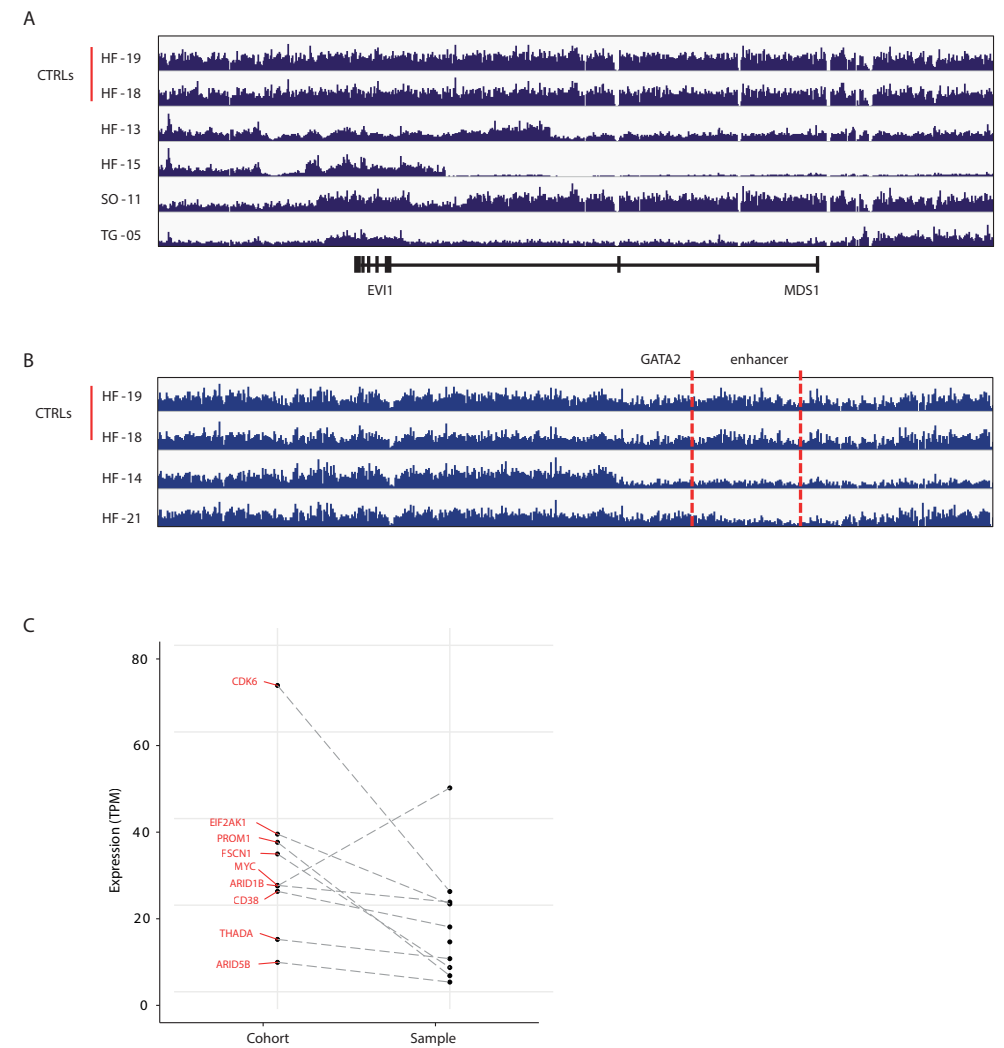

Figure S3. Copy number variants in *MECOM/GATA2* locus and decreased expression of genes in translocated region

IGV screenshots of 3q-capture tracks of atypical 3q26 cases with clear gains or losses in either the *MECOM* locus (**A**) or *GATA2* locus (**B**). The first two tracks of each panel show cases with a balanced 3q21 and 3q26 region (CTRLs). (**A**) In #HF-13 and #HF-15 clear gains of *EVI1* are observed but not in *MDS1*. In #SO-11 losses 3'and 5'of *EVI1* are found, but *EVI1* remains balanced. #TG-05, losses of the *MDS1* exons and the region 3' of *EVI1* are observed. Again, exactly *EVI1* remains intact. (**B**) In #HF-14 a loss of the *GATA2* gene and the region where the enhancer is found is lost. In patient #HF-21 the *GATA2* gene itself is balanced, but the enhancer is lost. Data of the cases shown here in S3A and S3B verifies what is observed in the SNP-array data, shown in figure 2 and 4C respectively. (**C**) Comparison of expression levels of the genes present in translocated regions. For each gene involved in a translocation, the first column shows its average expression in the whole cohort, whereas the second column shows its average expression in samples that carry that translocation.

METHODS

Patient material

Samples of the selected patients presenting with MDS or AML were collected either from the Erasmus MC Hematology department biobank (Rotterdam, The Netherlands) or from the MLL Munich Leukemia Laboratory biobank (Munich, Germany). Leukemic blast cells were purified from bone marrow or blood by standard diagnostic procedures. All patients provided written informed consent in accordance with the Declaration of Helsinki.

Cytogenetics: karyotype and FISH

Diagnostic cytogenetics for all samples was performed by each of the institutes mentioned above. For this study, samples were selected based on 3q26 rearrangements (other than recurrent or classic 3q26 rearrangements) detected by karyotyping or *MECOM* interphase fluorescence *in situ* hybridization (FISH). FISH and classic metaphase karyotyping were performed and reported according to standard protocols based on the International System of Human Cytogenetics Nomenclature (ISCN) 2017[40]. *MECOM* FISH was performed according to the manufacturer's protocol, using the *MECOM* t(3;3); inv(3)(3q26) triple color probe (Cytocell, LPH-036).

RNA isolation and qPCR

RNA was isolated using phenol-chloroform extraction followed by DNase digestion or using the Qiagen Allprep DNA/RNA kit and protocol (Qiagen, #80204). cDNA synthesis was done using the SuperScript II Reverse Transcriptase kit (Invitrogen). Quantitative real-time PCR was performed by using primers as described previously[13,41] on the 7500 Fast Real-time PCR System (Applied Biosystems). Relative levels of gene expression were calculated using the $\Delta\Delta Ct$ method[7,8,42].

SNP-Array

Patient blasts were stored at -80°C in RLT+ buffer (Qiagen) and DNA was isolated using the AllPrep DNA/RNA mini kit (Qiagen, #80204). All SNP-arrays were performed at the Erasmus MC Department of Clinical Genetics (Rotterdam, The Netherlands) as previously described [43,44]. In summary, per sample, 50-200 ng DNA was used for a single Illumina Global Screening Array (GSAMD)(San Diego, CA, USA). The array profiles were analyzed with a 0.15 Mb resolution in UCSC (Human Mar. 2006 (NCBI36/hg18) Assembly) by using Genome Studio (Illumina) and different versions of Nexus Copy Number Software (BioDiscovery: versions 5.0 and higher (Hawthorne, CA, USA)).

Targeted chromosomal region 3q21.1-3q26.2 DNA sequencing (3q-capture)

DNA was isolated as mentioned above. 3q-capture DNA sequencing was performed as we described previously [13]. In summary, genomic DNA was fragmented using the Covaris shearing device (Covaris), and sample libraries were assembled following the TruSeq DNA Sample Preparation Guide (Illumina). After ligation of adapters and an amplification step, target sequences of chromosomal regions 3q21.1-q26.2 were captured using custom in-solution oligonucleotide baits (Nimblegen SeqCap EZ Choice XL). The design of target sequences was based on the human genome assembly hg19: chr3q21.1:126036241-130672290 - chr3q26.2:157712147-175694147. Amplified captured sample libraries were paired-end sequenced (2x100 bp) on the HiSeq 2500 platform (Illumina) and aligned against the hg19 reference genome using the Burrows-Wheeler Aligner (BWA)[45]. Chromosomal breakpoints were determined using Breakdancer [46]. All chromosomal aberrations found using this program were visually confirmed in the Integrated Genome Viewer (IGV)[47].

RNA sequencing

Sample libraries were prepped using 500 ng of input RNA according to the KAPA RNA HyperPrep Kit with RiboErase (HMR) (Roche) using Unique Dual Index adapters (Integrated DNA Technologies, Inc.). Amplified sample libraries were paired-end sequenced (2x100 bp) on the Novaseq 6000 platform (Illumina) and aligned against the human genome (hg19) using STAR version 2.5.4b. A description of the quantification and differential expression analysis is provided in the Supplementary Material.

Exome sequencing

DNA was isolated as described above. The Genomic DNA Clean & Concentrator kit (ZYMO Research) was used to remove EDTA from the DNA samples. Sample libraries were prepared using 100 ng of input according to the KAPA HyperPlus Kit (Roche) using Unique Dual Index adapters (Integrated DNA Technologies, Inc.). Exomes were captured using the SeqCap EZ MedExome (Roche Nimblegen) according to SeqCap EZ HyperCap Library v1.0 Guide (Roche) with the xGen Universal blockers – TS Mix (Integrated DNA Technologies, Inc.). The amplified captured sample libraries were paired-end sequenced (2x100 bp) on the Novaseq 6000 platform (Illumina) and aligned to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA)[45]. A description of the variant calling and allele expression analysis is provided in the Supplementary Material.

Whole-genome sequencing

DNA isolation, whole genome library preparation and sequencing was performed at the Munich Leukemia Laboratory (MLL, Munich, Germany). Sequencing was performed on the Novaseq 6000 platform (Illumina). The experimental procedures are detailed in a previous report by the MLL laboratory[48]. WGS data were aligned to the hg19 reference genome using the Burrows-Wheeler Aligner (BWA)[45].

SUPPLEMENTARY DATA

Differential expression analysis

Salmon[49] was used to quantify expression of individual transcripts, which were subsequently aggregated to estimate gene-level abundances with tximport[50]. Human gene annotation derived from RefSeq[51] was downloaded from UCSC[52] (RefGene) as a GTF file. Both gene- and transcript-level abundances were normalized to counts per million (CPM) for visualization in the figures of this paper. Differential gene expression analysis of count estimates from Salmon was performed with DEseq2[53]. As control, in house RNA-seq data of a cohort representative of the genetic diversity of AML cases was used (referred to as non-3q26 AML).

Allele-specific expression

To discriminate expression from different alleles, single nucleotide variants (SNVs) were first detected at the DNA level, using either whole-genome or exome sequencing data if available, or 3q-capture sequencing data otherwise. This step was performed with a custom script that integrated variants called by multiple software tools, including HaplotypeCaller and MuTecT2 from GATK[54], VarScan2[55] and bcftools[18]. The combined list of SNVs was subjected to stringent filtering to remove low-quality positions, considering the following criteria: a) strand bias, b) sequencing depth, c) alignment and base calling score, d) mappability. A highly optimized in-house tool (*annotateBamStatistics*) was then used to compute DNA and RNA allele-specific read counts at every SNV position from their respective alignment (BAM) files. For every gene, counts from all SNVs were summed to create a 2x2 contingency table (variables MAJOR/MINOR and DNA/RNA) and a χ^2 test of independence was conducted. Finally, skewed expression was determined for genes with False Discovery Rate (FDR) < 0.05 and RNA minor allele frequency < 0.35. The results were validated by visual examination of the DNA-seq and RNA-seq BAM files in IGV[47].

Copy number variant (CNV) analysis in 3q-capture data

CNV analysis was performed with CNVkit[56] in two steps. First, a pooled reference was generated based on all the 3q-capture datasets, which averaged out possible differences between them. As suggested by the instructions of the program, 5 kb regions of poor mappability were excluded from the analysis. Subsequently, the reference was employed to compute log2 copy ratios and infer discrete copy number segments using the default settings of CNVkit. Finally, we derived absolute integer copy numbers of these segments with the function "cnvkit call". Regions with a copy number other than 2 in the vicinity of *GATA2* or *MECOM* were subjected to further scrutiny in the BAM file of the corresponding sample: depth and variant allele frequency (VAF) were visually checked using IGV to confirm the CNV reported by CNVkit.

ChIP-seq data and analysis

H3K27ac ChIP-seq data from the inv(3) cell line MOLM-1 was previously generated in our group and is publicly available[13]. Briefly, reads were aligned to the human reference genome build hg19 with BBMap[57] and bigwig files were generated for visualization with bedtools genomecov[58] and UCSC bedGraphToBigWig[59]. Putative super-enhancers were identified on the basis of the ranked H3K27ac signal with ROSE[60], with 5kb stitching distance and excluding peaks in promoter regions. Transcription factor binding profiles (ChIP-seq) in human CD34+ cells were retrieved from the BloodChip database[28] in bigwig format. These tracks were visualized using IGV combined with Molm-1 derived H3K27ac signal to infer the presence of myeloid driven putative super-enhancers.

Bisulfite sequencing

To investigate if skewed *GATA2* expression was due to methylation of the promoter at one allele, bisulfite-sequencing experiments were performed like previously described[61]. Three regions in the *GATA2* locus were incorporated in the experiments. Based on RNA-seq data we saw that two main isoforms of *GATA2* were expressed in this patient cohort: a long and a short transcript, the latter expressed the highest. For both forms, sequences in the promoter regions were analyzed. In addition, we sequenced a region upstream of *GATA2* marked by H3K4me3 in MOLM-1 cell line. Chromosomal coordinates and primers are indicated below.

'H3K4me3 region' **Chr.3:** 128497666-128497881	**Fw:** AGCCTCTGCAGCTGGGACAAGGATGT **Rv:** GGGATTAGCTCATCTCCAGGCAGGT
'Long form *GATA2*' **Chr.3:** 128492783-128492961	**Fw:** GAGCCCCAAAGGTAGGGGCCACAGGG **Rv:** GCCTGGAGTAGAGCTGGGAGCAGG
'Short form *GATA2*' **Chr.3:** 128487826-128488155	**Fw:** GGGTAGGAGCTGGGGGTAGA **Rv:** CACCACTAAGGGACCCTCACCCCAAGG

R. Mulet-Lazaro, S. Ottema, T. Grob), Skyline DX (S. Ottema) and the Daniel den Hoed Foundation (L. Smeenk).

Authorship Contributions

S.O., R.M., R.D. designed the study. S.O., C.E., S.H., R.H., M.H., T.G. E.B. L.S. carried out experiments. R.M., S.O., H.B.B., L.S., C.H., T.H., P.J.M.V. analyzed data. H.B.B., P.J.M.V., T.H., C.H. provided samples and/or data. R.D., R.M., S.O. wrote the manuscript.

Disclosure of Conflicts of Interest

T.H. and C.H.: MLL Munich Leukemia Laboratory: Employment, Equity Ownership. The other authors have nothing to disclose.

REFERENCES

1. Döhner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453-474.

2. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine*. 2016;374(23):2209-2221.

3. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine*. 2013;368(22):2059-2074.

4. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.

5. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.

6. Morishita K, Parganas E, William CL, et al. Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300-400 kilobases on chromosome band 3q26. *Proceedings of the National Academy of Sciences*. 1992;89(9):3937-3941.

7. Lugthart S, van Drunen E, van Norden Y, et al. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. *Blood*. 2008;111(8):4329-4337.

8. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, van Putten WLJ, et al. High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. *Blood*. 2003;101(3):837-845.

9. Lugthart S, Gröschel S, Beverloo HB, et al. Clinical, Molecular, and Prognostic Significance of WHO Type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and Various Other 3q Abnormalities in Acute Myeloid Leukemia. *Journal of Clinical Oncology*. 2010;28(24):3890-3898.

10. Mitelman F, Johansson B, Mertens F. Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nature Genetics*. 2004;36(4):331-334.

11. Mitelman F JBaMF. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2019). Nat. Genet., 36 (2004), pp. 331-334: Nature Genetics; 2019.

12. Fröhling S, Döhner H. Chromosomal Abnormalities in Cancer. *New England Journal of Medicine*. 2008;359(7):722-734.

13. Groschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014;157(2):369-381.

14. Yamazaki H, Suzuki M, Otsuki A, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell*. 2014;25(4):415-427.

15. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood*. 2011;118(10):2653-2655.

16. Hsu AP, Johnson KD, Falcone EL, et al. GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood*. 2013;121(19):3830-3837.

17. Hahn CN, Chong C-E, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nature Genetics*. 2011;43:1012.

18. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nature Genetics*. 2011;43:929.

19. Rodrigues NP, Janzen V, Forkert R, et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood*. 2005;106(2):477-484.

20. Gerhardt TM, Schmahl GE, Flotho C, Rath AV, Niemeyer CM. Expression of the Evi-1 gene in haemopoietic cells of children with juvenile myelomonocytic leukaemia and normal donors. *British Journal of Haematology*. 1997;99(4):882-887.

21. Privitera E, Longoni D, Brambillasca F, Biondi A. EVI-1 gene expression in myeloid clonogenic cells from juvenile myelomonocytic leukemia (JMML). *Leukemia*. 1997;11(12):2045-2048.

22. Nucifora G, Laricchia-Robbio L, Senyuk V. EVI1 and hematopoietic disorders: History and perspectives. *Gene*. 2006;368:1-11.

23. Lin P, Medeiros LJ, Yin CC, Abruzzo LV. Translocation (3;8)(q26;q24): a recurrent chromosomal abnormality in myelodysplastic syndrome and acute myeloid leukemia. *Cancer Genetics and Cytogenetics*. 2006;166(1):82-85.

24. Lennon PA, Abruzzo LV, Medeiros LJ, et al. Aberrant EVI1 expression in acute myeloid leukemias associated with the t(3;8)(q26;q24). *Cancer Genetics and Cytogenetics*. 2007;177(1):37-42.

25. De Braekeleer M, Guéganic N, Tous C, et al. Breakpoint heterogeneity in (2;3)(p15–23;q26) translocations involving EVI1 in myeloid hemopathies. *Blood Cells, Molecules, and Diseases*. 2015;54(2):160-163.

26. Trubia M, Albano F, Cavazzini F, et al. Characterization of a recurrent translocation t(2;3)(p15–22;q26) occurring in acute myeloid leukaemia. *Leukemia*. 2006;20(1):48-54.

27. Storlazzi CT, Anelli L, Albano F, et al. A novel chromosomal translocation t(3;7)(q26;q21) in myeloid leukemia resulting in overexpression of EVI1. *Annals of Hematology*. 2004;83(2):78-83.

28. Chacon D, Beck D, Perera D, Wong JWH, Pimanda JE. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Research*. 2013;42(D1):D172-D177.

29. Watt SM, Chan JYH. CD164-A Novel Sialomucin on CD34+ Cells. *Leukemia & Lymphoma*. 2000;37(1-2):1-25.

30. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications*. 2019;10(1):2395.

31. Zannettino ACW, Bühring H-J, Niutta S, Watt SM, Benton MA, Simmons PJ. The Sialomucin CD164 (MGC-24v) Is an Adhesive Glycoprotein Expressed by Human Hematopoietic Progenitors and Bone Marrow Stromal Cells That Serves as a Potent Negative Regulator of Hematopoiesis. *Blood*. 1998;92(8):2613-2628.

32. Yin AH, Miraglia S, Zanjani ED, et al. AC133, a Novel Marker for Human Hematopoietic Stem and Progenitor Cells. *Blood*. 1997;90(12):5002-5012.

33. Matsuo Y1 AT, Tsubota T, Imanishi J, Minowada J. Establishment and characterization of a novel megakaryoblastic cell line, MOLM-1, from a patient with chronic myelogenous leukemia. *Human Cell*. 1991.

34. Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2015;47(1):8-12.

35. Lovén J, Hoke Heather A, Lin Charles Y, et al. Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell*. 2013;153(2):320-334.

36. Whyte Warren A, Orlando David A, Hnisz D, et al. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*. 2013;153(2):307-319.

37. Mucenski ML, Taylor BA, Ihle JN, et al. Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Molecular and cellular biology*. 1988;8(1):301-308.

38. Ott MG, Schmidt M, Schwarzwaelder K, et al. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nature Medicine*. 2006;12(4):401-409.

39. Katayama S, Suzuki M, Yamaoka A, et al. GATA2 haploinsufficiency accelerates EVI1-driven leukemogenesis. *Blood*. 2017;130(7):908-919.

40. International Standing Committee on Human Cytogenomic N, McGowan-Jordan J, Simons A, Schmid M. ISCN : an international system for human cytogenomic nomenclature (2016); 2016.

41. Gröschel S, Lugthart S, Richard FS, et al. High EVI1 Expression Predicts Outcome in Younger Adult Patients With Acute Myeloid Leukemia and Is Associated With Distinct Cytogenetic Abnormalities. *Journal of Clinical Oncology*. 2010;28(12):2101-2107.

42. Valk PJM, Verhaak RGW, Beijen MA, et al. Prognostically Useful Gene-Expression Profiles in Acute Myeloid Leukemia. *New England Journal of Medicine*. 2004;350(16):1617-1628.

43. Srebniak M, Boter M, Oudesluijs G, et al. Application of SNP array for rapid prenatal diagnosis: implementation, genetic counselling and diagnostic flow. *European Journal Of Human Genetics*. 2011;19:1230.

44. Srebniak MI, Diderich KEM, Joosten M, et al. Prenatal SNP array testing in 1000 fetuses with ultrasound anomalies: causative, unexpected and susceptibility CNVs. *European Journal Of Human Genetics*. 2015;24:645.

45. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

46. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*. 2009;6:677.

47. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature Biotechnology*. 2011;29:24.

48. Mack EKM, Marquardt A, Langer D, et al. Comprehensive genetic diagnosis of acute myeloid leukemia by next-generation sequencing. *Haematologica*. 2019;104(2):277-287.

49. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.

50. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521.

51. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-745.

52. Karolchik D, Hinrichs AS, Furey TS, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 2004;32(Database issue):D493-496.

53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.

54. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.

55. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576.

56. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.

57. Bushnell B. BBMap short-read aligner, and other bioinformatics tools; 2016.

58. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.

59. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010;26(17):2204-2207.

60. Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307-319.

61. Wouters BJ, Jordà MA, Keeshan K, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood*. 2007;110(10):3706-3714.

SUPPLEMENTARY REFERENCES

1. Groschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014;157(2):369-381.

2. Chacon D, Beck D, Perera D, Wong JWH, Pimanda JE. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Research*. 2013;42(D1):D172-D177.

3. Maicas M, Vázquez I, Alis R, et al. The MDS and EVI1 complex locus (MECOM) isoforms regulate their own transcription and have different roles in the transformation of hematopoietic stem and progenitor cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. 2017;1860(6):721-729.

3

Selective requirement of MYB for oncogenic hyperactivation of a translocated enhancer in leukemia

Leonie Smeenk[1,2], Sophie Ottema[1,2], Roger Mulet-Lazaro[1,2], Anja Ebert[5], Marije Havermans[1,2], Andrea Arricibita Varea[1,2], Michaela Fellner[5], Dorien Pastoors[1,2], Stanley van Herk[1,2], Claudia Erpelinck[1,2], Tim Grob[1], Remco Hoogenboezem[1], François G. Kavelaars[1], Daniel R. Matson[3], Emery H. Bresnick[3], Eric Bindels[1], Alex Kentsis[4], Johannes Zuber[5,6,7] and Ruud Delwel[1,2,7,*]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands, [2] Oncode Institute, Erasmus University Medical Center, Rotterdam, the Netherlands, [3] Department of Cell and Regenerative Biology, Madison, WI, USA, [4] Tow Center for Developmental Oncology, Sloan Kettering Institute, Department of Pediatrics, Weill Medical College of Cornell University, Memorial Sloan Kettering Cancer Center, New York, USA, [5] Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria, [6] Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria, [7] These authors contributed equally

*Correspondence: h.delwel@erasmusmc.nl

Keywords

AML, EVI1, MYB, hijacked enhancer, CRISPR/Cas9

Statement of significance

We show a novel paradigm in which chromosomal aberrations reveal critical regulatory elements that are non-functional at their endogenous locus. This knowledge provides a rationale to develop new compounds to selectively interfere with oncogenic enhancer activity

ABSTRACT

In acute myeloid leukemia (AML) with inv(3)(q21;q26) or t(3;3)(q21;q26), a translocated *GATA2* enhancer drives oncogenic expression of *EVI1*. We generated an EVI1-GFP AML model and applied an unbiased CRISPR/Cas9 enhancer scan to uncover sequence motifs essential for *EVI1* transcription. Using this approach, we pinpointed a single regulatory element in the translocated *GATA2* enhancer that is critically required for aberrant *EVI1* expression. This element contained a DNA binding motif for the transcription factor MYB which specifically occupied this site at the translocated allele and was dispensable for *GATA2* expression. *MYB* knockout as well as peptidomimetic blockade of CBP/p300-dependent MYB functions resulted in downregulation of EVI1 but not of GATA2. Targeting *MYB* or mutating its DNA-binding motif within the GATA2 enhancer resulted in myeloid differentiation and cell death, suggesting that interference with MYB-driven *EVI1* transcription provides a potential entry point for therapy of inv(3)/t(3;3) AMLs.
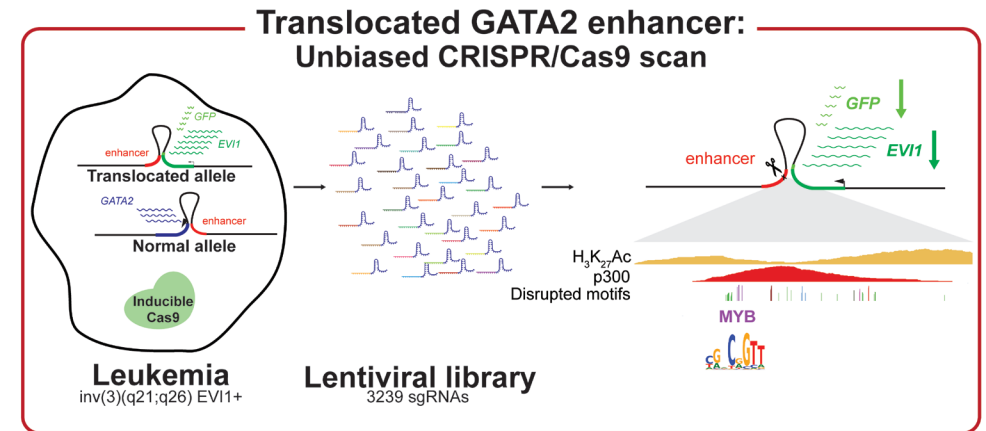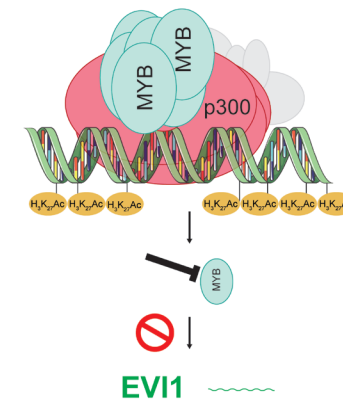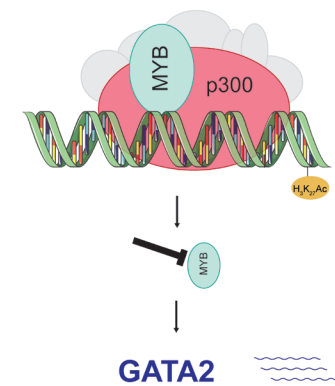
INTRODUCTION

Next-generation sequencing has greatly improved our knowledge about the location, distribution and frequency of recurrent gene mutations in cancer [1, 2]. The focus has previously been on the identification and understanding of mutations in protein coding regions. However, many mutations are found in intergenic regions as well [3], which now receive broad attention [4-8]. Those studies demonstrate that malignant transformation does not only rely on coding mutations in proto-oncogenes, but may also depend on aberrant regulation of oncogene expression. Well-described mechanisms of aberrant gene activation include generation of novel enhancers by nucleotide substitution, focal amplification of enhancers, loss of boundaries between topologically associated domains (TAD) or enhancer hijacking by chromosomal rearrangements [9-15].

Chromosomal inversion or translocation between 3q21 and 3q26 (inv(3)(q21;q26) or t(3;3)(q21;q26)) in AML result in the aberrant expression of the proto-oncogene *EVI1* located at the *MDS1* and *EVI1* complex locus (*MECOM*) at 3q26 [16-19]. Our group and others reported that hyper-activation of *EVI1* is caused by a *GATA2* enhancer translocated from chromosome 3q21 to *EVI1* on chromosome 3q26 [12, 20]. Upon translocation, this hijacked *GATA2* enhancer appears to behave as a super-enhancer and is marked by a broad stretch of H_3K_{27} acetylation [12, 20]. In the current study, we aimed to unravel the mechanism by which the hijacked *GATA2* super-enhancer leads to *EVI1* activation. We generated a model to study *EVI1* regulation in inv(3)/t(3;3) AML cells by inserting a *GFP* reporter 3' of endogenous *EVI1* and introduced an inducible Cas9 construct. To uncover important elements in this hijacked enhancer, we applied CRISPR/Cas9 scanning and identified motifs essential for driving *EVI1* transcription. We demonstrated a single regulatory element in the translocated GATA2 enhancer that is critical for the regulation of *EVI1* expression, with an essential role for MYB through binding to the translocated enhancer. Treatment of inv(3)/t(3;3) AML cells with peptidomimetic MYB:CBP/p300 inhibitor decreased EVI1 expression, and induced leukemia cell differentiation and cell death.

RESULTS

Expression of *EVI1* in inv(3)/t(3;3) AML is reversible

In inv(3)/t(3;3) AMLs the *GATA2* super-enhancer is translocated to *MECOM*, driving expression of *EVI1* [12, 20]. We investigated whether *GATA2* enhancer-driven transcription of *EVI1* in inv(3)/t(3;3) is reversible in leukemia cells. In primary inv(3)/t(3;3) AML, immature CD34+CD15- cells can be discriminated from more mature CD34-CD15- and CD34-CD15+ cells (Figure 1A,C and Figure S1A,C). Whereas EVI1 is highly expressed in CD34+CD15- cells, mRNA and protein levels decline in the CD34-CD15- fraction and are almost completely lost in CD34-CD15+ cells in inv(3)/t(3;3) primary AML as well as in MUTZ3 cells, an inv(3) AML model (Figure 1B-E, Figure S1B,D). Since 3q26 rearrangements are present in all fractions, as determined by three-colored Fluorescent in-situ hybridization (FISH) (Figure S1E), we conclude that transcription of *EVI1* can be reversed in AML cells despite the presence of a 3q26 rearrangement. *In vitro* culture of sorted MUTZ3 cells revealed that only the *EVI1*-expressing CD34+CD15- cells were competent to proliferate (Figure S1F), in agreement with previous observations showing that *EVI1* depletion results in loss of colony formation and induction of differentiation [12]. Thus, although AML cells with inv(3)/t(3;3) depend on *EVI1*, transcription of this gene remains subject to regulation and can be repressed, with major consequences for cell proliferation and differentiation.

Generation of an EVI1-GFP inv(3) AML model

Our findings indicate that interference with *EVI1* transcription may be an entry point to specifically target inv(3)/t(3;3) AMLs. To study the molecular mechanisms of *EVI1* transcriptional activation by the hijacked *GATA2* enhancer, we introduced a *GFP* reporter 3' of *EVI1* in MUTZ3 cells, with a T2A self-cleavage site separating the two proteins (Figure 2A, Figures S2A and S2B). Knockdown of *EVI1* using two unique *EVI1*-specific shRNAs (Figure 2B) resulted in a reduction of the GFP signal (Figure 2C). Subsequently, a construct with tight doxycycline (Dox) controlled expression of Cas9 was introduced into MUTZ3-EVI1-GFP cells (Figure S2C-D) and used to target the translocated *GATA2* enhancer and study *EVI1* regulation. Deletion of approximately 1000 bp of the translocated *GATA2* enhancer [12, 20] using two specific sgRNAs (Table S3) resulted in a severe decrease in *GFP* expression upon Dox treatment (Figure 2D). We sorted the GFP expressing cells into three fractions and observed that enhancer deletion was most pronounced in the GFPlow FACS-sorted cells (Figure 2E lower band). The GFPlow fraction also contained the lowest *GFP* and *EVI1* mRNA levels (Figure 2F-G). Cells from the GFPlow fraction, which showed reduced EVI1 expression, formed less colonies than GFPhigh cells in methylcellulose (Figure 2H). Only colonies obtained from the GFPhigh fraction consisted of cells able to multiply when placed in liquid culture (Figure S2E). Immunophenotyping of the colonies revealed that GFPhigh fractions predominantly consisted of immature CD34+CD15- cells, while in contrast the GFPlow fraction

contained the highest number of differentiated CD15+CD34- cells (Figure S2F). Together, this established a Dox-inducible Cas9 expressing inv(3)/t(3;3) AML model (MUTZ3-EVI1-GFP) for studying the transcriptional control of *EVI1* via a GFP reporter.

Unbiased CRISPR/Cas9 enhancer scan reveals one 1 kb region to be essential for *EVI1* activation

The minimally translocated region of the *GATA2* super-enhancer is 18 kb long [12]. In MUTZ3 and MOLM1, which are both inv(3) AML models, this highly H_3K_{27} acetylated region (Figure 3A; yellow) contains four loci of open chromatin determined by ATAC-seq (Figure 3A; orange), of which two show strong p300 occupancy (Figure 3A; red). To identify, in an unbiased fashion, which elements of the 18 kb translocated region control *EVI1* transcription, we employed a CRISPR/Cas9-based enhancer scanning approach (Figure 3A). We constructed a lentiviral library containing 3239 sgRNAs covering the 18 kb translocated region (Figure 3A, Table S1) and transduced it into MUTZ3-EVI1-GFP cells at a low multiplicity of infection. After neomycin selection and cell expansion, the cells were treated with Dox to induce Cas9 expression and cells displaying reduced GFP reporter expression (GFPlow) were selected by flow cytometric sorting at day 5 and day 7. sgRNAs were amplified from genomic DNA and deep-sequenced to identify the sgRNAs that were enriched in the GFPlow fraction. The log2fold change of 3 independent experiments were combined as shown in Figure 3B, which demonstrated a strong correlation between the sgRNAs enriched in GFPlow cells at day 5 and day 7 (Figure 3B).

Five sgRNAs targeting *EVI1* were the top scoring hits in the GFPlow fraction (indicated in blue), whereas sgRNAs targeting the safe harbor AASV1 locus (in red) were not enriched, emphasizing the specificity and sensitivity of the assay (Figure S3A). sgRNAs with a minimum of 3-fold enrichment in the GFPlow fraction all clustered in a small region of approximately 700 bp (Figure 3C). This region is a known p300-interacting region, which belongs to the -110 kb (-77 kb in mouse) **distal** *GATA2* enhancer [21, 22]. Both p300-interacting regions are occupied by a heptad of transcription factors (SCL, LYL1, LMO2, GATA2, RUNX1, FLI1 and ERG) that regulate genes in hematopoietic stem and progenitor cells (HSPCs) [23, 24] (Figure S3E). Only the first p300-interacting region appears to be important for *EVI1* expression in inv(3) AML. Approximately 40 sgRNAs within this region, with at least a 2-fold enrichment in the GFPlow fraction, were selected and cloned into a lentiviral construct with iRFP720 for individual testing. The loss of GFP signal at day 7 in the iRFP+ fraction (gating strategy, see figure S3D) highly correlated with the enrichment of those 40 sgRNAs in the GFPlow fraction as observed in the enhancer scan (Figure 3D). An efficiently cutting sgRNA that was not enriched in the enhancer scan did not affect GFP signal upon Dox exposure (Figure S3B,C). Deep amplicon sequencing of the -110 kb enhancer region upon targeting by 36 individual sgRNAs revealed frequent mutations in motifs for MYB, GATA, RUNX-, MEIS-, XBP- and ETS-binding sites, which were among the highest conserved (Figure 3E, Figure S3F, Table S2).

A MYB binding motif is essential for *EVI1* rather than for *GATA2* transcription

Four sgRNAs, i.e. # 3, 8, 11 and 16, generating the highest GFPneg (EVI1neg) fraction in the single guide validation experiments, all targeted the same region containing a potential MYB-binding motif (Figure 4A). The strong reduction of GFP expression, as tested for three of those guides (Figure 4B), was accompanied by loss of EVI1 protein (Figure 4C) and mRNA (Figure 4D). EVI1 loss was accompanied by differentiation into CD34-CD15+ cells in the sgRNA8-targeted GFPlow fraction (Figure 4E), in line with the findings in primary AML cells (Figure 1A,C and Figure S1A,C). Strikingly, sgRNA8-directed mutations within the enhancer did not affect GATA2 protein (Figure 4C) or mRNA levels (Figure 4D). Western blot analysis on sorted fractions of sgRNA8-treated cells revealed a strong reduction of EVI1 but not of GATA2 in GFPlow cells (Figure 4F). Amplicon-seq within the GFPlow sorted fraction of sgRNA8-treated cells revealed that almost 97% of the aligned sequences, including the translocated and non-translocated allele, were mutated (Figure 4G). In approximately 86% of all aligned sequences, the MYB motif was mutated. In 14%, a 20 bp deletion fully eliminated the predicted MYB DNA-binding motif (Figure 4H). We carried out pulldown experiments in which equal amounts of MUTZ3 nuclear lysates (Figure S4) were exposed to beads with immobilized 100 bp enhancer DNA fragments representing WT or MYB-motif mutant enhancer DNA, as defined in Figure 4H. Western Blot analysis confirmed MYB binding to the 100 bp WT enhancer fragment (Figure 4I). MYB binding to the M1 or M2 mutants was severely reduced, but it was preserved in the M3 mutant, in which the MYB DNA-binding motif was retained (Figure 4I). We conclude that in inv(3)/t(3;3) AML transcription of *EVI1* depends on the presence of a MYB DNA-binding motif in the translocated enhancer. Strikingly, this MYB motif appears less relevant for the transcription of *GATA2* in the non-translocated allele.

Differential MYB binding and H_3K_{27} acetylation at the hijacked *GATA2* enhancer

ChIP-seq revealed MYB occupancy at the -110 kb *GATA2* enhancer in MUTZ3 and in inv(3)/t(3;3) AML patient cells (Figure 5A,C; green tracks). Based on a heterozygous SNP in the -110 kb *GATA2* enhancer in MUTZ3, the translocated allele (*EVI1*) can be discriminated from the non-translocated (*GATA2*) allele [12]. We found approximately 7 times more MYB occupancy at the translocated allele (Figure 5B), in agreement with the finding that p300 occupancy (Figure 5A; red track) was also detected predominantly at the translocated enhancer (Figure 5B). Furthermore, H_3K_{27}Ac signal (Figure 5B) and open chromatin (ATAC) (Figure S5C) were 5 times more prevalent at the translocated enhancer. No SNPs were present in primary AMLs to discriminate MYB binding to the different alleles. However, based on two SNPs in the 18 kb region (Figure 5C), we observed a strong H_3K_{27}Ac allelic skewing of the primary inv(3)/t(3;3) AML, predicted to be biased to the translocated allele (Figure 5D). These data suggest that MYB and p300 interact with the -110 kb enhancer preferentially at the translocated allele. In sgRNA8-treated MUTZ3 cells (+Dox) MYB binding to the -110 kb

site was significantly decreased compared to control (−Dox) cells (Figure 5E). This loss was *GATA2* enhancer-specific, since genome-wide MYB chromatin occupancy, which includes the MYB target gene *BCL2*, did not change in +Dox cells (Figure S5A and S5B). Importantly, the decrease of MYB-binding at the -110 kb enhancer upon sgRNA8 treatment was greater within the translocated allele (Figure 5F). Using Cut&Run we demonstrated that $H_3K_{27}Ac$ was severely decreased at the enhancer in GFP^{low} sorted cells (Figure 5G, blue track) compared to GFP^{high} sorted cells (Figure 5G, green track) following sgRNA8 treatment. Moreover, SNP analysis revealed that the remaining H_3K27_{Ac} at the enhancer in GFP^{low} cells occurred predominantly at the non-translocated allele (*GATA2*) (Figure 5H). These data demonstrate that mutating the MYB binding motif at the translocated -110 kb enhancer decreases MYB binding, thus inactivating the enhancer and reducing *EVI1* transcription.

MYB interference downregulates EVI1 but not GATA2

To study whether MYB is important for *EVI1* expression, MYB-specific sgRNAs were introduced into MUTZ3-EVI1-GFP cells. At day 3 and 6 post-Dox induction, loss of MYB expression was evident, which was accompanied by a decrease of EVI1 protein (Figure 6A). In contrast, and similar to measurements made after mutation of the MYB binding motif, knockout of MYB did not decrease GATA2 protein expression (Figure 6A). The activity of MYB can be repressed using the peptidomimetic inhibitor MYBMIM, which impairs the assembly of the MYB:CBP/p300 complex [25]. In MUTZ3 cells, treatment with 25 µM MYBMIM caused a 50% reduction of viable cells, whereas the inactive MYBMIM analog TG3 showed no effect (Figure 6B). Treatment of MUTZ3 cells with 20 µM MYBMIM strongly reduced EVI1 protein levels (Figure 6C) without impacting MYB levels (Figure 6C). Consistent with the MYB knockout experiment (Figure 6A), MYBMIM treatment did not alter GATA2 protein levels (Figure 6C). A two-day exposure of MUTZ3 cells to MYBMIM reduced the number of colonies in methylcellulose (Figure 6D). Flow cytometric analysis of MYBMIM-treated colony cells revealed increased maturation (CD34$^-$CD15$^+$ cells) in comparison with TG3-treated controls (Figure 6E). Moreover, whereas MYBMIM treatment did not reduce MYB protein, it decreased MYB occupancy at the *GATA2* enhancer (Figure 6F). p300 occupancy also decreased, but to a lesser extent than MYB (Figure 6F). MYB binding was reduced at several sites, including at the *BCL2* enhancer (Figure 6G). MYBMIM, but not TG3, reduced viability of inv(3)/t(3;3) AML patient cells (n=3) (Figure 6H), and treatment of AML primary cells with MYBMIM reduced EVI1 protein levels without affecting levels of MYB or GATA2 (Figure 6I). Thus, targeting MYB represents a promising therapeutic possibility in the context of inv(3)/t(3;3) AMLs with *EVI1* overexpression.
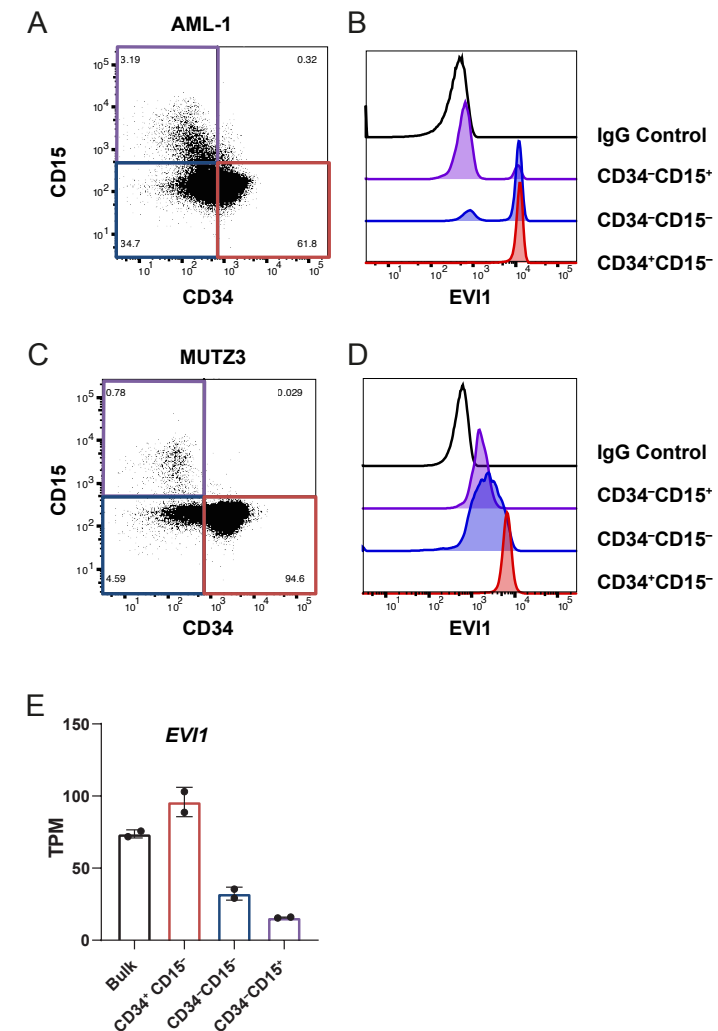

Figure 1. Expression of *EVI1* in inv(3)/t(3;3) AML is reversible
(A) Flow cytometric analysis of CD34- and CD15-stained inv(3;3) primary AML cells (AML-1). **(B)** Intracellular EVI1 staining in AML cells in the gated fractions as indicated in A. **(C)** Flow cytometric analysis of MUTZ3 cells stained with CD34 and CD15. **(D)** Intracellular EVI1 staining in the gated fractions as indicated in C. **(E)** Bar plot showing relative expression of *EVI1* in Transcripts Per Million (TPM) in sorted fractions of MUTZ3 cells. Error bars represent standard deviation of two biological replicates
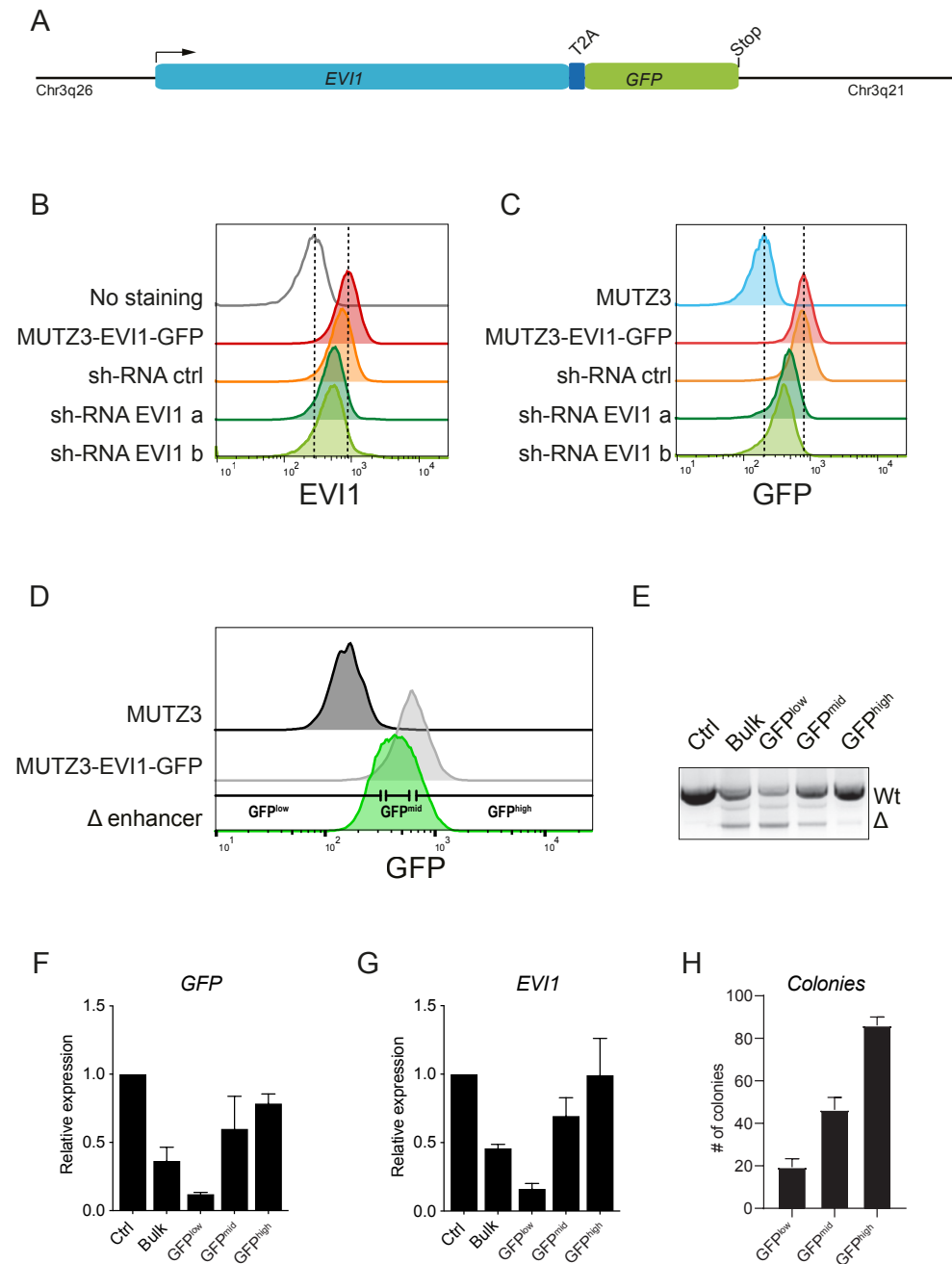
Figure 2. Generation of an EVI1-GFP inv(3) AML model
(A) Schematic representation of EVI1-GFP knock-in with a T2A self-cleavage site in the MUTZ3 cells at the endogenous *EVI1* locus. **(B)** Flow cytometric analysis of intracellular EVI1 after shRNA-mediated knockdown of *EVI1* using two different shRNAs. The effects on EVI1 protein were measured 48 hours after transduction. Scrambled shRNAs were used as control. **(C)** Flow cytometric analysis of GFP in the same experiment indicated in (B). **(D)** Representative flow cytometric plot showing the effect of the -110kb *GATA2* enhancer deletion in MUTZ3-EVI1-GFP cells (Δ enhancer). Cas9 was induced with Dox 24h before nucleofection of two sgRNAs. The effect on EVI1 was measured by GFP levels using flow cytometric analyses. Cells were sorted 48h after nucleofection of subsequent sgRNAs into three fractions: GFP^low^, GFP^mid^ and GFP^high^. **(E)** Genotyping PCR showing a wild type (WT) band (1500 bp) or a band for the enhancer deleted(Δ) (900 bp), either in bulk (before sorting) or in sorted fractions. Control (Ctrl) represents PCR after nucleofection of the sgRNAs without Dox induction. **(F)** Bar plot showing relative *GFP* expression of bulk and sorted fractions analyzed by qPCR. The expression levels of *PBGD*, a housekeeping gene, were used as control for normalization. Relative expression is calculated as fold over Ctrl (nucleofection of the sgRNAs without Dox). Error bars represent standard deviation of two biological replicates. **(G)** Bar plot showing relative *EVI1* expression of MUTZ3-EVI1-GFP bulk and sorted fractions analyzed by qPCR. For details see Figure 2F legend. **(H)** Bar plot showing the number of colonies grown in methylcellulose from each sorted fraction. Colonies were counted 1.5 weeks after plating. Error bars represent standard deviation of three plates.
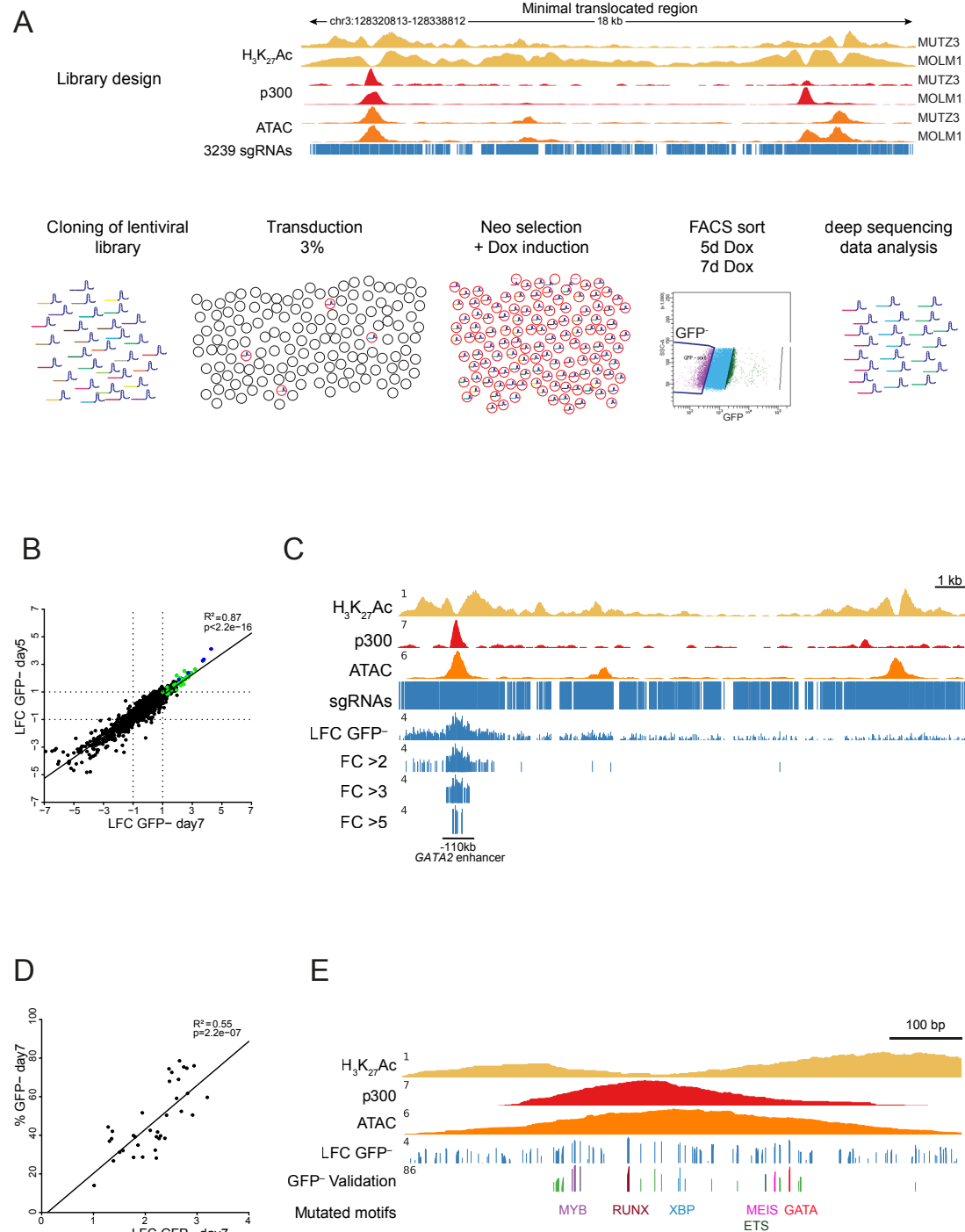
Figure 3. Unbiased CRISPR/Cas9 enhancer scan reveals one 1 kb region to be essential for *EVI1* activation
(A) ChIP-seq to determine $H_3K_{27}Ac$ pattern and p300 binding as well as open chromatin analysis using ATAC-seq in MUTZ3 and MOLM1 cells. The locations of the >3200 sgRNAs targeting the enhancer are indicated as vertical blue lines. A schematic overview of the enhancer scanning strategy is depicted below. **(B)** Scatter plot of enrichment of sgRNAs in sorted GFP^{low} fractions at day 5 and day 7 upon Dox induction. The average of three independent experiments for each dot is depicted. For every sgRNA detected in the GFP^{low} fractions the log2fold change (LFC) of the +Dox relative to −Dox was calculated. Five sgRNAs targeting *EVI1* were added to the sgRNA library as positive controls and are indicated in blue. The sgRNAs selected for further validation are indicated in green. The fitted linear regression and corresponding R-squared and p-value are indicated. The LFC enrichment at day 7 of all sgRNAs and of sgRNAs with >2, >3 or >5 fold enrichment of sgRNAs in the GFP^{low} fractions at the 18 kb region of the *GATA2* super-enhancer in MUTZ3 cells is depicted. The $H_3K_{27}Ac$ pattern, p300 binding, open chromatin (ATAC) and location of all sgRNAs are indicated to visualize which sgRNAs were enriched in the GFP^{low} fraction. The -110 kb **distal** *GATA2* enhancer is indicated. **(C)** Scatter plot showing enrichment of sgRNAs in sorted GFP^{low} fractions at day 7 compared to %GFP^{neg} cells at day 7 for individually validated sgRNAs (based on two independent biological experiments). The sgRNAs used for validation are indicated by dots. The fitted linear regression and corresponding R-squared and p-value are indicated. **(D)** Zoom-in of the -110 kb *GATA2* enhancer (chr3:128322411-128323124) showing $H_3K_{27}Ac$ pattern, p300 binding and open chromatin (ATAC), LFC enrichment of sgRNAs at day 7 and the %GFP^{neg} cells at day 7 of the individually validated sgRNAs. Mutations in motifs for known transcription factors identified in the individually validated sgRNAs are indicated.
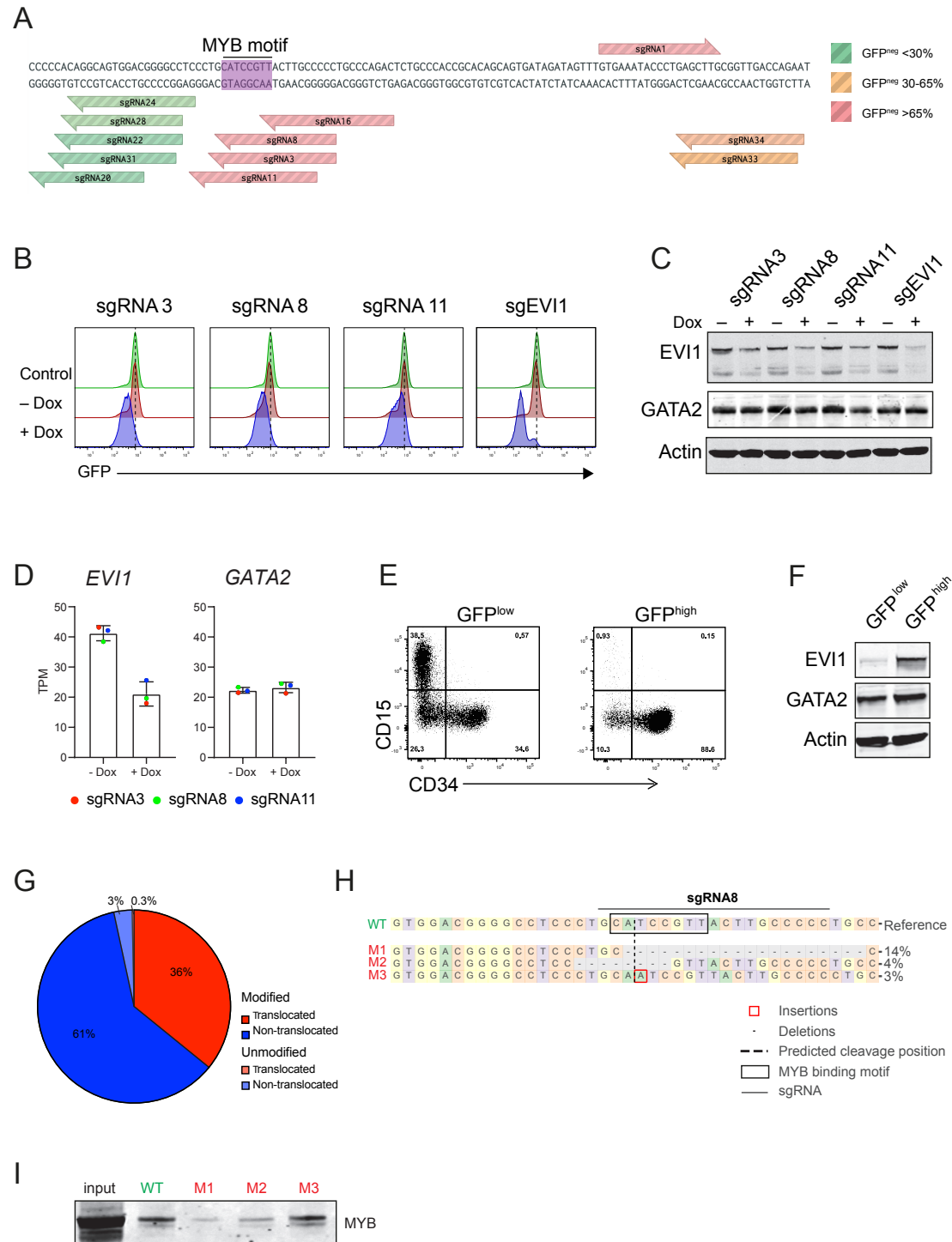
Figure 4. A MYB binding motif is essential for *EVI1* rather than for *GATA2* transcription

(A) Nucleotide sequence of the region targeted by sgRNAs 3,8,11 and 16, as well as other nearby sgRNAs, with the corresponding MYB DNA binding motif highlighted in purple. Colors of sgRNAs represent differences in percentage of recovery in the GFP^neg fraction. sgRNAs indicated in red are the most highly enriched in the GFP^neg fraction.

(B Flow cytometric analysis of MUTZ3-EVI1-GFP cells upon sgRNA treatment. GFP signal shifts are shown upon transduction with lentivirus containing sgRNAs 3, 8, 11 or an EVI1-specific sgRNA. Cells were analyzed by flow cytometry 7 days after induction of Cas9. **(C)** Western blot using EVI1- and GATA2-specific antibodies upon transduction with lentivirus containing sgRNAs 3, 8, 11 or an *EVI1* specific sgRNA (EVI1.4) analyzed 7 days after induction of Cas9. Actin was used as loading control. **(D)** Bar plot showing relative expression of *EVI1* and *GATA2* in transcripts per million (TPM) in MUTZ3-EVI-GFP cells treated with sgRNAs 3, 8 or 11, −Dox or +Dox. The cells treated with sgRNAs 3, 8, or 11 were considered replicates and standard deviation is shown. **(E)** CD34/CD15 flow cytometric analyses of MUTZ3 EVI1-GFP cells transduced with sgRNA8 (+Dox), sorted for GFP^low or GFP^high and analyzed two weeks after sorting. **(F)** EVI1 and GATA2 western blot upon treatment with sgRNA 8, sorted into GFP^low or GFP^high fractions, 7 days after induction of Cas9. Actin was used as loading control. **(G)** Editing frequency in the GFP^low fraction of sgRNA8-treated cells. Modified reads exhibited variations with respect to the reference human sequence. The percentages of reads that align to each allele were determined based on a heterozygous SNP in the sequenced region. **(H)** Visualization of the distribution of mutations identified around the sgRNA8 target site in the GFP^low sorted fraction. The sgRNA8 target site is indicated (GGGGGCAAGTAACGGATGC) as well as the MYB binding motif (black rectangle). **(I)** Western blot using anti-MYB antibody in MUTZ3 cell lysates following pulldowns using WT, mutated M1, M2 or M3 100bp DNA fragments.
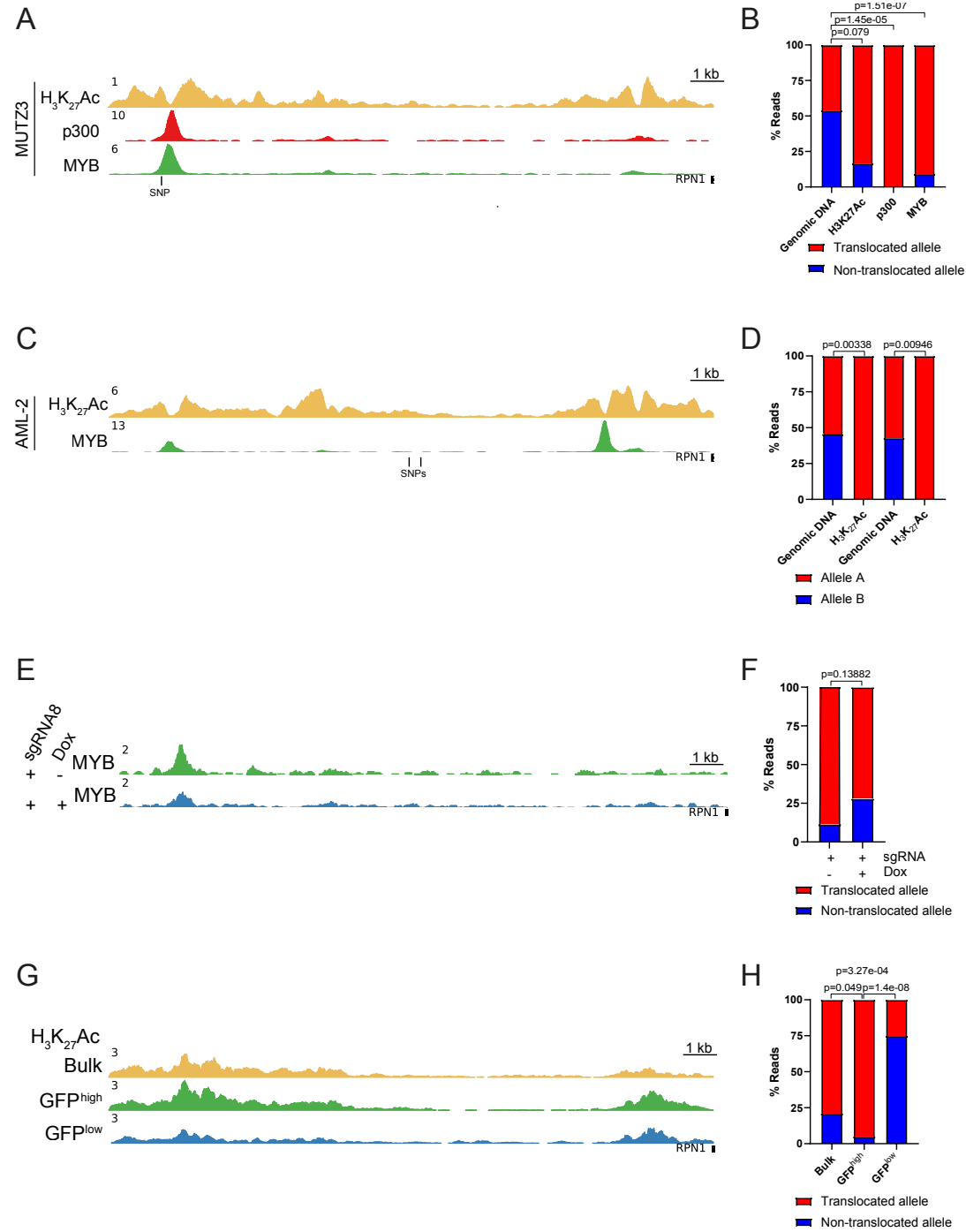
Figure 5. Differential MYB binding and H_3K_{27} acetylation at the hijacked *GATA2* enhancer
(A) H_3K_{27}Ac, p300 and MYB ChIP-seq profiles of the 18 kb super-enhancer region in MUTZ3 cells. **(B)** Bar plot showing allelic bias towards the translocated allele for H_3K_{27}Ac, p300 and MYB occupancy by ChIP-seq analysis based on a SNP (rs553101013). Previous sequencing showed that G represents the translocated allele and A the wild type allele (Groschel et al., 2014). P-values were calculated using a χ^2 test. **(C)** H_3K_{27}Ac and MYB ChIP-seq profiles of the 18 kb super-enhancer in an AML patient with inv(3) (AML-2). **(D)** Bar plot showing discrimination between H_3K_{27}Ac at the two *GATA2* enhancer alleles based on two SNPs (rs2253125 and rs2253144). P-values were calculated using a χ^2 test. **(E)** MYB ChIP-seq profile of the 18 kb super-enhancer in sgRNA8-treated MUTZ3-EVI1-GFP cells plus or minus Dox treatment. **(F)** Bar plot showing allelic distribution of MYB binding in sgRNA8 treated MUTZ3-EVI1-GFP cells plus or minus Dox treatment. P-values were calculated using a χ^2 test. **(G)** H_3K_{27}Ac profile of the 18 kb super-enhancer in sgRNA8-treated MUTZ3-EVI1-GFP cells, determined by Cut&Run in bulk, in GFP[high] and in GFP[low] sorted fractions. **(H)** Bar plot showing allelic bias for H_3K_{27}Ac in the bulk, GFP[high] and GFP[low] fractions. P-values were calculated using a χ^2 test.
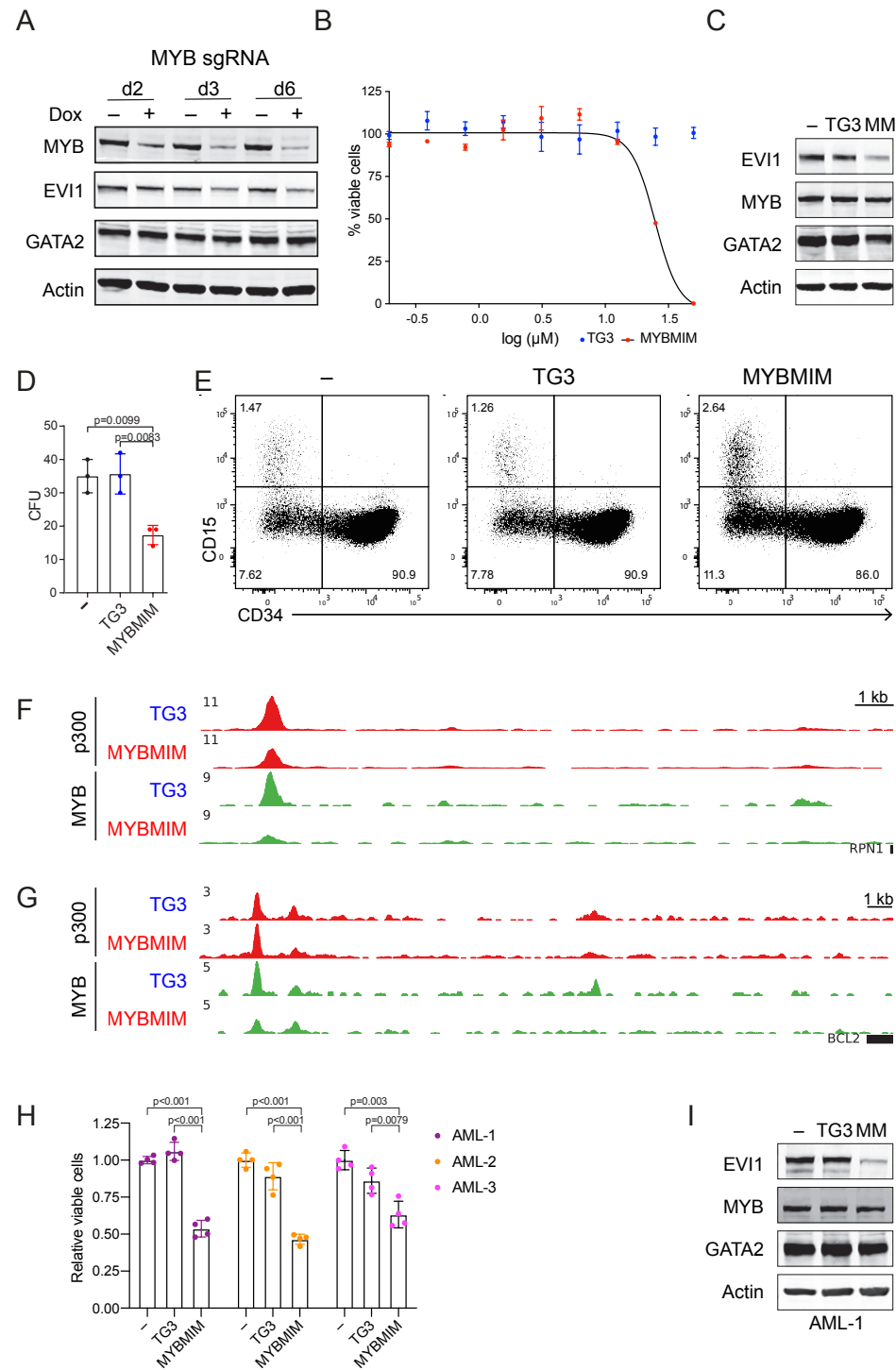
Figure 6. MYB interference downregulates EVI1 but not GATA2

(A) Western blot for MYB, EVI1 and GATA2 in MUTZ3-EVI1-GFP upon sgRNA-mediated MYB knockdown (MYB.30) at indicated days after induction of Cas9. Actin was used as loading control. **(B)** MUTZ3 cells were treated with either TG3 (blue) or MYBMIM (red) at indicated concentrations and cell viability was determined by CellTiter-Glo three days after plating the cells in triplicate. **(C)** Western blot for MYB, EVI1 and GATA2 in untreated cells (-) or cells treated for two days with 20 μM of TG3 or MYBMIM (MM). Actin was used as loading control. **(D)** Colony forming units (CFU) of MUTZ3 cells cultured without peptide or treated with 20 μM TG3 or MYBMIM for two days and subsequently plated in methylcellulose. Error bars show standard deviation across three plates. P-values were calculated using a one-way ANOVA test. **(E)** Flow cytometric analysis of MUTZ3 cells stained with CD34 and CD15. Cells studied by flow cytometry were either untreated or treated with 20 μM TG3 or MYBMIM for two days and subsequently grown for nine days in methylcellulose. **(F)** p300 and MYB ChIP-seq profiles of the 18 kb region in MUTZ3 cells treated with either 20 μM TG3 or MYBMIM for 48 h. **(G)** p300 and MYB ChIP-seq profiles of the BCL2 region in MUTZ3 cells treated with either 20 μM TG3 or MYBMIM for 48 h. **(H)** Cell-viability test of inv(3)/t(3;3) AML primary cells determined by CellTiter-Glo three days after culturing the cells in a 96-well plate with 20 μM TG3 or MYBMIM. Error bars show standard deviation across four biological replicates. P-values were calculated using a one-way ANOVA test. **(I)** Western blot for MYB, EVI1 and GATA2 in untreated AML cells or in AML cells treated with 20 μM TG3 or MYBMIM for 48h. Actin was used as loading control.

DISCUSSION

Although multiple examples of hijacked enhancers causing uncontrolled expression of proto-oncogenes have been reported in various types of cancer [8, 10, 12, 26, 27], insight into their altered biological function remains limited. Elucidating these functions could provide opportunities for tailored interference and tools for therapeutic exploitation. Our unbiased CRISPR/Cas9 scan of the translocated 18 kb region in inv(3)/t(3;3) AMLs revealed a single region of approximately 1 kb essential for *EVI1* activation and leukemogenesis. This distal *GATA2* enhancer contained several conserved transcription factor DNA binding motifs, including an element preferentially occupied by MYB at the translocated allele (Figure 7, upper panel). Strikingly, mutating this MYB binding motif in the enhancer at both alleles strongly decreased the expression of *EVI1,* but not of *GATA2.* Together, these findings support a unique role for MYB in driving *EVI1* expression via the translocated enhancer, and suggest a potential vulnerability in inv(3)/t(3;3) AMLs. Indeed, peptidomimetic inhibition of MYB:CBP/p300 assembly in inv(3)/t(3;3) AML cells reduced EVI1 but not GATA2 protein levels, causing myeloid differentiation and cell death (Figure 7, lower panel). This fortifies the hypothesis that interfering with *EVI1* expression via MYB may constitute a new entry point for targeting these AMLs. The fact that targeting MYB specifically compromises *EVI1* expression points to the possibility of selectively target leukemia cells while sparing normal HSPCs, in which GATA2 is a vital regulator.

Although *MYB* encodes a transcription factor essential for normal hematopoiesis [28], there is also overwhelming evidence that it plays a critical role in malignant transformation. MYB was first discovered as an oncogene (v-*myb*) within the avian myeloblastosis virus (AMV) genome which generated myeloid leukemias in chickens [29, 30]. Its critical involvement in super-enhancer activity was previously shown in human T-cell acute lymphoblastic leukemia (T-ALL) [9, 31]. Mutations in non-coding regions near *TAL1* or *LMO2* create *de novo* binding sites for MYB, leading to the formation of new MYB-bound super-enhancers which drive uncontrolled transcription of those target genes. Furthermore, MYB binds to a translocated super-enhancer driving *MYB* expression in adenoid cystic carcinoma, creating a positive feedback loop sustaining its own expression [27]. MYB is also frequently overexpressed in human myeloid leukemias [32, 33] and AML cells can be addicted to high levels of MYB and thus be more vulnerable to MYB inhibition than normal hematopoietic progenitor cells [34]. However, the mechanisms whereby MYB drives transformation to AML are not fully understood. To our knowledge, our results in this study represent the first example of a mechanism by which MYB drives oncogene activation in AML (Figure 7).

MYB occupies the translocated *GATA2* enhancer at a level considerably higher than the non-translocated enhancer. This may reflect increased chromatin accessibility as determined by H$_3$K$_{27}$Ac ChIP-seq and ATAC-seq. The mechanisms driving this open chromatin pattern at the translocated locus remain a focus of future studies. However, translocation of the enhancer to a new location places it in proximity to distinct promoters and regulatory elements which may ultimately impact chromatin accessibility and MYB binding. In support of this hypothesis, mutating the MYB DNA binding site or interference with MYB function causes reduced expression of EVI1 but not GATA2.

The coactivators CBP and p300 are major mediators of MYB transcriptional activity [35, 36]. Therefore, specifically targeting the MYB:CBP/p300 interaction has been the focus of most small molecules seeking to inhibit MYB activity [25, 37-40]. Experiments using the peptidomimetic inhibitor MYBMIM, which blocks the formation of MYB:CBP/p300 complex, showed a severe loss of EVI1 activity. As reported by Ramaswamy et. al., [25], we also observed that MYBMIM caused loss of MYB binding to the enhancer, with largely preserved total cellular levels of MYB. Concurrently, we observed that MYBMIM treatment did not inhibit p300 occupancy at the enhancer to the same extent as MYB occupancy. This partially retained p300 binding could be explained by the presence of other transcription factors bound at the *GATA2* enhancer that also recruit CBP/p300 (Figure S3E). Finally, while initial results with MYBMIM peptide treatment of inv(3)/t(3;3) AML cells are promising, future studies will be necessary to evaluate whether other MYB target genes are impacted in healthy cells under these conditions.

Our CRISPR/Cas9 scan identified one p300-interacting region containing a MYB DNA binding motif to be important for *EVI1* expression. Although mutations in the MYB DNA-binding motif had the biggest impact on *EVI1* expression, other mutations also reduced *EVI1* levels. These included mutations in consensus DNA binding sites for GATA-, RUNX-, MEIS-, XBP- and ETS-factors. Interestingly, some of these factors have been demonstrated to occupy the -110 kb enhancer in CD34+ cells, including RUNX1, ERG and GATA2 [24]. MYB binding and activity at the -110 kb *GATA2* enhancer most likely occurs in conjunction with p300 as well as transcription factors like RUNX1 and ERG. This is in accordance with other studies showing co-localization and potential cooperation between these factors and MYB [25, 41, 42]. Therefore, combinatorial targeting of MYB and other transcription factors may synergistically impact *EVI1* expression. This knowledge provides a rationale to develop new compounds to treat inv(3)/t(3;3) AML, which can be tested in our newly developed model.

Our findings provide important insight into the mechanisms of oncogenic enhancer-driven gene activation in AML. The selective MYB motif requirement for enhancer function at the translocated but not the normal allele constitutes a paradigm in which chromosomal aberrations reveal critical motifs that are non-functional at their endogenous locus. In principle, this paradigm may be extrapolated to other enhancer-driven cancers and even non-malignant pathologies.
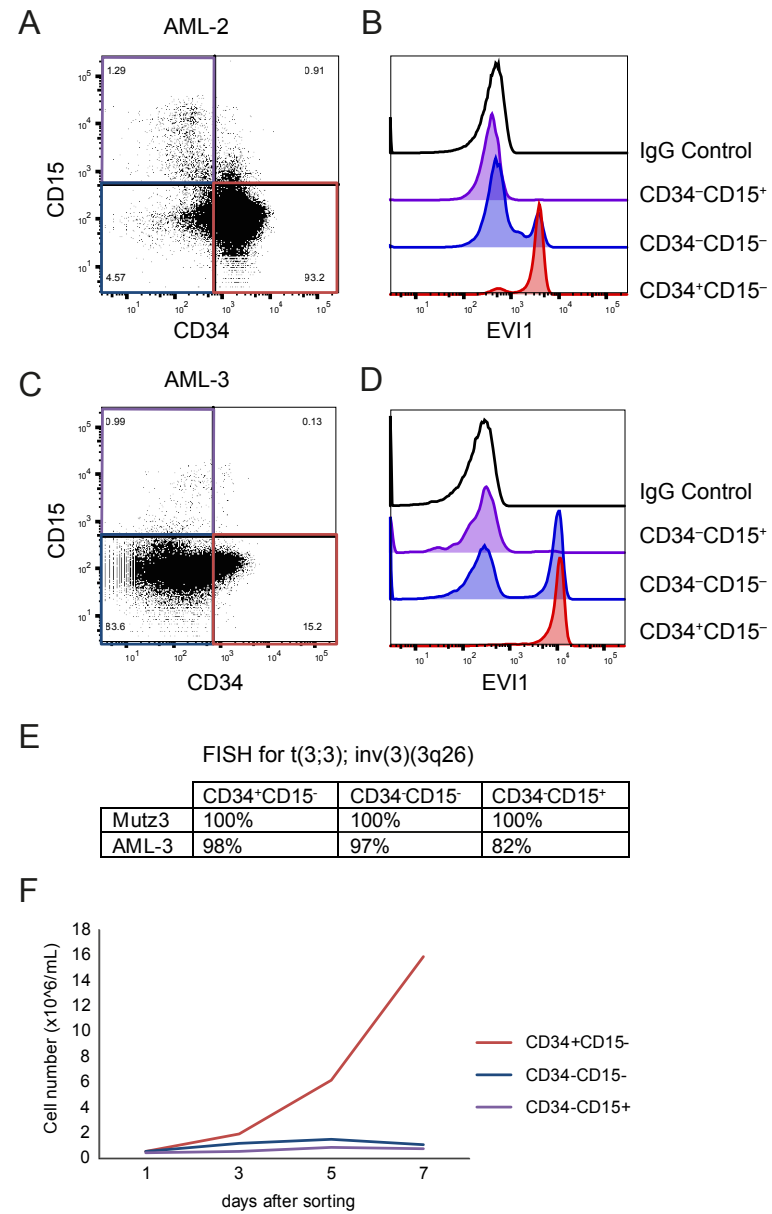
Figure S1. Expression of *EVI1* in inv(3)/t(3;3) AML is reversible (A) Flow cytometric analysis of CD34- and CD15-stained inv(3;3) primary AML cells (AML-2) **(B)** Intracellular EVI1 staining in AML cells in the gated fractions as indicated in A. **(C)** Flow cytometric analysis of CD34- and CD15-stained inv(3;3) primary AML cells (AML-3). **(D)** Intracellular EVI1 staining in the gated fractions as indicated in C. **(E)** Percentage of cells that were found positive for *EVI1*/3q26 rearrangements determined by three-colored FISH. **(F)** Line graph showing numbers of MUTZ3 cells sorted into the indicated fractions and cultured for seven days.
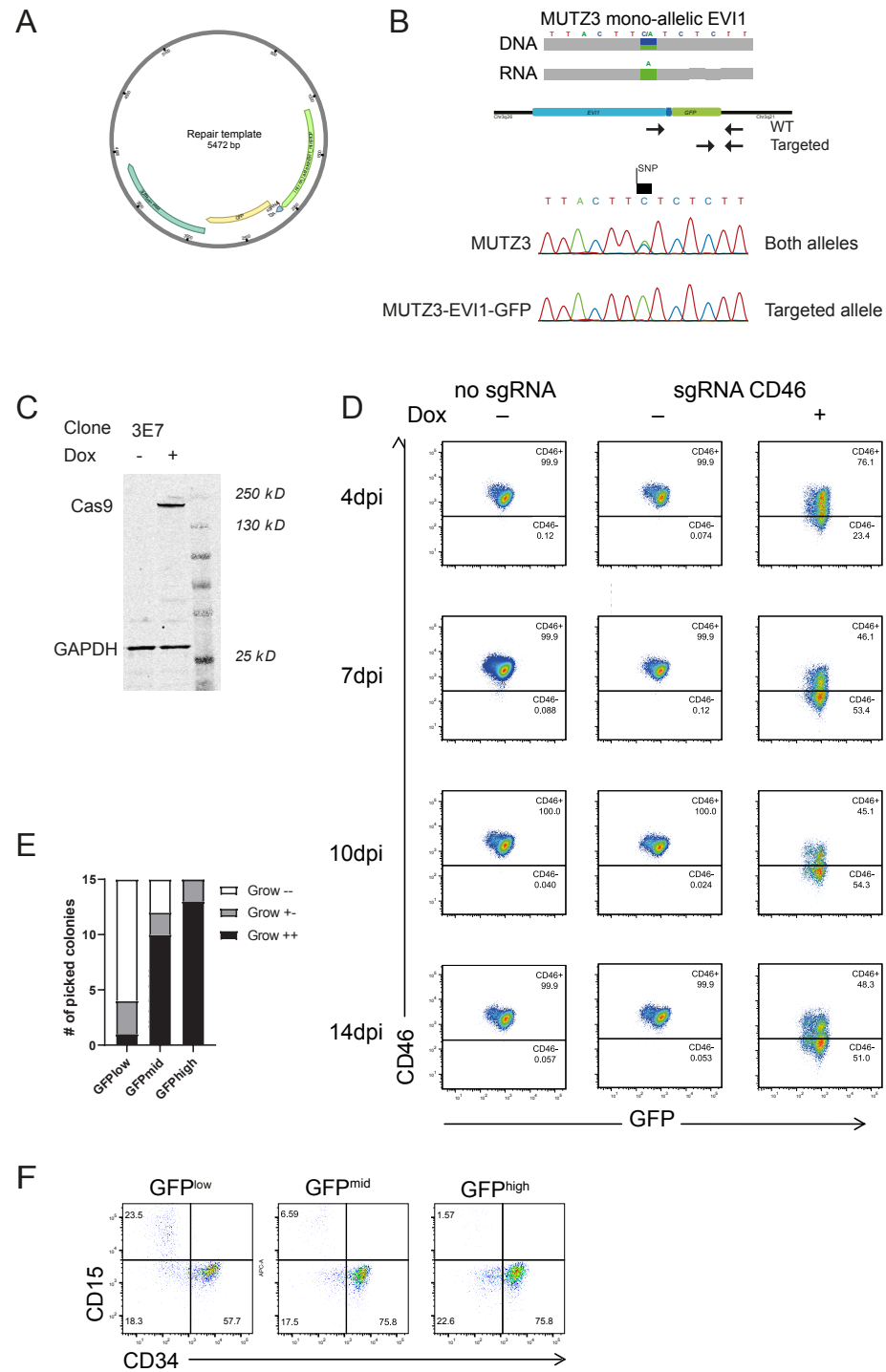
Figure S2. Generation of an EVI1-GFP inv(3) AML model

(A) Repair template consisting of a PUC19 backbone with a homology arm of the intron and last exon of *EVI1* minus the STOP, a T2A site and GFP and the second homology arm. The PAM sequence of sgRNA was omitted. **(B)** Mono-allelic expression of *EVI1* in MUTZ3 cells based on SNP differences (upper panel). PCR strategy (middle panel) for Sanger sequencing of genomic DNA of MUTZ3 WT and MUTZ3-EVI1-GFP (lower panel). **(C)** Western blot showing Dox-inducible Cas9 protein expression in MUTZ3-EVI1-GFP. Cas9 was induced with 1ug/ml Dox for 48h. **(D)** Testing tightness of the system by sgRNA-mediated knockdown of the cell surface marker CD46. Cells were transduced and followed up by flow cytometric analysis for two weeks at indicated days post infection (dpi). Without Dox no knockdown of CD46 is detected, whereas upon Dox exposure a strong effect on CD46 levels was observed. **(E)** Growth in liquid cultures of fifteen colonies picked from methylcellulose (from experiment Figure 2H) from each sorted fraction. For each well, growth was defined as no growth (--), slow growth (+/-) or normal growth (++). **(F)** Flow cytometric analysis of cells of sorted fractions 12 days after plating in methylcellulose stained with CD34 and CD15.
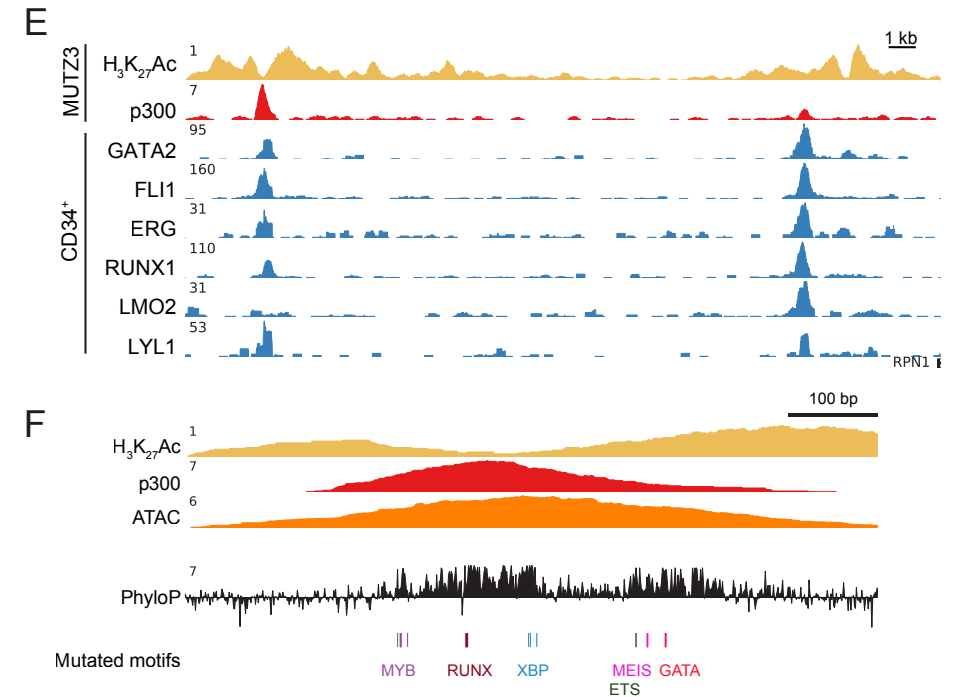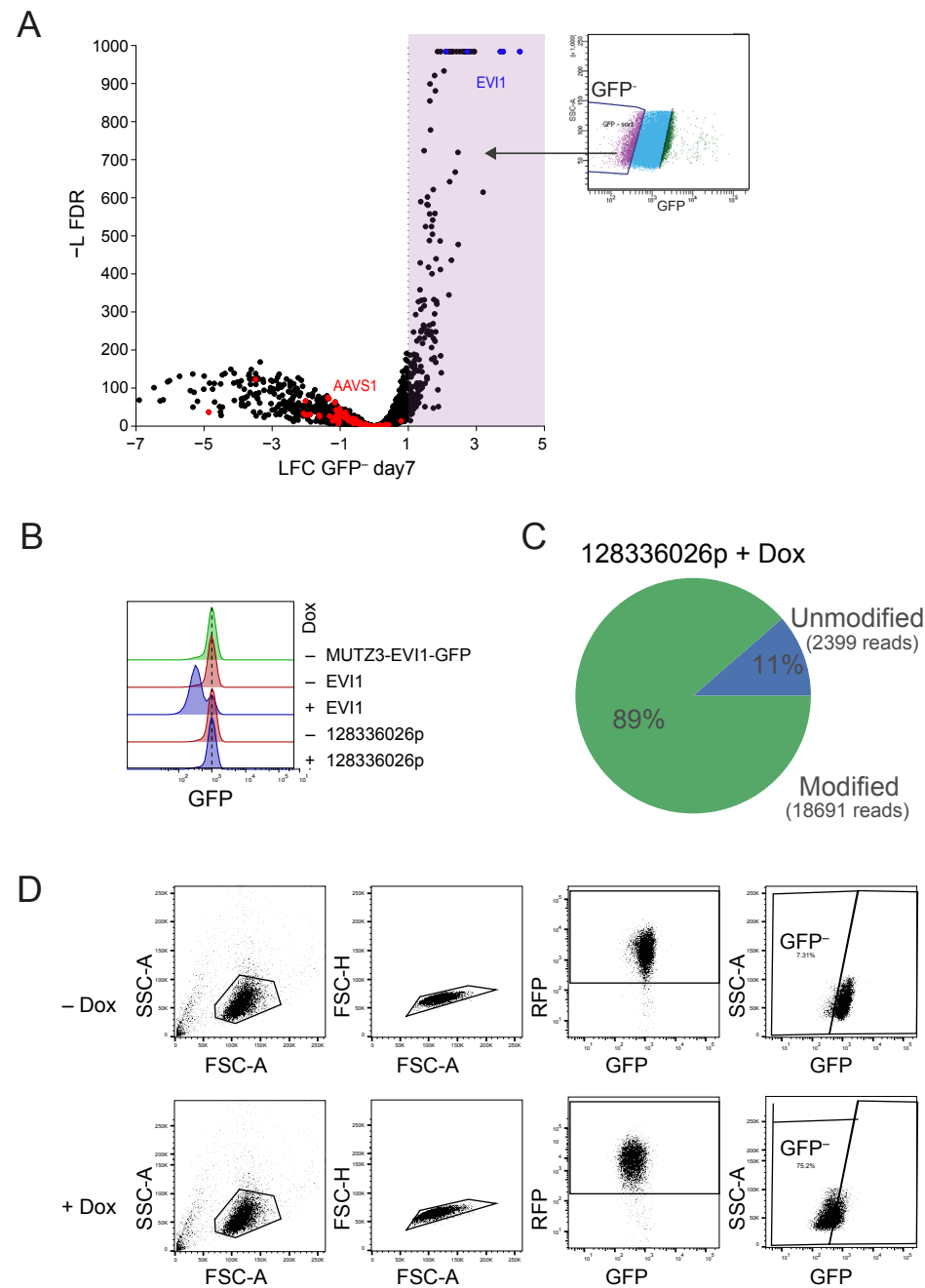
Figure S3. Unbiased CRISPR/Cas9 enhancer scan reveals one 1 kb region to be essential for *EVI1* activation
(A) Volcano plot of enrichment of sgRNAs in sorted GFP^{low} fractions at day 7, based on three independent experiments. For every sgRNA detected in the GFP^{low} fractions, the log2 fold change (LFC) compared to –Dox values was plotted as a dot. Five sgRNAs targeting *EVI1* were added to the sgRNA library as positive controls and are indicated in blue, sgRNAs targeting the AAVS1 region were added as negative controls indicated in red. The area highlighted in purple shows sgRNAs enriched at least two-fold (L2FC>1). **(B)** Flowcytometric analysis of MUTZ3-EVI1-GFP cells upon sgRNA treatment. GFP signal shifts are shown upon transduction with lentivirus containing sgRNAs targeting *EVI1* or a sgRNA targeting a region (Chr3:128336026) which was not enriched in the enhancer scan. Cells were analyzed by flow cytometry 7 days after Dox induction of Cas9. **(C)** Editing frequency of the sgRNA 128336026p. Modified reads exhibited variations with respect to the reference human sequence. **(D)** Flow cytometry-gating strategy for validation experiments. A lentiviral vector containing sgRNA8 was transduced, Dox was added (+/- 4 days after transduction) and the cells were analyzed by flow cytometry at day 7 after Dox admission. **(E)** Genome profile of the 18 kb translocated region with ChIP-seq showing binding of H$_3$K$_{27}$Ac, p300 and MYB in MUTZ3 cells and ChIP-seq of heptad transcription factors in CD34$^+$ cells (Beck et al., 2013). **(F)** Zoom-in of the -110 kb *GATA2* enhancer showing binding of H$_3$K$_{27}$Ac, p300 and open chromatin (ATAC), a track containing conservation scoring by phyloP (phylogenetic p-values) from the PHAST package (http://compgen.bscb.cornell.edu/phast/) and the mutations affecting motifs for known transcription factors in the GFP^{neg} fractions.
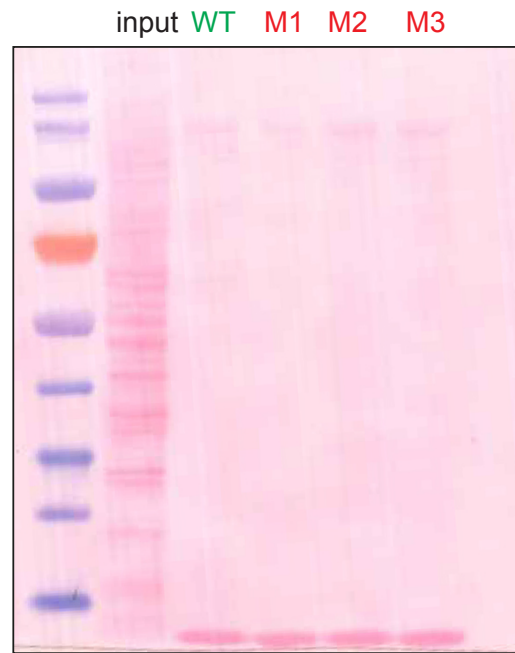
Figure S4. A MYB binding motif is essential for *EVI1* rather than for GATA2 transcription
Ponceau red staining showing equal loading of protein of MUTZ3 protein lysates after DNA pulldown with immobilized WT, M1, M2 or M3 sequences.
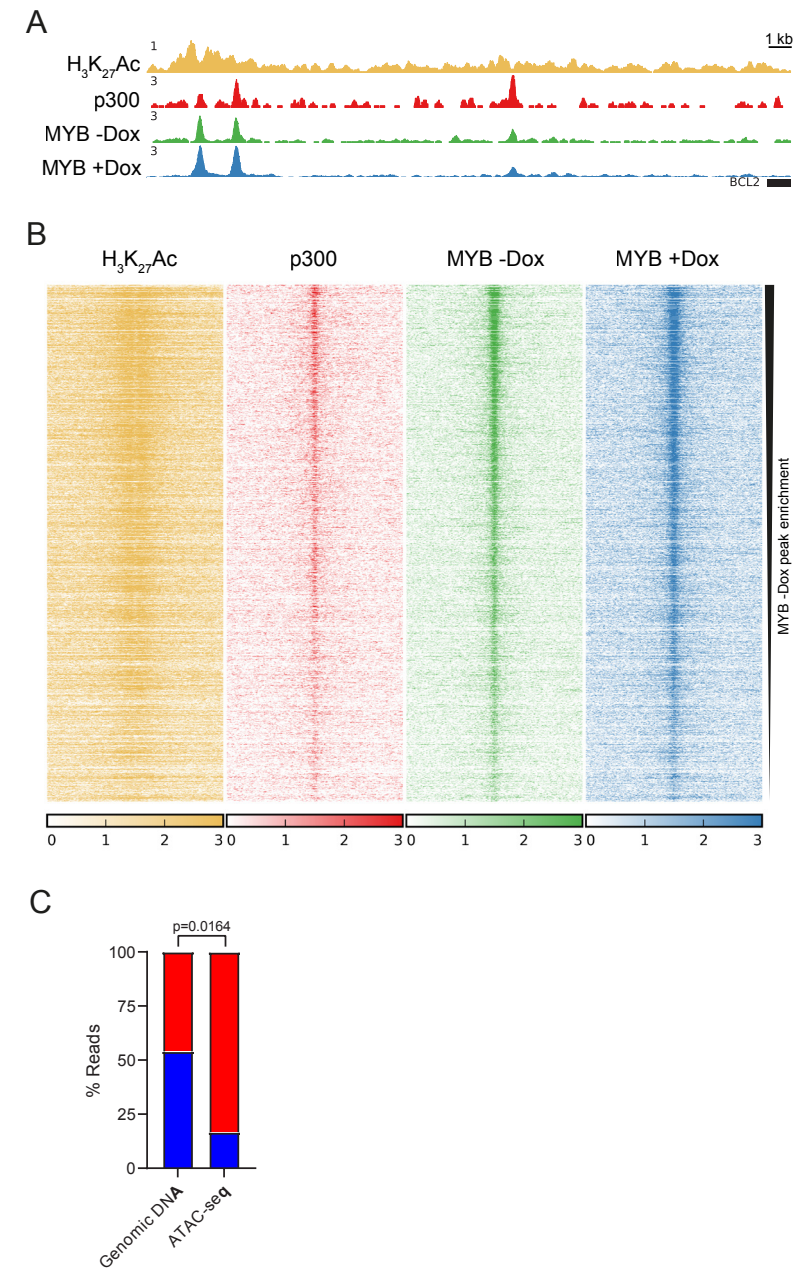
Figure S5. Differential MYB binding and H_3K_{27} acetylation at the hijacked *GATA2* enhancer
(A) Pattern of H_3K_{27}Ac, p300 ChIP-seq in MUTZ3 cells at the BCL2 enhancer as well as MYB binding in MUTZ3-EVI1-GFP cells transduced with sgRNA8 with (+) or without (-) Dox induction of Cas9 for 7 days. **(B)** Heatmap showing genome-wide MYB binding peaks as determined by the MACS2 peak-calling algorithm, sorted for fold enrichment of MYB without Dox exposure. H_3K_{27}Ac and p300 binding at those sites are shown as well. **(C)** Bar plot showing allelic bias towards the translocated allele for ATAC-seq analysis based on a SNP (rs553101013). P-values were calculated using a χ^2 test.

METHODS

Data and Code Availability

Sequence data generated in this study have been deposited at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/), which is hosted by the European Bioinformatics Institute (accession number EGAS00001004839). This study did not generate any unique codes. All software tools used in this study are freely or commercially available and listed in the Key Resources table.

Cell culture

The MUTZ3 cell lines were cultured in αMEM (HyClone) with 20% fetal calf serum (FCS) and 20% conditioned 5637 medium. The 293T were cultured in DMEM (Gibco) with 10% FCS. All cell lines were supplemented with 50 U/mL penicillin and 50 µg/mL streptomycin. Viable frozen AML cells were thawed and resuspended in IMDM medium supplemented with: 20% BIT medium (Stem cell technologies), 1x β-mercaptoethanol (1000x Life technologies), 6 µg/ml LDL (Sigma Aldrich), human IL6, IL3, G-CSF, GM-CSF at 20 ng/ml and FLT3, SCF at 50 ng/ml (Peprotech).

Generation of model lines

The repair template was generated using Gibson Assembly (NEB). Both homology arms were PCR amplified from MUTZ3 genomic DNA using Q5 polymerase (NEB). The first homology arm consists of a part of the intron and last exon of *EVI1* minus the STOP codon. The second homology arm consists of part of the 3'UTR with the PAM sequence of sgRNA omitted. The T2A-eGFP was PCR amplified from dCAS9-VP64_2A_GFP (Key Resources table). All fragments were cloned using Gibson assembly into the PUC19 backbone. sgRNA sequence AGCCACGTATGACGTTATCA was cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9. Cells were nucleofected with pX330 vector containing the sgRNA and Cas9 and the repair template using the Nucleofector 4D (Lonza) with Kit SF and program DN-100. GFP$^+$ cells were sorted using a FACS AriaIII (BD Biosciences). In a second sorting round, GFP$^+$ cells were single cell sorted and tested for proper integration. Clone 1A5 was transduced with lenti pCW-Cas9, puromycin selected (1 µg ml^{-1}) and subsequently single cell sorted based on GFP positivity and tested for inducible Cas9 expression. Clone 3E7 was used for the screen, which we called MUTZ3-EVI1-GFP.

Patient material

Samples of the selected patients presenting with AML were collected from the Erasmus MC Hematology Department biobank (Rotterdam, the Netherlands). Leukemic blast cells were purified from bone marrow or blood by standard diagnostic procedures. All patients provided written informed consent in accordance with the Declaration of Helsinki.

The Medical Ethical Committee of the Erasmus MC has approved usage of the patient rest material for this study.

Western Blotting

Cells were lysed in lysis buffer (20 mM Tris-HCL, 138 mM NaCl, 10 mM EDTA, 50 mM NaF, 1% Triton, 10% glycerol, 2 mM NA-vanadate) containing Complete protease inhibitors (CPI, Roche #4693159001). Protein levels were detected using antibodies against EVI1 (Cell Signaling, #2265), MYB (Merck Millipore, clone1-1, #05-175), B-Actin (Sigma, clone AC15, #A5441), GAPDH (Santa Cruz, #A5441), CAS9 (Biolegend, clone 7A9, #7A9) or GATA2 (kind gift of E.H. Bresnick, Department of Cell and Regenerative Biology, Madison, WI). Proteins were visualized using the Odyssey infrared imaging system (Li-Cor).

Flow cytometric analysis

Cell Sorting was performed using the FACS Aria flow cytometer (BD Biosciences) into a 96-well plate format (StemCell Technologies) or into batch culture. Flow Cytometric analysis on MUTZ3 cells was done with GFP/RFP or antibody stainings for CD34-PE-CY7 (BD, clone 8G12, #348811) and CD15-APC (Sony, #2215035) or CD15-BV510 (Biolegend, clone W6D3, #323028). Intracellular stainings with EVI1 (Cell Signaling, #2256) or Rabbit (DA1E) mAb IgG XP® Isotype Control (Cell Signaling, #3900) were performed using Foxp3/Transcription Factor Staining Buffer Set (00-5523-00, eBioscience). Cells were measured on a BD Canto or BD LSR II flow cytometer (BD Biosciences), and data was analyzed using FlowJo.

DNA pulldown

Nuclear lysates for pulldown experiments were prepared as described [43]. Oligo nucleotides for affinity purification were ordered as custom-synthesized oligos from Integrated DNA Technologies (IDT) (see Table S4). DNA pulldown was performed as described by Karemaker and Vermeulen with minor changes. Essentially, per DNA pulldown, 500 pmole of annealed oligos were diluted to 600 µL in DNA binding buffer (DBB: 1 M NaCl, 10 mM Tris pH 8.0, 1 mM EDTA, 0.05% NP40) and incubated with washed beads (10 uL Streptavidin Sepharose High performance bead slurry (GE Healthcare #17511301), washed once with PBS + 0.1% NP-40 and once with DBB) for 30 minutes at 4°C while rotating. After washing once with 1mL DBB and twice with 1 mL protein incubation buffer (PIB: 150 mM NaCl, 50 mM Tris pH 8.0, 0.25% NP40, 1 mM DTT with Complete protease inhibitors (CPI, Roche #4693159001)) the immobilized oligos on beads were combined with 500 µg nuclear extracts in a total volume of 600 µL PIB with 10 µg competitor DNA (5 µg poly-dIdC (Sigma #81349_500ug) and 5 µg poly-dAdt (Sigma #P0883_50UN)) and incubated for 90 minutes at 4°C while rotating. Beads

were washed three times with 1mL PIB and twice with 1 mL PBS. To elute proteins from the oligo probes, beads were resuspended in 20 uL 1x western blot protein sample buffer and incubated at 95°C for 15 minutes while shaking. The beads were spun down and the eluate was loaded on a protein gel. A 40 µg nuclear extract sample was prepared directly from the nuclear lysate as input sample for western blot.

Peptide treatment of AML cells

MUTZ3 or primary AMLs were cultured in medium as described above, plus MYBMIM or control peptide TG3 at indicated concentrations. For measuring viability of MUTZ3 or primary AMLs, cells were seeded in an opaque colored 96-well plate at 15.000 cells/well in a total volume of 100 µl medium containing MYBMIM or control peptide TG3 at indicated concentrations (20 µM MYBMIM or control peptide TG3 for primary AMLs). Cell viability was assessed 72 hours after treatment using CellTiter-Glo cell viability assay according to manufacturer's protocol (Promega). Luminescence was measured on the Victor X3 plate reader (Perkin Elmer). For protein lysates and ChIP experiments cells were cultured containing 20 µM MYBMIM or control peptide TG3 and harvested after 48 hours of peptide treatment.

Fluorescence in situ hybridization (FISH)

FISH was performed and reported according to standard protocols based on the International System of Human Cytogenetics Nomenclature (2016) [44]. *MECOM* FISH was performed according to the manufacturer's protocol using the *MECOM* t(3;3); inv(3)(3q26) triple-color probe (Cytocell, LPH-036).

Genome editing

The sgRNAs (Table S3) were either cloned into pLentiV2_U6-IT-mPgk-iRFP720 using BsmBI restriction sites, px330 using BbsI or were *in vitro* transcribed using the T7 promoter. Lentiviruses were prepared by transfecting 293T cells with lentiviral packaging constructs pSPAX2 / pMdelta2.G and sgRNA cloned into pLentiV2_U6-IT-mPgk-iRFP720. Transfections were performed using Fugene 6 (Promega) according to manufacturer's protocol. For *in vitro* transcribed sgRNAs oligo's containing the T7 promoter, target sequence and the Tail annealing sequence were annealed, filled in and transcribed using the Hi-scribe T7 kit (NEB). Turbo DNAse (Invitrogen) was added and sgRNAs were cleaned up using RNA clean&concentrator kit (Zymo). Concentration of sgRNAs was estimated using Qubit (Invitrogen). RNP complexes were formed incubating sgRNA and Cas9 (IDT) for 20-30 at RT before nucleofection using the Neon (Thermofischer) with buffer R using program 3. Genomic DNA was extracted 48 or 72 h after transfection using Quick Extract buffer (Epicenter) and checked for targeting by PCR using Q5 polymerase (NEB).

Pooled sgRNA Enhancer scanning

To design a high-resolution sgRNA library for the enhancer scan, we considered all possible sgRNA target sites containing a canonical Cas9 PAM site (NGG) on both strands of the minimal 18 kb translocated region. sgRNAs containing a G in positions 1-3 of the 20nt target site were trimmed at this position to favor 20-, 19- or 18-mers (in this order of priority) containing a natural G at the 5'end as previously described [45]. For all other sgRNAs, a G was added to the 5'end (resulting in a 21-mer). Subsequently, all sgRNAs showing (1) a high number of target sites in the human genome (>5 with no mismatch, or >20 with 1 mismatch), (2) a BsmBI site (interfering with cloning), or (3) a polyA signal (interfering with packaging) were filtered out. In addition, we added a number of negative controls (82 sgRNAs targeting the *AAVS1* region) as well as positive controls (5 sgRNAs targeting *EVI1* as well as 313 sgRNAs covering 5 kb of the breakpoint in MUTZ3 cells). The final library of 3239 sgRNAs (Table S1) was synthesized with overhangs for PCR amplification and cloning as one oligo pool (Twist Bioscience) and cloned into the lentiviral vector sgETN as previously described [45]. The pool of 3239 sgETN-sgRNAs was transduced in triplicate into MUTZ3-EVI1-GFP. For each replicate, a total of 120 million cells were infected with 3-4% transduction efficiency to ensure that each sgRNA is represented predominantly as a single lentiviral integration in >1000 cells. After neomycin drug selection (1 mg ml^{-1}) for 7 days, T0 samples were obtained (5 million cells per replicate), and cells were subsequently cultured in the presence of 1 µg ml^{-1} doxycycline (Dox). Culture medium was exchanged every 2 days. After 5 days (T5) and 7 days (T7), about 1 million sgRNA-expressing (GFP^{low}) cells were sorted for each replicate using a FACS AriaII (BD Biosciences). Genomic DNA from T0, T5 and T7 samples was isolated by two rounds of phenol extraction using PhaseLock tubes (5PRIME), followed by isopropanol precipitation. Deep-sequencing libraries were generated by PCR amplification of sgRNA guide strands using primers that tag the product with standard Illumina adapters and a 4 bp sample barcode in a 2 step-PCR protocol. For each sorted sample, all DNA was used as template in multiple parallel 50-µl PCR reactions, each containing 250-500 ng template, 1x AmpliTaq Gold buffer, 0.2 mM of each dNTP, 2 mM MgCl$_2$, 0.3 µM of each primer and 1U AmpliTaq Gold (Invitrogen), which were run using the following cycling parameters: 95 °C for 10 min; 28 cycles of 95 °C for 30 s, 52 °C for 45 s and 72 °C for 30 s; 72 °C for 7 min. PCR products (367 bp) were combined for each sample and Ampure purified. For the T0 samples and a DNA-pool sample the amount of input DNA necessary to get a 1000x coverage was used as input in the PCRs. For the second PCR 10ng of input was used per PCR using the following cycling parameters: 95 °C for 10 min; 8 cycles of 95 °C for 30 s, 57 °C for 45 s and 72 °C for 30 s; 72 °C for 7 min. PCR products (448 bp) were combined for each sample and Ampure-purified. Libraries were sequenced equimolarly on an Illumina HiSeq 2500 (Illumina) by the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria. Multiple experiments (different time points and sorted fractions) were sequenced

simultaneously, each identified by a unique barcode. Sequencing data were processed by converting unaligned BAM files into FASTA using bam2fastx. Experiment-specific barcodes (positions 7-10) were extracted together with the sgRNA sequence (positions 31-) into a new FASTA file, which was subsequently reverse-complemented with seqtk seq. Next, the barcodes were used to demultiplex the FASTA file into experiment-specific files with ngs-tools split-by-barcode, using parameters –s 4 –d 1, i.e. barcode size 4 and maximum 1 mismatch. For each of these files, we counted the number of identical sgRNA sequences with fastx_collapser and we assigned them to their known identifiers. These counts were employed for downstream data analysis. To provide a sufficient baseline for detecting sgRNA enrichment in experimental samples, we aimed to acquire >1000 reads per sgRNA in the sequenced sgRNA pool to compensate for variation in sgRNA representation inherent in the pooled plasmid preparation or introduced by PCR biases. Reads were normalized to the total number of library-specific reads per lane for each condition. To ensure a proper sgRNA representation in the initial plasmid pool, we used a cutoff of more than 10% average reads/sgRNA sequenced in the Plasmid-Pool (resulting in passing of 3050 out of 3239 sgRNAs). Enrichment analyses were performed using MAGeCK [46].

ChIP sequencing

$H_3K_{27}Ac$ and p300 ChIP-seq data from the inv(3) cell line MOLM1 as well as p300 ChIP-seq data from MUTZ3 were previously generated by our group and are available at ArrayExpress E-MTAB-2224 [12]. $H_3K_{27}Ac$ (Abcam, AB4729) ChIPs were performed according to the standard ChIP protocol from Upstate. ChIP with antibodies direct against MYB (Merck Millipore, clone1-1, #05-175) or p300 (Diagenode, #C15200211) were performed by first crosslinking for 45 minutes with DSG before formaldehyde crosslinking. ChIP samples were processed according to the Illumina TruSeq ChIP Sample Preparation Protocol (Illumina) or Diagenode Library V3 preparation protocol (Diagenode) and either sequenced single-end (1x 50 bp) on the HiSeq 2500 platform (Illumina) or paired-end (2x100 bp) on the Novaseq 6000 platform (Illumina). Briefly, reads were aligned to the human reference genome build hg19 with bowtie [47] for single-end runs and bowtie2 [48] for paired-end runs, and bigwig files were generated for visualization with bedtools genomecov [49] and UCSC bedGraphToBigWig [50]. Peaks were determined using the MACS2 program with default parameters [51]. The tracks were normalised per million reads (RPM) and visualized as genome browser profiles using the Fluff package [52].

Cut&Run

$H_3K_{27}Ac$ (Abcam, #AB4729) Cut&Run libraries for the MUTZ3 bulk and sorted fragments were generated with an input of 200.000 cells. The protocol described by the Henikoff group was used to generate these tracks [53], using a 0.04% Digitonin buffer and with the addition of cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche) and 1M Sodiumbutyrate (Sigma

Aldrich) to all the buffers. Isolation was done according to the standard Phenol Chloroform protocol. Cut&Run samples were processed according to the protocol described by the Fazzio group [54] and sequenced paired-end (2x100 bp) on the Novaseq 6000 platform (Illumina). Reads were aligned similarly to ChIP-seq.

ATAC sequencing

Open chromatin regions were mapped by the ATAC-seq method as described [55] with a modification in the lysis buffer (0.30 M sucrose, 10 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA, 0.1% NP40, 0.15 mM Spermine, 0.5 mM Spermidine, 2 mM 6AA) to reduce mitochondrial DNA contamination. ATAC-seq samples were sequenced paired-end (2x 50 bp) on the HiSeq 2500 platform (Illumina) and aligned against the human genome (hg19) with bowtie2, allowing for a maximum 2000 bp insert size. Mitochondrial reads and fragments with mapping quality below 10 were removed.

RNA sequencing

RNA was isolated either using Trizol or the Qiagen Allprep DNA/RNA kit and protocol (Qiagen, #80204). cDNA synthesis was done using the SuperScript II Reverse Transcriptase kit (Invitrogen). Quantitative real-time PCR was performed by using primers (Table S4) as described previously [15] on the 7500 Fast Real-time PCR System (Applied Biosystems). For RNA sequencing, sample libraries were prepped using 500 ng of input RNA according to the KAPA RNA HyperPrep Kit with RiboErase (HMR) (Roche) using Unique Dual Index adapters (Integrated DNA Technologies, Inc.). Amplified sample libraries were paired-end sequenced (2x100 bp) on the Novaseq 6000 platform (Illumina) and aligned against the human genome (hg19) using STAR version 2.5.4b. Salmon [56] was used to quantify expression of individual transcripts, which were subsequently aggregated to estimate gene-level abundances with tximport [57]. Human gene annotation derived from RefSeq [58] was downloaded from UCSC [59] (RefGene) as a GTF file. Transcript-level abundances were normalized to transcripts per million (TPM) for visualization.

Amplicon sequencing

For amplicon sequencing we used a PCR-based NGS library preparation method in combination with the TruSeq Custom Amplicon index kit (Illumina). The first PCR for target selection (Table S4) was performed using Q5 polymerase (NEB), the second nested PCR, to add the index-adapters, with KAPA HiFi HotStart Ready mix (KapaBiosystems). Libraries were sequenced paired-end (2x 250 bp) on the MiSeq platform (Illumina). Reads were trimmed with trimgalore [60] to remove low-quality bases and adapters, and subsequently aligned to the human reference genome build hg19 with BBMap [61] allowing for 1000 bp indels. Mutations introduced by genome editing were analysed and visualised using CRISPResso2 [62]. Mutated sequences consisting of up to 5% of sequenced reads were next analysed for differential binding with CIS- BP [63].

Author contributions

Acknowledgments

Declaration of Interests

The authors declare no competing financial interests. A patent application related to MYBMIM has been submitted by A.K. to the U.S. Patent and Trademark Office entitled ''Agents and methods for treating CREB binding protein-dependent cancers'' (application PCT/US2017/059579).

REFERENCES

1. Garraway, L.A. and E.S. Lander, *Lessons from the cancer genome.* Cell, 2013. **153**(1): p. 17-37.

2. Vogelstein, B., et al., *Cancer genome landscapes.* Science, 2013. **339**(6127): p. 1546-58.

3. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

4. Khurana, E., et al., *Role of non-coding sequence variants in cancer.* Nat Rev Genet, 2016. **17**(2): p. 93-108.

5. Zhu, H., et al., *Candidate Cancer Driver Mutations in Distal Regulatory Elements and Long-Range Chromatin Interaction Networks.* Mol Cell, 2020. **77**(6): p. 1307-1321 e10.

6. Rahman, S. and M.R. Mansour, *The role of noncoding mutations in blood cancers.* Dis Model Mech, 2019. **12**(11).

7. Fredriksson, N.J., et al., *Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types.* Nat Genet, 2014. **46**(12): p. 1258-63.

8. Bresnick, E.H. and K.D. Johnson, *Blood disease-causing and -suppressing transcriptional enhancers: general principles and GATA2 mechanisms.* Blood Adv, 2019. **3**(13): p. 2045-2056.

9. Mansour, M.R., et al., *Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element.* Science, 2014. **346**(6215): p. 1373-7.

10. Herranz, D., et al., *A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia.* Nat Med, 2014. **20**(10): p. 1130-7.

11. Hnisz, D., et al., *Activation of proto-oncogenes by disruption of chromosome neighborhoods.* Science, 2016. **351**(6280): p. 1454-1458.

12. Groschel, S., et al., *A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia.* Cell, 2014. **157**(2): p. 369-381.

13. Northcott, P.A., et al., *Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma.* Nature, 2014. **511**(7510): p. 428-34.

14. Affer, M., et al., *Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma.* Leukemia, 2014. **28**(8): p. 1725-1735.

15. Ottema, S., et al., *Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia.* Blood, 2020. **136**(2): p. 224-234.

16. Barjesteh van Waalwijk van Doorn-Khosrovani, S., et al., *High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients.* Blood, 2003. **101**(3): p. 837-45.

17. Lugthart, S., et al., *Clinical, Molecular, and Prognostic Significance of WHO Type inv(3)(q21q26.2)/t(3;3) (q21;q26.2) and Various Other 3q Abnormalities in Acute Myeloid Leukemia.* Journal of Clinical Oncology, 2010. **28**(24): p. 3890-3898.

18. Lugthart, S., et al., *High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated.* Blood, 2008. **111**(8): p. 4329-4337.

19. Morishita, K., et al., *Activation of EVI1 gene expression in human acute myelogenous leukemias by translocations spanning 300-400 kilobases on chromosome band 3q26.* Proceedings of the National Academy of Sciences, 1992. **89**(9): p. 3937-3941.

20. Yamazaki, H., et al., *A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression.* Cancer Cell, 2014. **25**(4): p. 415-27.

21. Grass, J.A., et al., *Distinct functions of dispersed GATA factor complexes at an endogenous gene locus.* Mol Cell Biol, 2006. **26**(19): p. 7056-67.

22. Johnson, K.D., et al., *Cis-regulatory mechanisms governing stem and progenitor cell transitions.* Sci Adv, 2015. **1**(8): p. e1500503.

23. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.* Cell Stem Cell, 2010. **7**(4): p. 532-44.

24. Beck, D., et al., *Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes.* Blood, 2013. **122**(14): p. e12-22.

25. Ramaswamy, K., et al., *Peptidomimetic blockade of MYB in acute myeloid leukemia.* Nat Commun, 2018. **9**(1): p. 110.

26. Loven, J., et al., *Selective inhibition of tumor oncogenes by disruption of super-enhancers.* Cell, 2013. **153**(2): p. 320-34.

27. Drier, Y., et al., *An oncogenic MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma.* Nat Genet, 2016. **48**(3): p. 265-72.

28. Sakamoto, H., et al., *Proper levels of c-Myb are discretely defined at distinct steps of hematopoietic cell development.* Blood, 2006. **108**(3): p. 896-903.

29. Beug, H., et al., *Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation.* Cell, 1979. **18**(2): p. 375-90.

30. Weston, K. and J.M. Bishop, *Transcriptional activation by the v-myb oncogene and its cellular progenitor, c-myb.* Cell, 1989. **58**(1): p. 85-93.

31. Rahman, S., et al., *Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia.* Blood, 2017. **129**(24): p. 3221-3226.

32. Nguyen, N., et al., *Myb expression is critical for myeloid leukemia development induced by Setbp1 activation.* Oncotarget, 2016. **7**(52): p. 86300-86312.

33. Ramsay, R.G. and T.J. Gonda, *MYB function in normal and cancer cells.* Nat Rev Cancer, 2008. **8**(7): p. 523-34.

34. Zuber, J., et al., *An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance.* Genes Dev, 2011. **25**(15): p. 1628-40.

35. Kasper, L.H., et al., *A transcription-factor-binding surface of coactivator p300 is required for haematopoiesis.* Nature, 2002. **419**(6908): p. 738-43.

36. Sandberg, M.L., et al., *c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation.* Dev Cell, 2005. **8**(2): p. 153-66.

37. Best, J.L., et al., *Identification of small-molecule antagonists that inhibit an activator: coactivator interaction.* Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17622-7.

38. Uttarkar, S., et al., *Naphthol AS-E Phosphate Inhibits the Activity of the Transcription Factor Myb by Blocking the Interaction with the KIX Domain of the Coactivator p300.* Mol Cancer Ther, 2015. **14**(6): p. 1276-85.

39. Uttarkar, S., et al., *Small-Molecule Disruption of the Myb/p300 Cooperation Targets Acute Myeloid Leukemia Cells.* Mol Cancer Ther, 2016. **15**(12): p. 2905-2915.

40. Walf-Vorderwulbecke, V., et al., *Targeting acute myeloid leukemia by drug-induced c-MYB degradation.* Leukemia, 2018. **32**(4): p. 882-889.

41. Roe, J.S., et al., *BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia.* Mol Cell, 2015. **58**(6): p. 1028-39.

42. Diffner, E., et al., *Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia.* Blood, 2013. **121**(12): p. 2289-300.

43. Karemaker, I.D. and M. Vermeulen, *ZBTB2 reads unmethylated CpG island promoters and regulates embryonic stem cell differentiation.* EMBO Rep, 2018. **19**(4).

44. McGowan-Jordan, J., A. Simons, and M. Schmid, *ISCN : an international system for human cytogenomic nomenclature (2016).* 2016: Basel ; New York : Karger.

45. Michlits, G., et al., *Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles.* Nat Methods, 2020. **17**(7): p. 708-716.

46. Li, W., et al., *MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens.* Genome Biology, 2014. **15**(12): p. 554.

47. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

48. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.

49. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

50. Kent, W.J., et al., *BigWig and BigBed: enabling browsing of large distributed datasets.* Bioinformatics, 2010. **26**(17): p. 2204-7.

51. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

52. Georgiou, G. and S.J. van Heeringen, *fluff: exploratory analysis and visualization of high-throughput sequencing data.* PeerJ, 2016. **4**: p. e2209.

53. Skene, P.J., J.G. Henikoff, and S. Henikoff, *Targeted in situ genome-wide profiling with high efficiency for low cell numbers.* Nat Protoc, 2018. **13**(5): p. 1006-1019.

54. Hainer, S.J., et al., *Profiling of Pluripotency Factors in Single Cells and Early Embryos.* Cell, 2019. **177**(5): p. 1319-1329 e11.

55. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.

56. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.

57. Soneson, C., M.I. Love, and M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.* F1000Res, 2015. **4**: p. 1521.

58. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.

59. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.

60. Krueger, F., *Trim Galore*. 2012.

61. Bushnell, B., *BBMap short-read aligner, and other bioinformatics tools*. 2016.

62. Clement, K., et al., *CRISPResso2 provides accurate and rapid genome editing sequence analysis.* Nat Biotechnol, 2019. **37**(3): p. 224-226.

63. Weirauch, M.T., et al., *Determination and inference of eukaryotic transcription factor sequence specificity.* Cell, 2014. **158**(6): p. 1431-1443.

SUPPLEMENTARY REFERENCES

1   Beck, D., Thoms, J.A., Perera, D., Schütte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Göttgens, B._, et al._ (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. Blood _122_, e12-22.

2   Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K._, et al._ (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell _157_, 369-381.

4

De-repression of retro-elements in acute myeloid leukemia with 3q aberrations

Jagoda Mika[1], Sophie Ottema[2], Leonie Smeenk[2], Sandra Kiehlmeier[1], Sabrina Kruse[1], Judith Müller[1], Sabrina Schweiggert[1], Carl Herrmann[3], Mathijs Sanders[2], Ruud Delwel[2], Stefan Gröschel[1,4]

[1]Molecular Leukemogenesis Group, German Cancer Research Center, Heidelberg, Germany, [2]Department of Hematology, Erasmus University Medical Center, Oncode Institute, Rotterdam, The Netherlands, [3]Health Data Science Unit, Medical Faculty Heidelberg and BioQuant, Germany, [4]Internal Medicine V, Heidelberg University Hospital, Heidelberg, Germany, [5]Oncology Center Worms, Germany

Acute myeloid leukemia (AML) with chromosomal rearrangements inv(3)/t(3;3) (3q-AML) is a rare but highly fatal subtype of leukemia. It is characterized by an aberrant transcription of the proto-oncogene *EVI1* (Ecotropic Viral Integration Site 1, *MECOM*) as a result of the chromosomal 3q21q26 rearrangements that lead to the relocation of a master *GATA2* hematopoietic enhancer (*G2DHE*) to the *EVI1* locus and deregulation of both genes.[1,2] To date, little is known about what triggers chromosomal rearrangements in 3q-AML that ultimately lead to the deregulation of *EVI1* via the repositioning of *G2DHE.* However, recent studies have shown evidence for the involvement of endogenous transposable elements of RNA family (retroelements, REs) in the formation of complex chromosomal aberrations, including translocations, large-scale duplications and amplifications through retrotransposition across different cancer entities.[3] Furthermore, hypomethylation of retroelements have been linked to their pathogenic mobility in epithelial tumors.[4-6]

Here we present the results of functional genomics analysis of a cohort of 3q-AML patients. Based on our data, we hypothesized that breakpoint-associated retroelements (breakpoint-REs) could play an important regulatory and activating role in this AML subtype. Therefore, we performed an array of *in vitro* studies using CRISPR-Cas9 approach to dissect their role in 3q-AML.

Targeted chromosome 3q-capture sequencing of 3q-AML patient samples and cell lines with *EVI1* overexpression previously revealed a characteristic 3q21q26 pattern of patient-specific breakpoints, demarcating a leukemogenic *EVI1*-activating super-enhancer that is found uniquely in inv(3)/t(3;3) AML and contains *G2DHE*.[1] In order to identify a commonality between breakpoints relevant for super-enhancer formation, we reanalyzed 3q-capture sequencing data and found that in 38 of 41 samples, chromosomal breakpoints at 3q21.3 and 3q26.2 mapped to sequences of retroelements, including LINE (long interspersed element), SINE (short interspersed element) and LTR (long terminal repeat) (**Fig. 1A**). Of note, RNA sequencing (RNA-Seq) of 3q-AML patients revealed a characteristic RNA readthrough spanning the large super-enhancer region at 3q21.3 (**Fig. 1B** top panel). Similar enhancer RNA (eRNA) signature was observed in non-3q-rearranged AML cases (**Fig. 1B**, top panel, **Fig. S3**), indicating active *G2DHE* regulating *GATA2* in its native environment. In 3q-AML, however, the RNA readthrough frequently originated at 3q21.3 breakpoint sites, extending beyond the super-enhancer region. Additionally, allele-specific bisulfite amplicon sequencing performed on selected AML cases revealed focal demethylation of CpG sites around the chromosomal breakpoints exclusively on the rearranged allele, whereas the intact allele in 3q-AML and both alleles in non-rearranged leukemic cell lines did not show any hypomethylation pattern (**Fig. 1B**, bottom panel). Focal hypomethylation around breakpoints on the rearranged allele could be the consequence of chromosomal rearrangements and super-enhancer-related epigenetic reprogramming, including the deposition of active chromatin marks and physical interaction between the *EVI1* promoter and *G2DHE*.[1]

Based on our RNA-Seq and bisulfite sequencing data, we investigated whether

derepression of breakpoint-REs could possibly represent a priming event for an enhancer rearrangement by relaxation of the local chromatin compaction and may play a role in the ectopic activation of *EVI1* by the super-enhancer. To test this hypothesis, we performed a CRISPR-Cas9 gene editing experiment using a homology-directed repair (HDR) template to insert selected 3q21.3 breakpoint-RE sequences or *G2DHE* in the vicinity of the *EVI1* locus in the *EVI1*-positive myeloid leukemia reporter cell line K562 that does not harbor inv(3)/t(3;3) rearrangements (**Fig. 2A**, top panel) (Ottema *et al*., 2021, under review). The presence of a *T2A-eGFP* fusion sequence inserted downstream of *EVI1* allows for correlation of the EVI1 expression with the synchronously expressed GFP. The parental reporter cell line is tolerant of increased EVI1 levels given that its baseline expression is already increased in K562. The insertion sequences were derived from 3q21.3 breakpoints of two leukemia cases: AML 3071, a patient with inv(3) AML and MOLM-1, a near-triploid myeloid leukemia cell line harboring two chromosome 3 alleles with inv(3) (**Fig 2A**, bottom panel).[7] The HDR templates were inserted in the corresponding 3q26.2 breakpoint loci as found in AML 3071 and MOLM-1, that is downstream of *EVI1* and within the last *EVI1* intron, respectively.

Single-cell clones validated by PCR and Sanger sequencing (**Fig S1A**) harboring the ectopic *G2DHE* showed a shift in the GFP fluorescence indicating successful *EVI1* activation, whereas clones with 3q21.3 breakpoint-RE sequences showed no change in the GFP signal compared with untreated cells (**Fig 2B**). Furthermore, single-cell clones were analyzed by quantitative PCR (qPCR) and western blot, which showed results consistent with the flow cytometry analysis (**Fig 2C** and **2D**, respectively), suggesting that the ectopic activation of *EVI1* occurs via *G2DHE*, whereas breakpoint-REs themselves are not sufficient to induce *EVI1* transcriptional activation in the K562 reporter cell line.

In order to further dissect a potential regulatory role of breakpoint-REs in 3q-AML, a reciprocal experimental CRISPR-Cas9 approach was applied to delete the original breakpoint-REs in MOLM-1 and UCSD-AML1, the latter being a t(3;3) AML cell line. We expressed pairs of single guide RNAs (sgRNAs) in stably Cas9-expressing cells to induce a segmental deletion of a fragment containing either the inverted (MOLM-1) or translocated (UCSD-AML1) breakpoint-RE at 3q26.2 on the rearranged alleles located within the last *EVI1* intron in MOLM-1, and upstream of the *EVI1* promoter in UCSD-AML1 (**Fig 3A**). Targeting on the non-rearranged allele was expected to result only in generation of indels at the 3q21.3 and 3q26.2 site but not segmental RE deletions. In total, two MOLM-1 and six UCSD-AML1 clones harboring the desired deletion validated by PCR and Sanger sequencing (**Fig S1B**) were derived successfully from single cells. Together with the nontargeting control (NTC) clones (targeted with sgRNAs against mCherry and eGFP) and the wild-type (WT) cell line, we performed phenotypic analysis of obtained deletion clones. We observed no differences in proliferation between deletion and control samples (**Fig 3B**). Slightly reduced EVI1 expression on mRNA and protein level was observed exclusively in the MOLM-1 deletion clones (**Fig 3C,** left panel).

To identify potential genome-wide effects of CRISPR-Cas9-induced RE deletion, we performed genomic and epigenomic analyses of the MOLM-1 and UCSD-AML1 deletion and control clones using circularized chromatin conformation capture sequencing (4C-Seq) and chromatin immunoprecipitation followed by sequencing (ChIP-Seq). Neither 4C-Seq nor ChIP-Seq revealed an impact of RE deletion on the interaction frequency of *G2DHE* with the *EVI1* promoter or on the deposition of active chromatin marks, such as H3K27ac and H3K4me3, in any of the two edited cell lines (results for MOLM-1 shown in **Fig S2A**), making it unlikely that the expression changes in EVI1 are caused by changes in the regulatory function of *G2DHE*.

Since the reduction of EVI1 expression was specific only to the MOLM-1, but not UCSD-AML1 deletion clones, we speculated that this effect might be due to other features present in the sequence deleted in MOLM-1, rather than the consequence of the breakpoint-RE-deletion. To this end, we reanalyzed the deleted sequences in the two cell lines. The original 3q21.3 fragment overlapping with *LINC01565* relocated to the 3q26.2 site and deleted in the MOLM-1 clones displays high degree of conservation (**Fig. 3A,** left panel), which might indicate functional importance of this sequence. ENCODE transcription factor (TF) ChIP data generated in K562 cells revealed a plethora of TFs binding to this region (**Fig. S2B**). Some of these TFs have predicted binding sites within the *G2DHE* sequence, including IKZF1, MAX, TAL1 and MAZ (Kiehlmeier *et al*., 2021, under review). Furthermore, EVI1 has been functionally linked to some of these TFs, including MTA1/2, HDAC1/2 and GATAD2B. All these proteins belong to the nucleosome remodeling and deacetylase (NuRD) complex[8], which was shown to specifically interact with the MDS-EVI1 (PRDM3) but not the EVI1 protein[9], while both EVI1 protein isoforms were shown to interact with HDAC1.[10] Contrary to MOLM-1, no TF binding in the region deleted in the UCSD-AML clones was found (**Fig. S2B**).

Taken together, the observed reduction in EVI1 expression upon breakpoint-RE deletion in MOLM-1 is more likely the consequence of TF binding loss within the *G2DHE* super-enhancer structure. However, we could not observe any specific pattern of TF binding sites commonly relocated to the *EVI1* locus in 3q-AML patient and cell line data. Meanwhile, we conclude that breakpoint-REs are not essential for the regulation and maintenance of *EVI1* expression in 3q-AML.

In summary, our data show that REs are highly enriched at inv(3)/t(3;3) breakpoint sites in AML and represent a source of genomic vulnerability without providing additional regulatory or activating signal for *EVI1* in this leukemia subtype, as evidenced by CRISPR-Cas9 editing experiments in 3q-rearranged cell lines. However, we cannot exclude the involvement of full-length source RE sequences in the formation of 3q rearrangements in earlier stages of malignant transformation. Since many retrotransposition-competent REs often become truncated or undergo internal reshuffling upon insertion in a new genomic location[11,12], the presence of resulting chimeric breakpoint-RE sequences would bear no functional consequences, which stands in line with the lack of effects observed upon

CRISPR-Cas9 editing of breakpoint-RE fusion sequences in our cell line models.

Additionally, the data from the K562 reporter cell line provide an orthogonal confirmation of the minimal *G2DHE* being sufficient for *EVI1* transcriptional activation.[1]
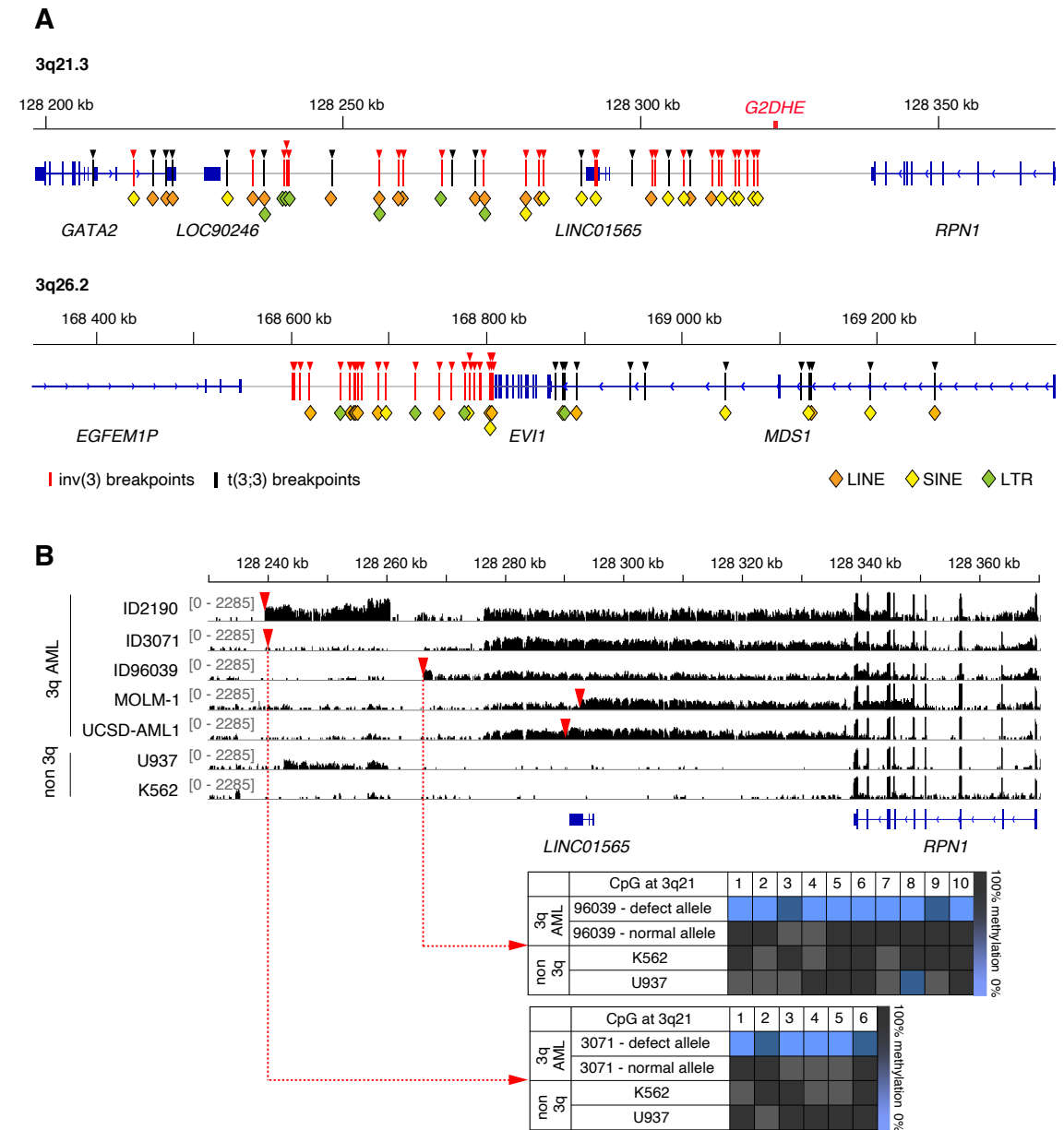
Fig. 1. Chromosomal breakpoints in AML inv(3)/t(3;3) (3q-AML) are enriched in retroelements
(**A**) 3q-capture sequencing revealed a characteristic breakpoint pattern (red and black arrowheads) at 3q21.3 (upper tracks) and 3q26.2 (lower tracks). At 3q21.3, a breakpoint-free region downstream (left) of *RPN1* was identified as a commonly translocated segment containing *G2DHE*. At 3q26.2, breakpoints of inversion cases map exclusively downstream (left) or within the last *EVI1* intron, whereas translocation cases have breakpoints upstream of *EVI1*. Color-coded diamonds indicate the position of retroelements: LINE, SINE or LTRs, annotated by RepeatMasker. (**B**) RNA sequencing of 3q-AML samples revealed a characteristic RNA readthrough signature at 3q21.3. Bisulfite amplicon sequencing on representative 3q-AML samples showed focal hypomethylation of breakpoint regions (red arrowheads) on the rearranged allele but not on the normal allele and non-3q samples.

Fig. 2. Ectopic activation of *EVI1* occurs via *G2DHE*, whereas breakpoint-associated retroelements do not display activating potential in the K562 cells

(**A**) Experimental strategy of CRISPR-Cas9-mediated genomic insertions. Donor templates containing either *G2DHE* (736 bp, red boxes) or 3q21.3 breakpoint-REs from selected 3q AML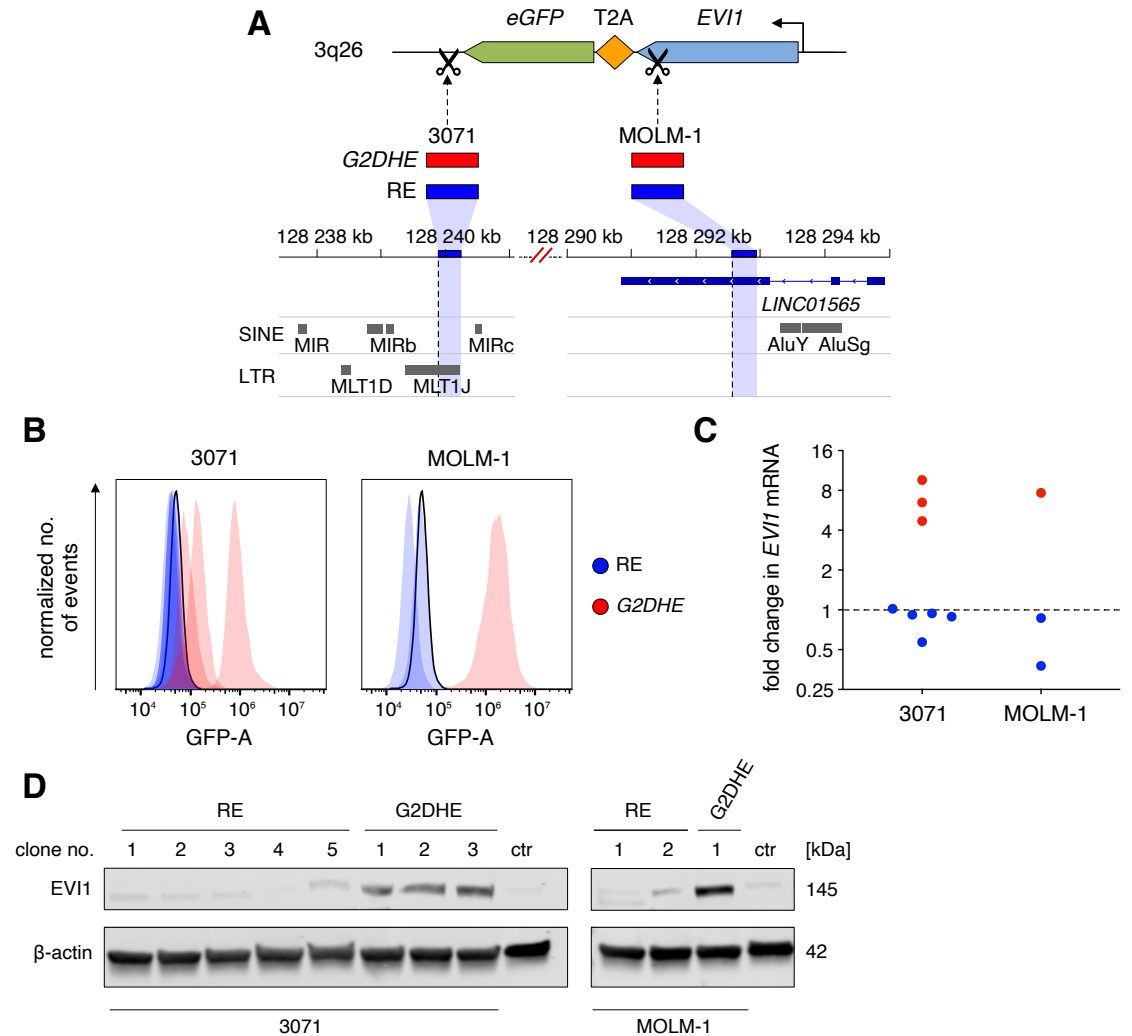 cases (3071: 353 bp and MOLM-1: 280 bp, blue boxes) were inserted in the non-3q K562 reporter cell line (K562 *eGFP-T2A-EVI1*) using CRISPR-Cas9 downstream of *EVI1* and within the last *EVI1* intron. A genomic view shows the origin of 3q21.3 breakpoint sequences used in CRISPR experiments, dashed lines indicate 3q21.3 breakpoints in 3071 (left) and MOLM-1 (right). 3071 HDR template contains a part of MLT1J LTR element. *G2DHE* (red) and 3q21.3 breakpoint-RE samples (blue) are consistently colored throughout the figure. (**B**) Flow cytometry on the K562 *eGFP-T2A-EVI1* single clones bearing the desired *G2DHE* or 3q21.3 breakpoint insertions. Peaks corresponding to the GFP signal from single clones targeted at the same region are presented together on one graph, with peaks for untreated cells shown as a black outline. (**C**) Quantitative PCR (qPCR) analysis of the *EVI1* mRNA levels in single clones shown in (**B**), relative to *HMBS* and normalized to the untreated *eGFP-T2A-EVI1* cells. (**D**) Representative western blot of the full-length EVI1 isoform from the clones and untreated cells (ctr) shown in (**B**) and (**C**). Data shown in (**C**) are means of three technical replicates from one independent experiment. RE: breakpoint-associated retroelement.
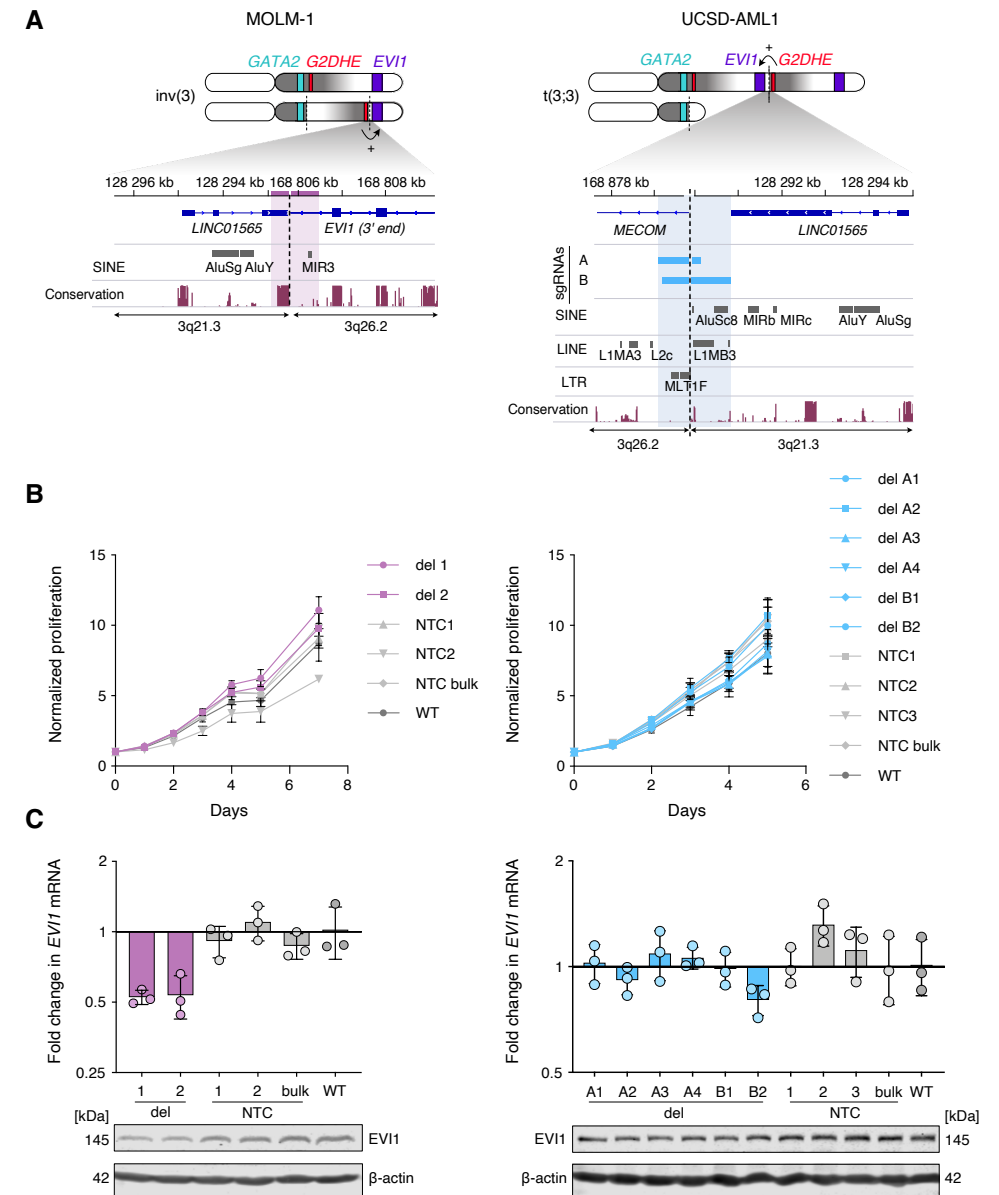
Fig. 3. MOLM-1 single clones harboring the deletion of breakpoint-associated retroelements exhibit no phenotypic changes but show reduced *EVI1* expression

(**A**) Illustration of the CRISPR-Cas9 mediated segmental deletion surrounding the breakpoint on the rearranged allele of chromosome 3 in MOLM-1 (left) and UCSD-AML1 (right). The deleted regions (MOLM-1: purple, UCSD-AML1: blue) encompass a MIR3 SINE element in MOLM-1, and representatives of all three RE subclasses in UCSD-AML1. Normalized proliferation (**B**) of deletion (del), nontargeting control (NTC) clones, NTC bulk of cells and wild-type (WT) cell line. (**C**) qPCR analysis of *EVI1* mRNA levels (top panel) and representative western blot (bottom panel) of the full-length EVI1 isoform from samples shown in (**B**). Data shown in (**B**) and (**C**, top panel) are mean ± SD from three independent experiments. Conservation across vertebrate genomes is taken from phastCons.[13] Supplementary figure legends

Author contributions

J.Mi., S.O., L.S., S.Ki., J.Mü., S.S. and S.G. performed experiments; S.O. and R.D. provided RNA-seq data; J.Mi., S.Kr., M.S. and C.H. analyzed results; J.Mi. made the figures; J.Mi. and S.G. designed the experiments; J.Mi. and S.G. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing financial interests.

REFERENCES

1. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 2014;157(2):369–381.

2. Yamazaki H, Suzuki M, Otsuki A, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. Cancer Cell 2014;25(4):415–427.

3. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet 2020;52(3):306–319.

4. Tubio JMC, Li Y, Ju YS, et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science 2014;345(6196):Article 1251343.

5. Lee E, Iskow R, Yang L, et al. Landscape of somatic retrotransposition in human cancers. Science 2012;337(6097):967–971.

6. Cajuso T, Sulo P, Tanskanen T, et al. Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. Nat Commun 2019;10(1):4022.

7. Matsuo Y, Adachi T, Tsubota T, Imanishi J, Minowada J. Establishment and characterization of a novel megakaryoblastic cell line, MOLM-1, from a patient with chronic myelogenous leukemia. Hum Cell 1991;4(3):261–264.

8. Torrado M, Low JKK, Silva APG, et al. Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. FEBS J 2017;284(24):4216–4232.

9. Ivanochko D, Halabelian L, Henderson E, et al. Direct interaction between the PRDM3 and PRDM16 tumor suppressors and the NuRD chromatin remodeling complex. Nucleic Acids Res 2019;47(3):1225–1238.

10. Vinatzer U, Taplick J, Seiser C, Fonatsch C, Wieser R. The leukaemia-associated transcription factors EVI-1 and MDS1/EVI1 repress transcription and interact with histone deacetylase. Br J Haematol 2001;114(3):566–573.

11. Sassaman DM, Dombroski BA, Moran J V, et al. Many human L1 elements are capable of retrotransposition. Nat Genet 1997;16(1):37–43.

12. Brouha B, Badge RM, Farley AH, et al. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci 2003;100(9):5280–5285.

13. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15(8):1034–1050.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY FIGURES
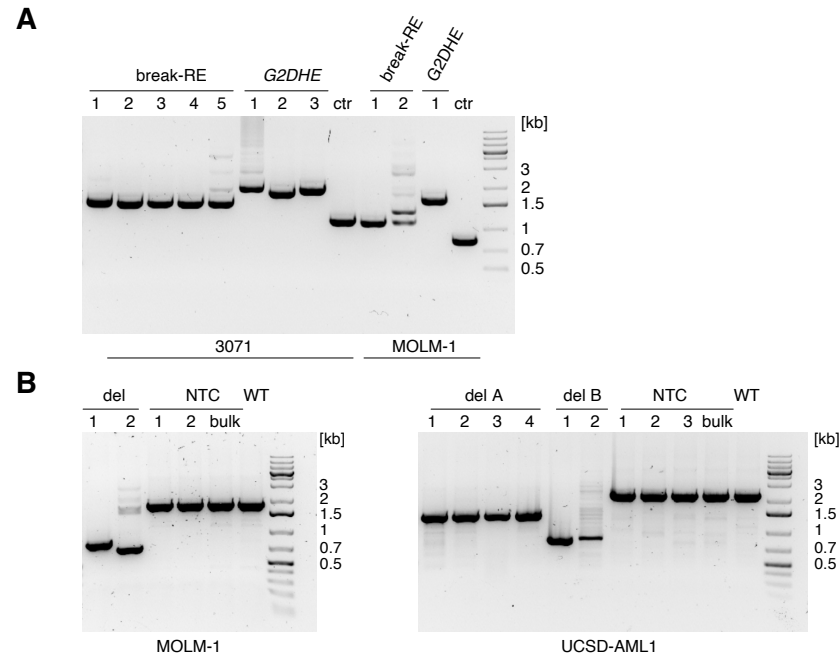
Figure S1. PCR validation of expanded single clones upon CRISPR-Cas9 targeting.
(**A**) PCR spanning the insertion sites in K562 shows successful *knock-in* of desired sequences (break-RE: breakpoint-RE) by the presence of higher-running bands on an agarose gel in respect to the control bands. (**B**) Rearranged allele-specific PCR of the MOLM-1 and UCSD-AML1 deletion clones (del, del A, del B) confirms the presence of desired deletions with smaller amplicons in comparison to control samples (NTC: nontargeting control, WT: wild-type). Number over each lane corresponds to a different single clone within a given condition.

Figure S2. Deletion of the breakpoint-RE does not affect neither the chromatin landscape nor the chromatin composition in the MOLM-1 deletion clones, TF binding sites overlapping with the deleted 3q21 region in MOLM-1
(**A**) Deletion of the breakpoint-RE affects neither the chromatin landscape nor the chromatin composition in the MOLM-1 deletion clones. 4C-Seq (top six tracks) and H3K4me3 and H3K27ac ChIP-Seq peaks (bottom Fig eight tracks) of selected MOLM-1 deletion (del) and control (NTC, WT) samples showing the 3q21 region around *G2DHE* (left) and the 3q26 region around *EVI1* promoter (right), which was used as a viewpoint in 4C-Seq. Chromosomal breakpoints in MOLM-1 are indicated with purple dashed lines. (**B**) TF binding sites overlap with the deleted 3q21 region in MOLM-1, but not UCSD-AML1. The dashed line indicates the position of the chromosomal breakpoints and the deleted 3q21 fragments are colored (MOLM-1: purple, UCSD-AML1: blue). No overlap of TF binding sites was found at the 3q26 breakpoint region in any of the two cell lines. TF ChIP-Seq data from K562 were taken from ENCODE 3 (grey-black) and ENCODE 2012 (blue) dataset.[1,2] The darkness of the boxes is proportional to the ChIP signal strength.
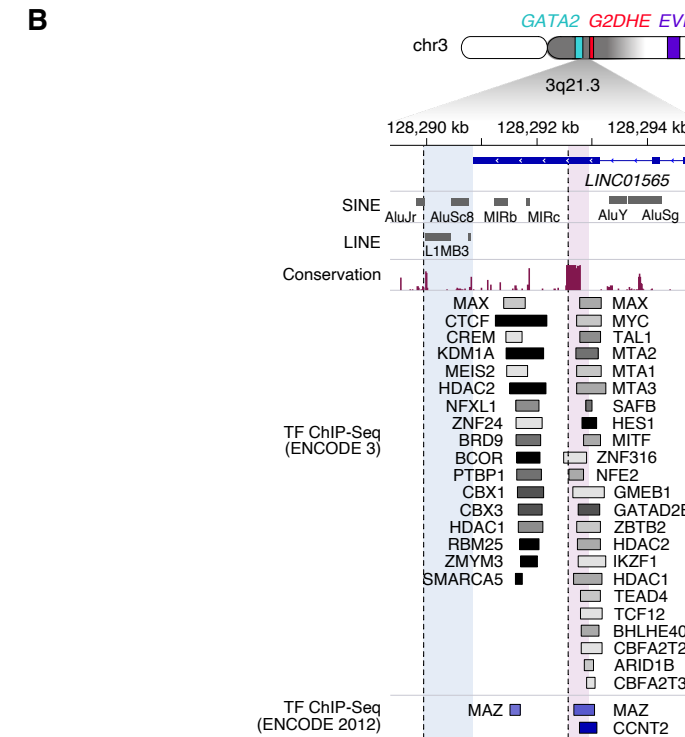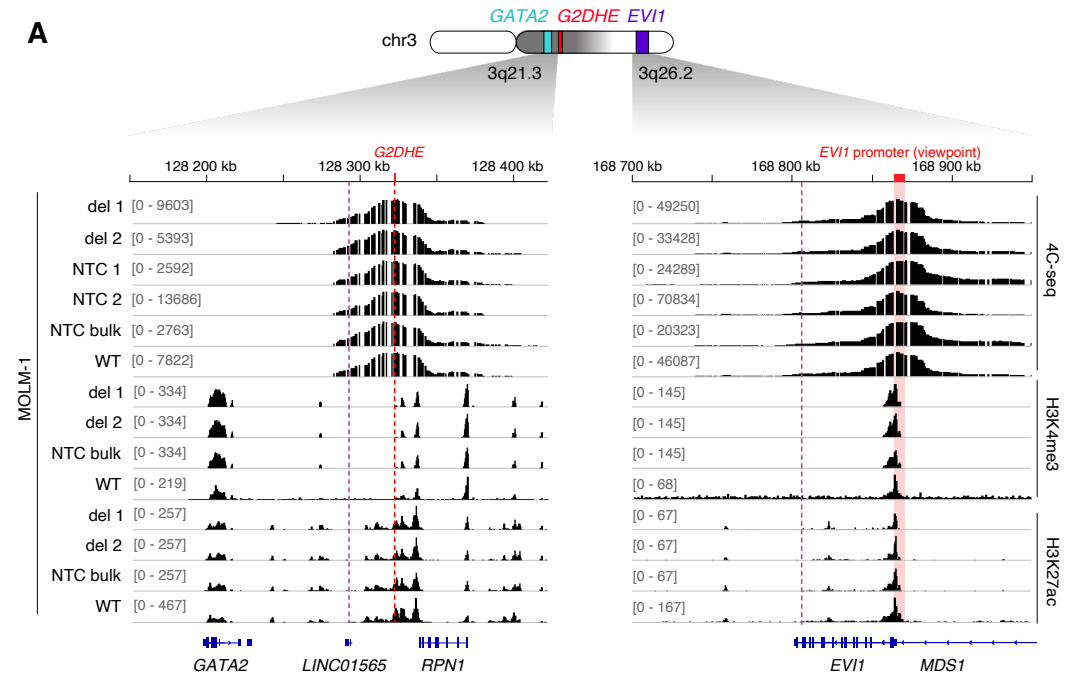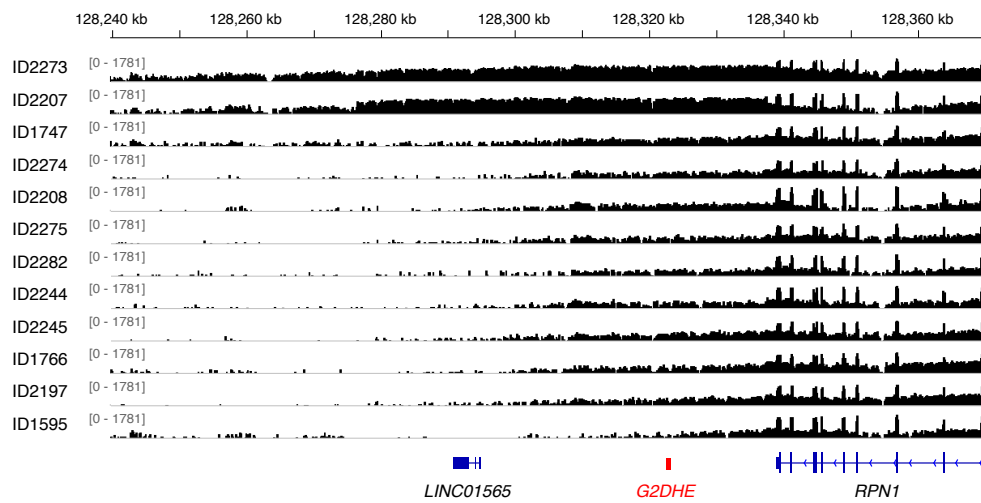
Figure S3. RNA-sequencing tracks from a selection of non 3q-AML patients showing the 3q21.3 region around *G2DHE*. Low-level RNA readthrough spanning the *G2DHE* is observed in these patients. Tracks are shown in a logarithmic scale. Sequencing data has been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under accession number EGAS00001004684 (Mulet-Lazaro *et al.*, 2021, under review).

SUPPLEMENTARY METHODS

Plasmids and cloning

pLKO5d.SFFV.hSpCas9.P2A.BSD vector (a gift from D. Heckl, University of Halle-Wittenberg, Halle (Salle), Germany) was used to generate cell lines stably expressing Cas9 with a lentiviral approach.

All sgRNAs used in this study were designed using the online sgRNA design tool (http://crispor.tefor.net/).[3] sgRNA oligonucleotides (Supplementary table 1) were cloned into lentiCRISPRv2 vector, a gift from F. Zhang (Addgene plasmid #52961). Resulting constructs were used as donors of the sgRNA cassettes and shuttled into pLKO.p-empty, a modular vector system originally created for simultaneous expression of multiple shRNA cassettes (a gift from P. Seelheim, University of Virginia, Charlottesville, VA, USA, described previously[4]), with following modifications: sgRNA cassettes were amplified by PCR from lentiCRISPR v2 using cloning primers (Supplementary table 1) and subsequently digested with MfeI and Bst1107I. Obtained fragments were sequentially ligated into pLKO.P-empty digested with EcoRI and Bst1107I, resulting in a pLKO.P plasmid with two sgRNA cassettes used to induce segmental deletions around the breakpoints in MOLM-1.

In vitro transcription of sgRNA

Double-stranded DNA templates for *in vitro* transcription were generated by PCR amplification with an sgRNA-specific forward primer containing T7 promoter, and a universal reverse primer binding to the sgRNA scaffold from lentiCRISPRv2. Purified PCR product was subsequently used in T7 *in vitro* transcription with HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). Resulting sgRNAs were purified using RNA Clean & Concentrator Kit (Zymo Research), quantified by Qubit (Invitrogen) and verified using RNA ScreenTape on 2200 TapeStation (Agilent). Primers and sgRNA sequences can be found in Supplementary table 1.

Generation of homology-directed repair (HDR) donor templates

Linear double stranded HDR donor templates were designed according to the protocol published previously[5] with modifications. Desired insert sequences with or without corresponding 35-bp homology arms flanked by HindIII and SalI restriction sites were ordered in a form of double-stranded GenParts (GenScript), digested with the aforementioned enzymes, ligated into pUC19 vector and transformed. Obtained colonies were verified by Sanger sequencing. Donor templates were generated by PCR amplification from pUC19 constructs with primer pairs listed in Supplementary table 1, including 35-bp homology arms as 5' overhangs where required. At least eight PCR reactions per donor template were pooled, purified using QIAquick PCR Purification kit (Qiagen) and concentrated by evaporation in a SpeedVac (Thermo Scientific) to a target concentration of 3-5 μM.

Cell lines and cell culture

Cell lines were cultured using the following media; MOLM-1: RPMI 1640 (Thermo Fisher Scientific), 20% FBS (Biochrom AG), HEK293T: DMEM (Thermo Fisher Scientific), 10% FBS. All media contained 1% penicillin/streptomycin mix (Thermo Fisher Scientific). All cell lines were cultured according to DSMZ recommendations.

Lentivirus packaging and transduction

Lentivirus was produced by co-transfection of HEK293T cells with Cas9 or sgRNA expression plasmid, pMD2.G and psPAX2 (Addgene plasmids #12259 and #12260, gifts from D. Trono) according to standard procedures. For lentiviral transduction, cells were seeded in 6-well plates (1x10^6 cells/well) and 0.4 ml of viral preparation was added to 1.6 ml medium with final concentration of 1 μg/ml polybrene (Merck). The cells were selected with 7.5 or 12 μg/ml blasticidin or 1 μg/ml puromycin (InvivoGen) beginning 48 hours after infection. To obtain MOLM-1 single-cell clones bearing retroelement deletions, 2 000 cells infected with pLKO.P sgRNA lentivirus after antibiotic selection were seeded in methylcellulose MethoCult H4230 (StemCell Technologies) supplemented with 20% conditioned medium from 5637 cells and 10 ng/ml GM-CSF where indicated. 10-14 days after seeding the colonies were transferred to a 96-well plate and expanded.

CRISPR-mediated *knock-in* of breakpoint-REs in the K562 reporter cell line using Cas9 ribonucleoprotein (RNP).

The K562 reporter cell line harboring *T2A-eGFP* fusion downstream of *EVI1* was described recently (Chapter 5)**.** Cas9 RNPs were prepared as described previously.[6] 10 pmoles of concentrated HDR template was then added to the RNP mixture. 200 000 K562 cells were harvested, washed once in PBS and resuspended in 20 μl nucleofection solution (SF Cell Line 4D-Nucleofector kit, Lonza). Cells with RNP and HDR template were combined in a Lonza 4D nucleofection strip and nucleofected using FF-120 program, incubated for 10 min at room temperature and transferred to a 96-well plate with the complete medium. To enhance the *knock-in* rate of 3q21 sequences, the recovery medium was supplemented with the 30 μM HDR Enhancer (Integrated DNA Technologies) and the medium was changed 24 hours after nucleofection.

To generate K562 breakpoint-RE single clones, the cells were seeded in methylcellulose 5 days after nucleofection. For K562 *G2DHE* clones, the cells were sorted for the GFP expression on a FACSAria Fusion (BD Biosciences) 5 days after nucleofection, expanded for the following 7 days and seeded in methylcellulose medium. 10-14 days after seeding the colonies were transferred to a 96-well plate and expanded. Single-cell clones were screened for the presence of desired edits by PCR.

Proliferation assay

Cell proliferation was assessed using CellTiter-Glo Luminescent Cell Viability Assay (Promega) following manufacturer's instructions in 96-well plates at 5 000 cells/well in 100 μl of complete medium. Luminescence was measured on a Victor X3 plate reader (Perkin Elmer) daily for at least 5 days. Results were normalized to values from day 0.

Colony formation assay

Cells were seeded in methylcellulose medium MethoCult H4230 at 1 500 cells/3.5 cm dish in technical duplicates. After 14 days the colonies were imaged using a Cell Observer (Zeiss) and colonies were quantified with an ImageJ macro developed in-house.

Genomic DNA extraction and PCR validation of single clones

Genomic DNA was isolated with DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions. Purified genomic DNA was amplified using Q5 High-Fidelity DNA polymerase (New England Biolabs) and primers (Supplementary Table 1) flanking the deletion or insertion site, gel-purified with QIAquick gel extraction kit (Qiagen) and verified by Sanger sequencing.

RNA isolation, reverse transcription and quantitative PCR (RT-qPCR)

RNA was isolated with RNeasy Plus Mini Kit (Qiagen). 1 μg of total RNA was reverse transcribed with the SuperScript III RT using random hexamers (Thermo Fisher Scientific). Quantitative PCR reactions were run on the CFX96 instrument (Bio-Rad Laboratories) as described previously.[7]

Western blotting

Cells were washed twice in PBS and lysed (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) containing Complete Protease Inhibitor Cocktail (Roche) for 20 min on ice. Lysates were pre-cleared by centrifugation at 15,000 rcf at 4°C for 20 min and protein was quantified with Protein Assay Dye Reagent (Bio-Rad Laboratories). 50 mg of protein lysate was resolved on SDS-PAGE gel and transferred onto Immobilon-FL PVDF 0.45 μm membranes (Merck). After blocking in Intercept (TBS) blocking buffer (LI-COR Biosciences) diluted 1:1 with 1xTBS, membranes were incubated with primary rabbit anti-EVI1 (2265, Cell signaling) and mouse anti-β-actin (AM1829B, Abcepta) antibodies overnight at 4°C, washed with TBS-T followed by incubation with donkey anti-rabbit IRDye 800CW and donkey anti-mouse IRDye 680RD (LI-COR Biosciences). Protein bands were visualized using the Odyssey CLx Imaging System (LI-COR Biosciences).

RNA-Seq and data analysis

RNA-Seq was described previously.[7] For MOLM-1 deletion clones, total RNA was extracted with RNAeasy Plus Mini kit (Qiagen) followed by ribosomal RNA depletion with RiboZero Gold kit (Illumina). Depleted RNA samples were processed according to the TruSeq Stranded mRNA kit (Illumina) excluding the poly(A) enrichment step. The libraries were multiplexed and paired-end sequenced (2 x 100 bp) on a HiSeq 4000 system (Illumina). Raw data were aligned to the human reference genome (hg19) using STAR aligner.[8] Transcript counts were produced using Featurecounts[9] based on hg19. The analysis of monoallelic expression based on informative SNPs was described previously.[7]

4C-Seq and data analysis

4C sample preparation, sequencing and data analysis was performed as described previously.[7,10] Multiplexed single-read sequencing was performed on a HiSeq 2000 v4 platform (1 x 50 bp, Illumina).

ChIP-Seq

ChIP experiments with histone marks H3K4me3 and H3K27ac were performed as previously described.[7] Immunoprecipitated DNA was processed according to the Illumina TruSeq ChIP Sample Preparation Protocol (Illumina) and single-read sequenced (1 x 0 bp) on the HiSeq 2000 v4 platform; Illumina).

3q capture-Seq

3q capture-Seq data was described previously.[7]

SUPPLEMENTARY TABLES

Name	Sequence (5'-3')	Purpose
gMOLM1del_q21_2	CGTCCACGAGCCACGCGCCA	sgRNA used for the deletion of a retroelement in MOLM-1
gMOLM1del_q26_2	AGACACGGAAAATGACTCTG	
MOLM-1_vdel_2F	AATGAATAATGGACTAGCACAGCC	Validation primers for RE deletion in MOLM-1
MOLM-1_vdel_2R	CATATAAGGCACTTCACAGATAGCC	
MunI_LKO.P_U6	tattcaattGGGCCTATTTCCCATGATTCC	Cloning primers
Bst1107I_CrV2_r	agctaagtataCCCACTCCTTTCAAGACCTAG	
gMOLM1-3q26_K562_1	GCAAATGTAATAATATCCCA	sgRNA used for the knock-in experiments in the K562 reporter cell line
g3071-3q26_K562_2	GACCAGGATGACTTCAACGG	
3071_HR-L_for	AAGCGCATATGATTATCACAGACC	HDR template amplification primers
3071_HR-R_rev	CATGACCTTGTCAGAAAATATGTCC	
3071_HR-L_Enh_for	**AAGCGCATATGATTATCACAGACCAGGAT GACTTCAA**CCTCCAGGTGTCCAGAGCCCG	
3071_HR-R_Enh_rev	**CATGACCTTGTCAGAAAATATGTCCTTACC TCCACCG**ACTCCAAGCACCTGCCAAGGC	
MOLM1_HR_L_for	TATCCCATTGACACAATG	
MOLM1_HR_R_rev	TGTCTGGGTTTGTCTTAAC	
MOLM1_HR_L_Enh_for	**TATCCCATTGACACAATGTTTTTCCCAACT TCCTTGG**CCTCCAGGTGTCCAGAGCCCG	
MOLM1_HR_R_Enh_rev	**TGTCTGGGTTTGTCTTAACAGCAAATGTAA TAATATC**ACTCCAAGCACCTGCCAAGGC	

Supplementary table 1: list of sgRNA and primer sequences used in this study. Restriction sites used for cloning are underlined. 35-bp homology arms introduced with 5' overhangs are shown in bold.

Name	Sequence (5'-3')
3071_R_HR	**AAGCGCATATGATTATCACAGACCAGGATGACTTCAA**TATCCTGTTATTTCTGCCGCA AGTAACAGAAAGCCTGGCGACTTCTGGCTTATGCATCAAAGAGTGGTTCTCTCATGTC ACAGGGAGTCTGGAGGCATGCATTCTTGGGCTGGTTTAAACAGTGGACGGGGGACAT CATCAAGAACCCCTTGACTGTCCATCTTCCCCTCTGCCAGCCTTGGCAGGTTCATTTT CCTCCCTAGGATTGGCTCTTCAAGGTCAGAAGGTGGCTGCATCGACTCCAGGATTCAT ATCACAGCCATGTCCAAAAGCCAGGTGAAGGCAGGGTCCCTTCCTGTGCACCTTTAG GAAGAGGAAAAAATTCCCCCAAAGCTCCCCAGAAGACCACCTCC**CGGTGGAGGTAAG GACATATTTTCTGACAAGGTCATG**
MOLM1_R_HR	**TATCCCATTGACACAATGTTTTTCCCAACTTCCTTGG**CCCGCTCTCAGCTCTCTGGGG AGGTCCCCTGTGAAGGGCCTGTCTTTCCTGTCCCCTCCCTCGGATCATTAATGAACCA CTCTCTCTCTCTTTTTCCTTTTCAGTGATTGGGCTGTCGGAATCAAAGAGGCCCTCCG GTGACTCAGGCGTTTCCCATTGCCCTATTCTAGCATCCTCTACTCTTAGCACTGAGAGT TTTTCACGTCCTATTTGGAACTGATAGGAAACCCTTTCATTGTTTCGCTACATGGATATT AGCACTGATGAAGGAAGCGCTCCA**GATATTATTACATTTGCTGTTAAGACAAACCCA GACA**
Enhancer_revc	CCTCCAGGTGTCCAGAGCCCGGCTGGCGCAGGTGGGGCAGCCAGGCTGAGGCCGC AGGGCAAGGCCGGGCTGAGCTGTCCTTTTGCCTCTCTCCGGGTGAGGCCTCCTCTGT CCTTCCTTTCCCAGTGCCCCCCCGCCCTGCTCCTCTGCGGGGGTACAGCATGGGGAG GTTGTTGACAATTTGTGCATTAATCTCTCTGCCGGGGGCCTCCTGGTCAGGTGGTGCT TATCAGGGCCCTGCTCCTGTCAGATGGGCCTTCCGGAGCCCGTGGGCGAGGGGGGG TGGCAGAAGCCCCTGCTGCAGGGCCTGGAGGCTGCCTCACCATCCCCAGGCCTTCA CATCCCCTTGCGCCCCCGCGGGCAGCCTGCTTTACCACATCAGGATACAGAACATTG CATGAATTCCGGTCTCAATGGAATTATTTCAGAATTCTGGTCAACCGCAAGCTCAGGG TATTTCACAAACTATCTATCACTGCTGTGCGGTGGGCAGAGTCTGGGCAGGGGGCAA GTAACGGATGCAGGGAGGCCCCGTCCACTGCCTGTGGGGGTGGCATGTCCCCCCCA GTGGCACCCCTGCCTCTGCAGACAGATGGCGTGTCCCCTGCCCCCCATCAGCAGGG GAGCTCAGGCCTGGAGTGGGGGCTTAGGAGAGGCAGCGGGAGGCCAGGCCCTGGC TTTGAGGGTACCTTCGAGAGGGCACCCAGGTTTGGGGCCCACTCCTACCAGCGTGGG GCCTTGGCAGGTGCTTGGAGT

Supplementary table 2. Sequences of HDR templates used in K562 *knock-in* experiments. Homology arms are shown in bold.

SUPPLEMENTARY REFERENCES

1. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489(7414):57–74.

2. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. Nucleic Acids Res 2018;46(D1):D794–D801.

3. Haeussler M, Schönig K, Eckert H, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol;17(1):.

4. Kreutzberger AJB, Kiessling V, Liang B, et al. Reconstitution of calcium-mediated exocytosis of dense-core vesicles. Sci Adv;3(7):.

5. Paix A, Folkmann A, Seydoux G. Precision genome editing using CRISPR-Cas9 and linear repair templates in C. elegans. Methods 2017;121–12286–93.

6. Richardson CD, Ray GJ, Bray NL, Corn JE. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. Nat Commun 2016;7(1):12463.

7. Gröschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. Cell 2014;157(2):369–381.

8. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15–21.

9. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30(7):923–930.

10. Van De Werken HJG, De Vree PJP, Splinter E, et al. 4C technology: Protocols and data analysis. 1st ed. Elsevier Inc.; 2012. 89–112 p.

5

The leukemic oncogene *EVI1* hijacks a *MYC* super-enhancer by CTCF facilitated loops

Sophie Ottema[1,2,*], Roger Mulet-Lazaro[1,2,*], Claudia Erpelinck[1,2,*], Stanley van Herk[1,2], Marije Havermans[1,2], Andrea Arricibita Varea[1,2], Michael Vermeulen[1], H. Berna Beverloo[3], Stefan Gröschel[4,5], Torsten Haferlach[6], Claudia Haferlach[6], Bas Wouters[1,2], Eric Bindels[1], Leonie Smeenk[1,2,**], and Ruud Delwel[1,2,**]

[1]Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands, [2]Oncode Institute, Erasmus University Medical Center, Rotterdam, The Netherlands, [3]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands, [4]German Cancer Research Center, A380, Heidelberg, Germany, [5]Department of Internal Medicine V, Heidelberg University Hospital, Heidelberg, Germany, [6]Munich Leukemia Laboratory, Munich, Germany

*/** shared authorship

Correspondence: h.delwel@erasmusmc.nl

Highlights
- A t(3;8) AML model was generated *in vitro* using CRISPR-Cas9
- The *MYC* super-enhancer hyperactivates *EVI1* expression in t(3;8) AML
- A single hematopoietic enhancer element is critical for *EVI1* expression
- CTCF-binding site upstream of *EVI1* hijacks enhancers in 3q26-rearranged AML

ABSTRACT

Chromosomal rearrangements are a frequent cause of oncogene deregulation in human malignancies. Overexpression of *EVI1* is found in a subgroup of acute myeloid leukemia (AML) with 3q26 chromosomal rearrangements, which is often therapy resistant. In AMLs harboring a t(3;8)(q26;q24), we observed the translocation of a *MYC* super-enhancer (*MYC* SE) to the *EVI1* locus. We generated an *in vitro* model mimicking a patient-based t(3;8)(q26;q24) using CRISPR-Cas9 technology and demonstrated hyperactivation of *EVI1* by the hijacked *MYC* SE. This *MYC* SE contains multiple enhancer elements, of which only one recruits transcription factors active in early hematopoiesis. This enhancer element is critical for *EVI1* overexpression as well as enhancer-promoter interaction. Multiple CTCF sites in the *MYC* SE facilitate this enhancer-promoter interaction, which also involves a CTCF site upstream of the *EVI1* promoter. We hypothesize that this CTCF site acts as an enhancer-docking site in t(3;8) AML. Genomic analyses of other 3q26-rearranged AML patient cells point to a common mechanism by which *EVI1* uses this docking site to hijack enhancers active in early hematopoiesis.
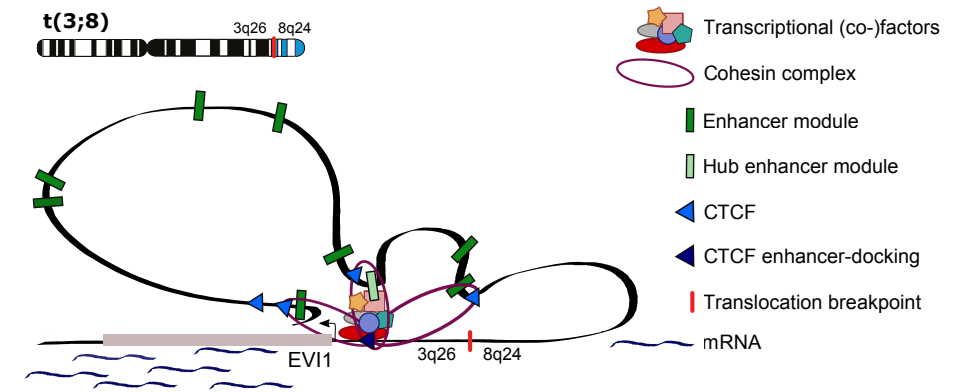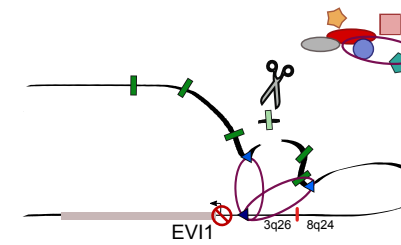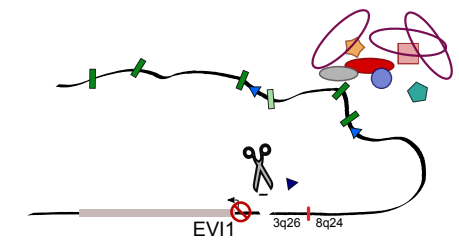
INTRODUCTION

The expression of cell lineage specific genes is highly regulated. Specific enhancer-promoter interactions and transcription factor binding to regulatory elements delineate gene expression profiles that define cell identity and function [1]. Physical interactions between enhancers and promoters primarily occur within chromosome segments enclosed by chromatin loops known as topologically associated domains (TADs) [2]. TADs are separated from each other by boundaries typically containing convergent CTCF (CCCTC-binding factor) occupied sites [3]. According to the loop extrusion model, the cohesin complex catalyzes the formation of loops and CTCF dimers act as anchors to these loops [4]. CTCF and the cohesin complex, but also other factors like Ying Yang 1 (YY1), may also contribute to enhancer-promoter looping [5-8]. However, not all promoters or enhancers within a TAD interact with each other. The mechanisms by which promoters interact with certain enhancers and not with others are not fully understood [9, 10]. Transcriptional control of genes driven by particular enhancer-promoter combinations depends on the availability of transcription factors and their ability to bind specific regulatory elements [8, 11].

Chromosomal rearrangements frequently lead to changes in the expression or function of genes causing malignant transformation [12]. Often breakpoints are found within gene bodies, resulting in fusion oncogenes driving tumorigenesis [13]. Alternatively, when a regulatory element of a certain gene is translocated into the vicinity of another gene, it can lead to deregulation of both the donor and the acceptor genes. Well-described examples are the inv(3)(q21q26) or t(3;3)(q21;q26) rearrangements in acute myeloid leukemia (inv(3)/t(3;3) AML), in which a *GATA2* enhancer at 3q21 is hijacked by *EVI1* at 3q26, causing *EVI1* overexpression and *GATA2* haploinsufficiency [14, 15]. AML is a heterogeneous disease, with *EVI1* positive (*EVI1*+) inv(3)/t(3;3) patients being identified as a subgroup with a very poor response to therapy [16-19]. Besides inv(3)/t(3;3), many other *EVI1*+ AML cases with 3q26 rearrangements have been reported, including translocations t(2;3)(p21;q26), t(3;7) (q26;q24), t(3;6)(q26;q11) and t(3;8)(q26;q24) [18, 20-27]. We hypothesize that in all these rearrangements *EVI1* overexpression is induced by the repositioning of an enhancer that can interact with the *EVI1* promoter, as shown for inv(3)/t(3;3) AML [14, 15]. We performed targeted next generation sequencing (NGS) of the long arm of chromosome 3 (3q-seq) in translocation t(3;8)(q26;q24) AML harboring an *EVI1/MYC* rearrangement [22, 27]. Applying CRISPR-Cas9 technology, we generated a human t(3;8) cell line model with an *eGFP* reporter cloned 3' of *EVI1*. This unique model was used to investigate how enhancer-promoter interactions drive oncogenic *EVI1* expression in leukemia. We demonstrate that CTCF in combination with transcription factors active in early hematopoiesis is essential in enhancer hijacking and oncogene activation.

RESULTS

MYC super-enhancer translocation and *EVI1* overexpression in t(3;8)(q26;q24) AML

Using 3q-seq, the exact chromosomal breakpoints were determined in 10 AML samples with a translocation t(3;8)(q26;q24), hereafter referred to as t(3;8) AML. All breakpoints at 3q26.2 occurred upstream of the *EVI1* promoter (Figure 1A). At chromosome 8, the breakpoints were downstream of the oncogene *MYC* at 8q24, leaving the gene intact at its original location. In all 10 cases a genomic region reported as a *MYC* super-enhancer (SE) had been translocated to *EVI1* (Figure 1B). The *MYC* SE harbors approximately 150 Kb of open chromatin enriched with histone mark H3K27 acetylation (H3K27ac) and is located 1.7 Mb downstream of *MYC* (Figure 1B). This locus has been reported to be essential for transcriptional control of *MYC* expression in normal hematopoiesis [28]. H3K27ac determined by ChIP-seq revealed *EVI1* promoter activity in t(3;8) AML patient cells, comparable to the promoter activity in AML with inv(3)(q21q26). H3K27ac was absent at the *EVI1* promoter in *EVI1* negative (*EVI1*-) non-3q26 AML (Figure 1A, lower panel). Accordingly, *EVI1* expression was found to be highly elevated in t(3;8) compared to non-3q26 rearranged AMLs (Figure 1C). The *EVI1* levels in t(3;8) AMLs were comparable to the levels found in AMLs with inv(3)/t(3;3). These data support the hypothesis that *EVI1* overexpression in t(3;8) AML is caused by the translocation of the *MYC* SE.

A t(3;8) cell model recapitulates *EVI1* overexpression in human AML

To study the transcriptional activation of *EVI1* by the *MYC* SE, we generated a human myeloid cell model with a translocation t(3;8)(q26;q24). We introduced e*GFP* in frame with a T2A self-cleavage site downstream of *EVI1* in K562 cells (Figure 2A). Successful integration of the insert is shown for two clones by flow cytometry and PCR (Figure 2B, S1A-C). Decreased eGFP levels were observed in the K562 *EVI1-eGFP* model after shRNA directed *EVI1* knockdown (Figure 2C-D and S1D-G). Next, sgRNAs were designed based on the genomic breakpoints of one of the t(3;8) AML patients in our cohort (Figure 1A). Double strand DNA breaks were generated at 3q26.2 and 8q24 (Figure 2E) using CRISPR-Cas9 technology. We hypothesized that the translocated *MYC* SE can activate *EVI1* transcription, which consequently leads to increased *eGFP* levels. As shown in Figure 2F, less than 0.1% of the sgRNA-treated K562 *EVI1-eGFP* cells showed increased eGFP levels. After two consecutive rounds of FACS sorting in combination with cell culture expansion, we obtained 95% eGFP positive cells of which single clones were isolated by single cell sorting (process done similarly for both clones 8 and 24, Figure 2F shows clone 24). The presence of a t(3;8) was demonstrated for four of these clones by PCR (Clone 24-7, Figure 2H) and Sanger sequencing (S2A). A combination of three separate diagnostic FISH probes for *MECOM*, *MYC* and centromere chromosome 8 confirmed the successful generation of a translocation t(3;8) in all four clones (Figure S2B-E). The translocation caused a strong increase of mRNA and protein levels of *EVI1* as

well as of *eGFP* expression (Figure 2G, J, K). No significant difference in *MYC* expression was observed between the parental K562 *EVI1-eGFP* and t(3;8) clones (Figure 2I). Upon *EVI1* knockdown by shRNA, *eGFP* and *EVI1* expression were reduced as shown for clone 24-7 and 8-4 (Figure 2L-M and S2F-G). We conclude that eGFP is a sensitive and reliable marker for *EVI1* expression in this *EVI1-eGFP* t(3;8) model, and that the translocated *MYC* SE strongly enhances *EVI1* transcription.

EVI1 promoter hyperactivation upon interaction with *MYC* SE in t(3;8) AML

4C-seq experiments taking the *EVI1* promoter (EVI1_PR) as a viewpoint revealed specific interaction with the *MYC* SE in *EVI1-eGFP* t(3;8) cells, which was not found in the parental K562 *EVI1-eGFP* line (clone 24-7 and clone 24 respectively, Figure 3A). This t(3;8)-specific interaction between the *EVI1* promoter and *MYC* SE was confirmed in t(3;8) clone 8-4 (Figure S3D) and by reciprocal 4C-seq using the *MYC* SE as a viewpoint (clone 24-7, Figure 3B). A comparable *EVI1* promoter – *MYC* SE interaction was found in a primary t(3;8) AML sample (Figure 3A-B), confirming that the K562 *EVI1-eGFP* t(3;8) model recapitulates primary AML. ChIP-seq for H3K4 trimethylation (H3K4me3, Figure 3C) indicated the presence of an active *EVI1* promoter in all K562 clones. However, H3K27 and H3K9 acetylation (H3K27ac and H3K9ac) levels were strongly increased at the promoter in all four t(3;8) clones, revealing a hyperactivated *EVI1* promoter (Figure 3D-E) upon interaction with the translocated *MYC* SE.

One critical enhancer element in the *MYC* SE drives *EVI1* transcription

The *MYC* SE is a cluster of multiple individual enhancer modules that may recruit different sets of transcription factors [28]. To investigate which of the enhancer elements are driving oncogenic *EVI1* transcription in t(3;8) AML, we designed sgRNAs to sequentially delete those individual elements. H3K27ac ChIP-seq data of a primary t(3;8) AML and of t(3;8) clone 24-7 were used to illustrate the different enhancer modules A-I described previously [28] (Figure 4A). The deletion of these modules by CRISPR-Cas9 using specific sgRNA pairs was shown by PCR and the effect on *EVI1* expression was determined by flow cytometry (Figure 4B). Only the deletion of module C caused loss of *EVI1/eGFP* expression. Due to existence of multiple alleles (K562 has trisomy 8) and partial efficiency of CRISPR-Cas9 in creating deletions, the translocated allele is exclusively targeted in a subpopulation of cells. As a consequence, not all cells lose *EVI1* expression and show a GFP shift in the flow cytometry plot. A loss of *EVI1* mRNA and EVI1 protein levels was observed in the eGFP negative sorted cell fraction when module C was deleted (Figure 4C-E and S3A). In a control clone in which *EVI1-eGFP* expression was increased due to the amplification of *EVI1* instead of the translocation of the *MYC* SE (Figure S4A-E), the expression of *EVI1*-eGFP was not affected by mutating the *MYC* SE (Figure S4G-H). ChIP-seq data revealed binding of early hematopoietic regulators (GATA2, FLI1, ERG, RUNX1, LMO2 and LYL1) to *MYC* SE module C in CD34+ hematopoietic stem and progenitor cells (HSPCs) [29] (Figure 4F). 4C-seq taking the *EVI1* promoter as a

viewpoint revealed that the strong interaction with the *MYC* SE was severely diminished in the eGFP negative fraction upon deletion of module C (Figure 4G). This loss of chromosomal interaction was also observed taking the *MYC* SE as a viewpoint (Figure 4H). Deletions of enhancer element D and I affected neither *EVI1* expression nor enhancer-promoter looping (Figure 4B and S3B-C). Our data demonstrate that aberrant *EVI1* expression in t(3;8) AML depends on a single enhancer module within the *MYC* SE that recruits a cluster of key hematopoietic transcription factors and facilitates promoter-enhancer looping.

CTCF binding sites in *MYC* SE are involved in the interaction with the *EVI1* promoter

The *EVI1* promoter interacts with the *MYC* SE over a long stretch of chromatin (275 Kb) with multiple zones of strong interaction indicative of a highly organized enhancer-promoter interaction (Figure 5A). These high interaction zones in the *MYC* SE were associated with enhancer elements, but also with CTCF/Cohesin binding based on ChIP-seq data (Figure 5A). Notably, CTCF binding motifs in the *MYC* SE are arranged in a convergent orientation to that of a CTCF binding site upstream of the *EVI1* promoter, suggesting the existence of a CTCF-facilitated enhancer-promoter loop. Using CRISPR-Cas9 technology, we sequentially deleted every CTCF binding site in the *MYC* SE. The deletions and their effect on *EVI1* expression were shown by PCR and *eGFP* flow cytometry (Figure 5B). A fraction of cells lost *eGFP* expression upon deletion of each of the CTCF binding sites in the *MYC* SE. The CTCF site closest to element C (CTCF2) was deleted and cells were sorted based on eGFP expression. A severe loss of promoter-enhancer interaction was observed in the eGFP negative cells (Figure 5C and S5A). This strongly supports a role for CTCF/cohesin in the promoter-enhancer complex formation and maintenance, and consequently in *EVI1* regulation in t(3;8) AML.

CTCF binding site upstream of the *EVI1* promoter hijacks the *MYC* SE in t(3;8) AML

Upstream of the *EVI1* promoter a CTCF binding site in the forward orientation (CTCF EVI1_PR) was found by ChIP-seq and motif analysis (Figure 5A and 6A). Deletion of this CTCF binding region caused loss of *EVI1* expression as determined by eGFP flow cytometry. This loss of eGFP expression was comparable to the loss of expression upon deletion of the *MYC* SE CTCF sites (Figure 5B). Deletion of this CTCF site also caused a severe loss of promoter-enhancer looping in eGFP negative cells, as measured by 4C-seq (Figure 5D and S5B). ChIP-seq showed that CTCF occupancy upstream of the *EVI1* promoter was indeed reduced upon deletion of this site (Figure 5E). CTCF occupancy at other CTCF binding sites, e.g. upstream of the *MYC* promoter (Figure 5F), was not affected. Aiming to specifically target the CTCF binding and not other transcription factor binding motifs within this genomic region, more subtle mutations were made close to the CTCF binding motif using a single sgRNA (Figure 6A). The mutations introduced by this single sgRNA strongly downregulated *eGFP/EVI1* expression (Figure 6B). A high mutation frequency was obtained in the eGFP negative sorted cells near the CTCF motif (Figure 6C-D). These mutations led to a decrease of CTCF binding specifically

at this site (Figure 6E-F) and a severe loss of enhancer-promoter interaction (Figure 6G) in the eGFP negative sorted cells. Taken together, these data demonstrate an important role for the CTCF binding site upstream of the *EVI1* promoter in the hijacking of the *MYC* SE and the hyperactivation of *EVI1*.

CTCF enhancer-docking site upstream of the *EVI1* promoter is preserved in 3q26-rearranged AML

The essential role of the CTCF binding site upstream of the *EVI1* promoter in mediating the interaction with a hijacked enhancer would predict that this site remains unaffected in 3q26-rearranged AMLs. Indeed, all breakpoints of t(3;8) AMLs analyzed were found upstream of this CTCF site, placing the *MYC* SE 5' of *EVI1* (Figure 1A and 7A). In t(3;3)(q21;q26) AML the *GATA2* enhancer similarly translocates 5' of the *EVI1* promoter and of the CTCF binding site (Figure 7A) [15]. In AML with inv(3)(q21q26) the *GATA2* enhancer translocates 3' of *EVI1* [15] (Figure 7A), leaving the enhancer-interacting CTCF site in position with respect to *EVI1* as well. We collected samples from AML patients with translocations t(2;3)(p21;q26), t(3;7)(q26;q24) or t(3;6)(q26;q11) and carried out 3q-seq. The exact breakpoints at chromosome 3q26 in each of those 3q26-rearranged AMLs are depicted in Figure 7A. Irrespective of whether a translocation had occurred 3' or 5' of *EVI1*, the CTCF site flanking the *EVI1* promoter was never disrupted, suggesting a key role for this binding site in this AML subtype. Enhancers of the genes *GATA2* [15] and *MYC* are respectively responsible for *EVI1* activation in inv(3)/t(3;3) and t(3;8) AML. Using 3q-seq, we observed that regions near the genes *CDK6* (6q11), *ARID1B* (7q24) and *THADA* (2p21) had been translocated to *EVI1* in t(3;6), t(3;7) or t(2;3) AML, respectively (Figure 7B-C and S6). All these genes are expressed in HSPCs (RNA-seq data not shown). Similar to the *MYC* SE in t(3;8) AML (Figure 4F), we found strong regulatory regions close to these, illustrated by H3K27ac and hematopoietic transcription factor binding (Figure 7B-D, S6A). These commonalities suggest a shared mechanism for *EVI1* activation in all 3q26-rearranged leukemias, whereby an active hematopoietic enhancer is hijacked by a CTCF-mediated loop with the *EVI1* promoter.
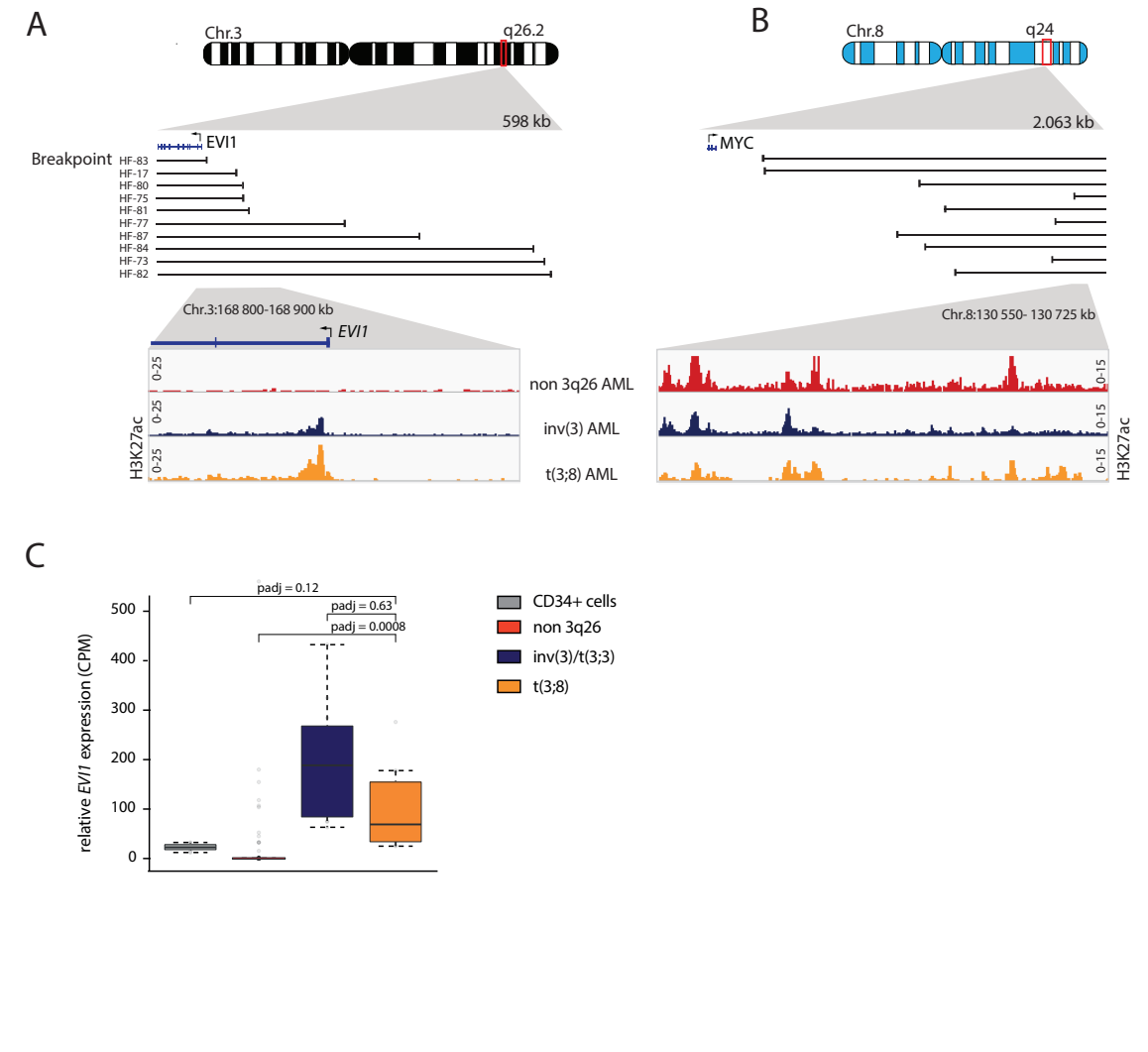
Figure 1. *MYC* super-enhancer translocation and *EVI1* overexpression in t(3;8)(q26;q24) AML
(A) Upper part, schematic depiction of Chr.3, zoomed in on 3q26.2. Black lines correspond to sample specific breakpoints detected by 3q-seq for each indicated t(3;8)(q26;q24) patient. Lower part: zoom-in on the *EVI1* promoter, H3K27ac ChIP-seq data for a primary non-3q26 AML sample in red (N=1, AML-185), an inv(3)(q21q26) in blue (N=1, AML-2190) and a t(3;8)(q26;q24) in orange (N=1, AML-17). **(B)** Similar to A, but here in the upper part a schematic depiction of Chr.8, zoomed in on 8q24. Lower part: H3K27ac ChIP-seq data as in A, but here a zoom-in on the +1.7 Mb *MYC* super-enhancer. **(C)** *EVI1* expression measured by RNA-seq in counts per million (CPM) for normal CD34+ HSPCs (N=9, grey), non-3q26 AMLs (N=114, red) (Mulet-Lazaro, 2020), inv(3)/t(3;3)(q21;q26) AMLs (N=11, blue), and t(3;8)(q26;q24) AMLs (N=10, orange).
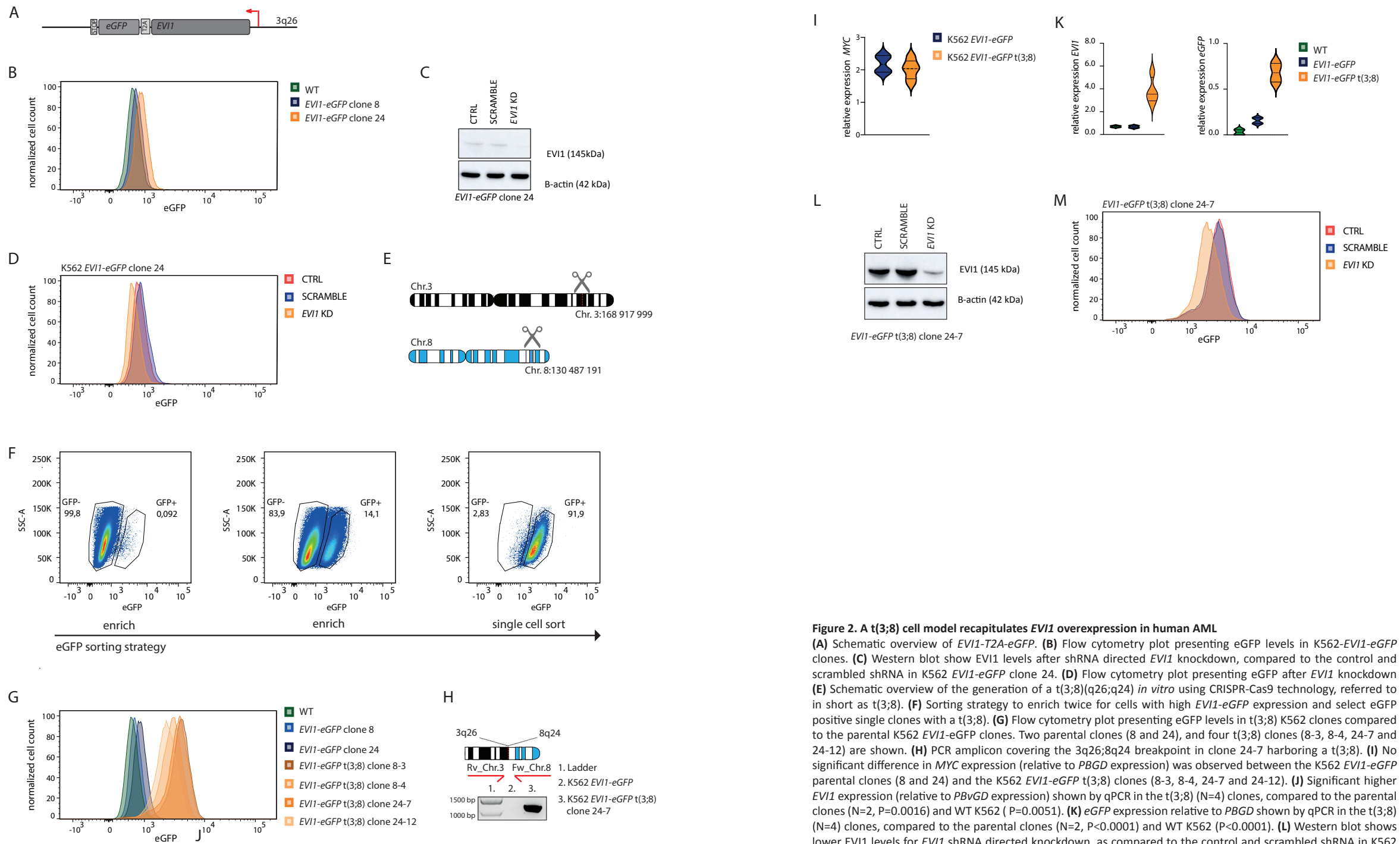
Figure 2. A t(3;8) cell model recapitulates *EVI1* overexpression in human AML

(A) Schematic overview of *EVI1-T2A-eGFP*. **(B)** Flow cytometry plot presenting eGFP levels in K562-*EVI1-eGFP* clones. **(C)** Western blot show EVI1 levels after shRNA directed *EVI1* knockdown, compared to the control and scrambled shRNA in K562 *EVI1-eGFP* clone 24. **(D)** Flow cytometry plot presenting eGFP after *EVI1* knockdown **(E)** Schematic overview of the generation of a t(3;8)(q26;q24) *in vitro* using CRISPR-Cas9 technology, referred to in short as t(3;8). **(F)** Sorting strategy to enrich twice for cells with high *EVI1-eGFP* expression and select eGFP positive single clones with a t(3;8). **(G)** Flow cytometry plot presenting eGFP levels in t(3;8) K562 clones compared to the parental K562 *EVI1*-eGFP clones. Two parental clones (8 and 24), and four t(3;8) clones (8-3, 8-4, 24-7 and 24-12) are shown. **(H)** PCR amplicon covering the 3q26;8q24 breakpoint in clone 24-7 harboring a t(3;8). **(I)** No significant difference in *MYC* expression (relative to *PBGD* expression) was observed between the K562 *EVI1-eGFP* parental clones (8 and 24) and the K562 *EVI1-eGFP* t(3;8) clones (8-3, 8-4, 24-7 and 24-12). **(J)** Significant higher *EVI1* expression (relative to *PBvGD* expression) shown by qPCR in the t(3;8) (N=4) clones, compared to the parental clones (N=2, P=0.0016) and WT K562 (P=0.0051). **(K)** *eGFP* expression relative to *PBGD* shown by qPCR in the t(3;8) (N=4) clones, compared to the parental clones (N=2, P<0.0001) and WT K562 (P<0.0001). **(L)** Western blot shows lower EVI1 levels for *EVI1* shRNA directed knockdown, as compared to the control and scrambled shRNA in K562 *EVI1-eGFP* t(3;8) clone 24-7. **(M)** Flow cytometry plot presenting eGFP after *EVI1* shRNA directed knockdown, as compared to the control and scrambled shRNA in K562 *EVI1-eGFP* t(3;8) clone 24-7.
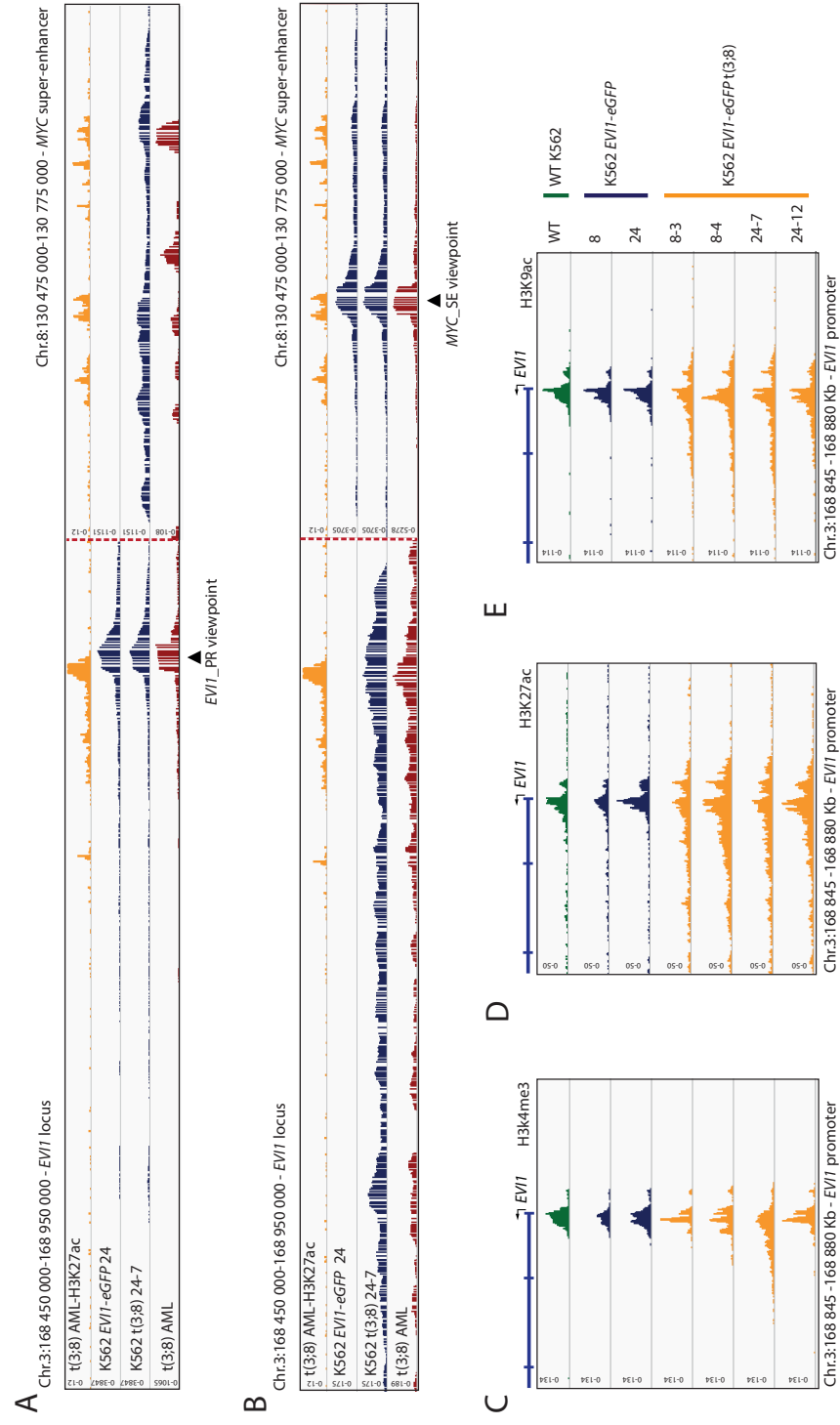
Figure 3. *EVI1* promoter hyperactivation upon interaction with *MYC* SE in t(3;8) AML

(A) Chromatin interaction shown by 4C-seq data, using the *EVI1* promoter as viewpoint (triangle symbol). The upper panel shows H3K27ac ChIP-seq data of a t(3;8) primary AML (HF-80). Indicated by H3K27ac signal peaks, on the left the *EVI1* promoter and on the right the -1.7 Mb *MYC* super enhancer, separated by a dotted red line. In the first 4C track (blue), parental K562 *EVI1-eGFP* clone 24; in the second, K562 t(3;8) *EVI1-eGFP* clone 24-7 (also blue); and in the bottom track (red), data of a primary t(3;8) AML (AML-17). **(B)** Similar to A, but using the *MYC* super enhancer as viewpoint (triangle symbol). The long stretch (500 kb) of chromosomal interaction shown for the K562 t(3;8) *EVI1-eGFP* clone 24-7 shows high resemblance with the interaction seen for the primary t(3;8) AML (HF-80) (second blue and red tracks respectively). **(C)** H3K4me3 ChIP-seq data for K562 *EVI1-eGFP* clones 8, 24 and t(3;8) clone 8-3, 8-4, 24-7, 24-12, highlighting the promoter region upstream of *EVI1*. **(D)** H3K27ac ChIP-seq data, comparing *EVI1* promoter activation of the four t(3;8) clones (orange) to the *EVI1-eGFP* parental lines (blue) or K562 WT (green). **(E)** H3K9ac ChIP-seq data, confirming the hyperactivation of the *EVI1* promoter in the t(3;8) clones (orange).
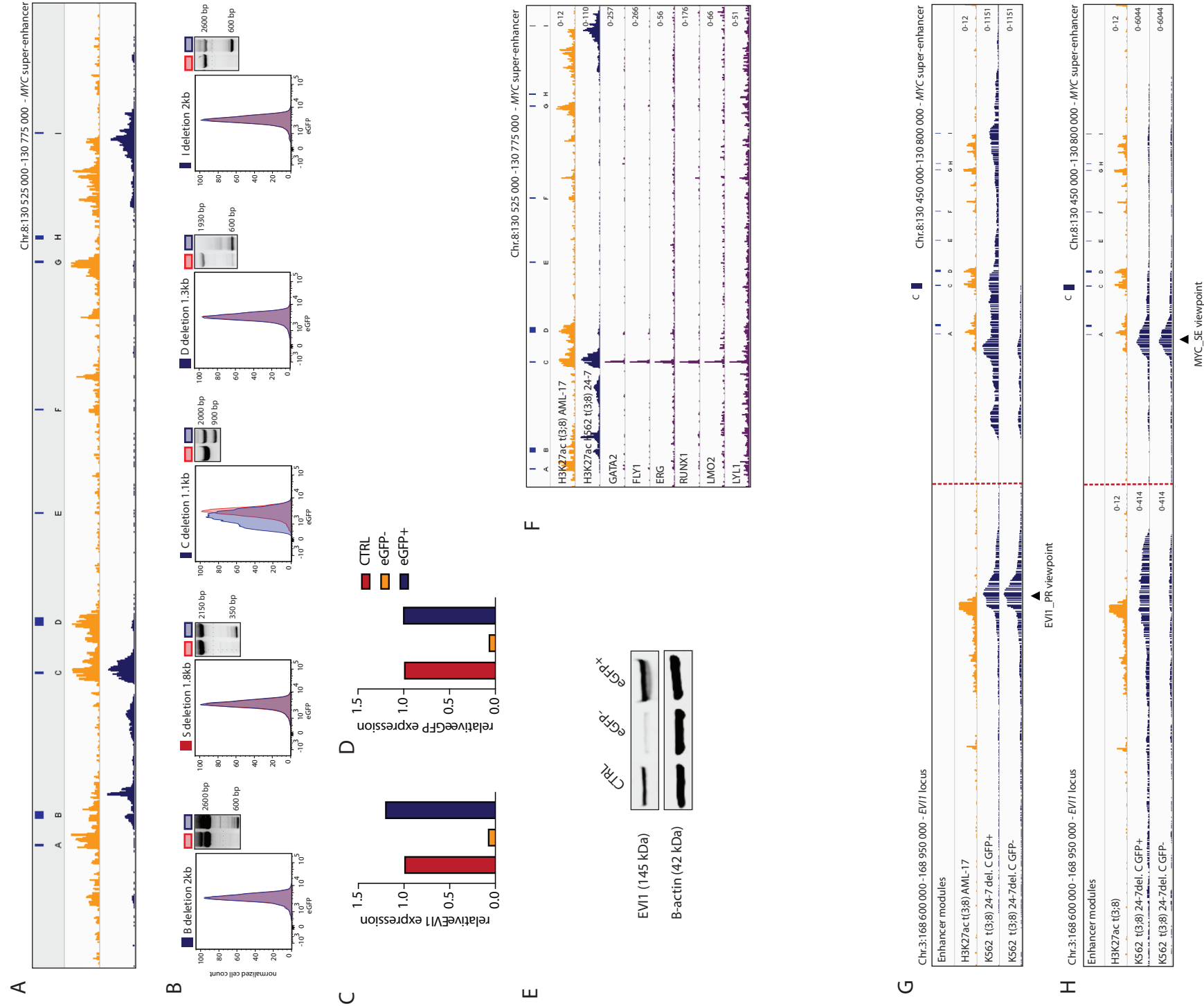
Figure 4. One critical enhancer element in the *MYC* SE drives *EVI1* transcription

(A) Overview of the *MYC* super-enhancer, with previously characterized individual enhancer elements A-I (Bahr et al., 2018) and added element S based on high H3K27ac signal at this location in all K562 *EVI1-eGFP* t(3;8) clones. Underneath, H3K27ac of a primary t(3;8) AML (AML-17, orange) and of K562 t(3;8) *EVI1-eGFP* clone 24-7. **(B)** Flow cytometry plots (clone 24-7) shown for each indicated enhancer element deletion. In red the control cells (no Cas9), and in blue the cells carrying the deletion. On the rivght of each graph the successful deletion of each element is shown by PCR. **(C)** *EVI1* expression relative to *PBGD* by qPCR in eGFP- and eGFP+ sorted cell fractions after deletion of enhancer element C. **(D)** *eGFP* expression relative to *PBGD* by qPCR in eGFP- and eGFP+ sorted cell fractions after deletion of enhancer element C. **(E)** EVI1 protein levels by Western blotting in eGFP- and eGFP+ sorted fractions after deletion of enhancer element C. **(F)** *MYC* SE element C recruits a set of HSPC-active transcription factors shown by ChIP-seq data of CD34+ cells (purple tracks (Beck et al., 2013)). H3K27ac of primary t(3;8) AML (AML-17, orange) and of a K562 t(3;8) (clone 24-7, dark blue) to illustrate enhancer elements. **(G)** Chromatin interaction shown by 4C-seq data, using the *EVI1* promoter as viewpoint (triangle symbol). The *EVI1* promoter and the -1.7 Mb *MYC* SE are shown on the left and right sections respectively, separated by a dotted red line. The upper panel shows H3K27ac ChIP-seq of a t(3;8) primary AML (AML-17, orange track). In blue, 4C-seq tracks of K562 *EVI1-eGFP* t(3;8) clone 24-7 cells in which the enhancer element C was deleted. In the upper blue track, eGFP+ sorted cells, and in the lower track, eGFP- cells. **(H)** Similar as G, but using the *MYC* super-enhancer as a viewpoint (triangle symbol).
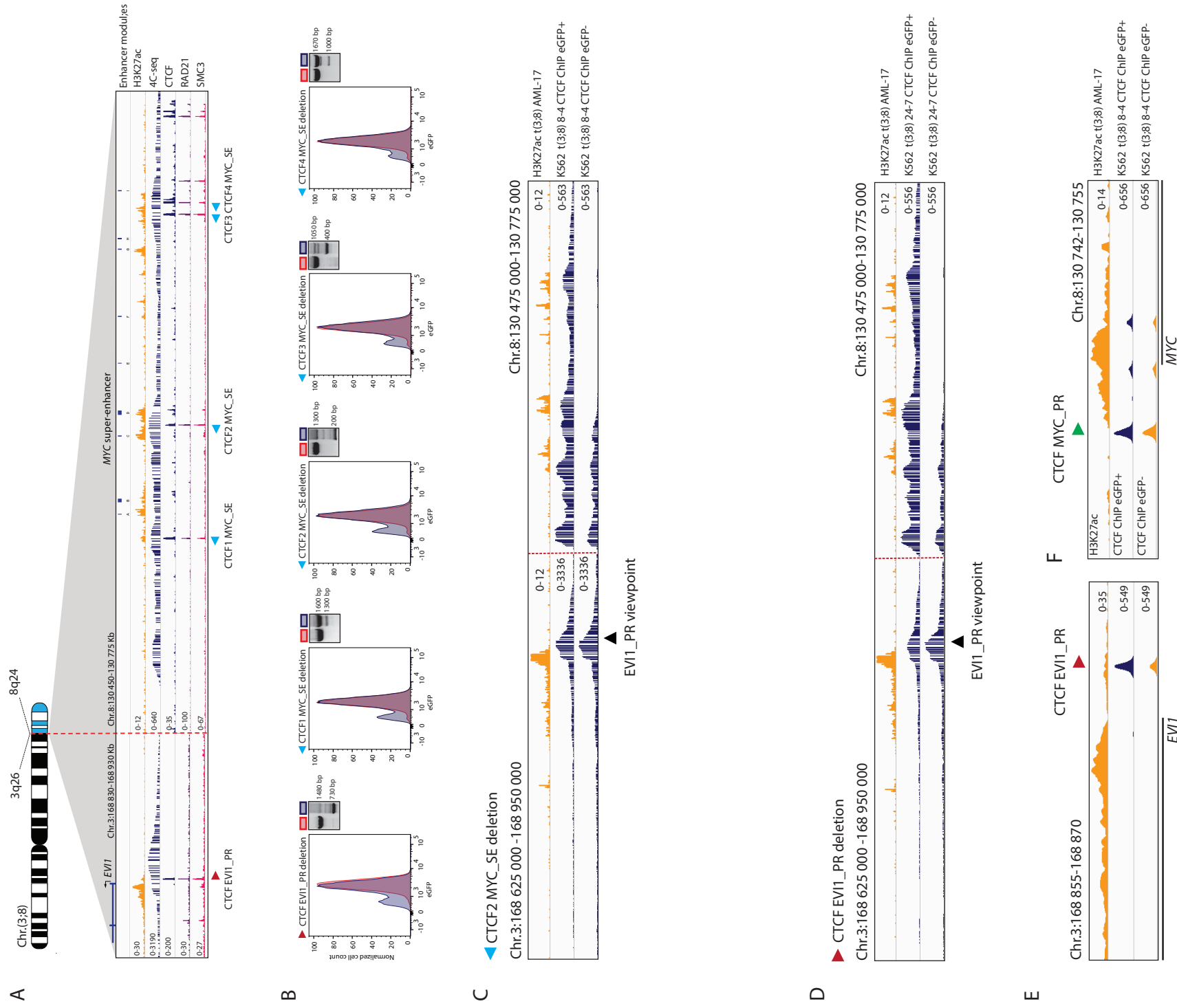
Figure 5. CTCF binding sites in *MYC* SE are involved in the interaction with the *EVI1* promoter

(A) Schematic overview of the t(3;8) model, zoom-in on the breakpoint with 3q26 on the left and 8q24 on the right (separated by a dotted red line). The *EVI1* promoter and the *MYC* super-enhancer are illustrated by ChIP-seq H3K27ac data of primary t(3;8) AML (AML-17) in orange on top. Below in blue, 4C-seq data showing the interaction pattern between the *EVI1* promoter and the *MYC* enhancer elements in K562 t(3;8) clone 24-7. The lower 3 tracks show ChIP-seq data of CTCF (blue in K562 t(3;8) clone 24-7) and the cohesin subunits RAD21 (purple) and SMC3 (pink), both in K562 and downloaded from Encode (Davis et al., 2018). **(B)** Flow cytometry plot of K562 *EVI1-eGFP* t(3;8) clone 8-4 cells after deletion of the indicated CTCF binding site (blue graph), and in red the control cells (no Cas9). On the right of the graph the deletion is shown by PCR. **(C)** Chromatin interaction at the *MYC* SE in eGFP+ and eGFP- cells, shown by 4C-seq with the *EVI1* promoter as viewpoint after deletion of the *MYC* SE CTCF2 binding site. Comparing chromatin interaction at the *MYC* SE for eGFP+ and eGFP- cells, shown by 4C-seq with the *EVI1* promoter as viewpoint after deletion of the *EVI1* PR CTCF binding site (indicated by red arrow, corresponding to the CTCF *EVI1*_PR locus in A). **(E)** CTCF ChIP-seq presenting CTCF occupancy in eGFP- cells (clone 8-4, orange) compared to eGFP+ cells (clone 8-4, blue) at the CTCF binding site upstream of the *EVI1* promoter after deletion of this CTCF *EVI1*_PR site. The top H3K27ac track indicates the presence of an active promoter. **(F)** The same CTCF ChIP-seq tracks are shown (clone 8-4), but now presenting unchanged CTCF occupancy at the CTCF binding site upstream of the *MYC* promoter.

Figure 6. CTCF binding site upstream of the *EVI1* promoter hijacks a *MYC* SE in t(3;8) AML

(A) ChIP-seq data with a zoom-in on the *EVI1* promoter binding site, indicating the exact location of the sgRNA#1 and the CTCF motif as described by JASPAR (Fornes et al., 2019) below. (B) Flow cytometry overlay plot after targeting the CTCF *EVI1_PR* binding site by sgRNA#1 (clone 8-4, blue graph) and in red the control cells (clone 8-4, no Cas9). On the right of the graph, the mutations introduced by the single sgRNA in the amplicon over the cutting site are shown by PCR. (C) Amplicon-seq data showing the percentage of modified and unmodified reads in the eGFP- sorted cell fraction after targeting CTCF *EVI1_PR* with sgRNA#1. (D Amplicon-seq data showing the mutations in the nucleotides around the Cas9 cleavage site, in the eGFP- sorted cell fraction after targeting CTCF *EVI1_PR* with sgRNA#1. The bars and numbers indicate percentage of reads found with the particular mutation, below the locations of the sgRNA (grey bar) and the CTCF motifs (black lines). (E) CTCF ChIP-seq presenting CTCF occupancy in the eGFP+ (clone 24-7, blue), and eGFP- (clone 24-7, orange) fractions after targeting CTCF *EVI1_PR* with sgRNA#1. The top H3K27ac track indicates the presence of an active promoter. (F) The same CTCF ChIP-seq tracks as in E are shown, but here presenting unchanged CTCF occupa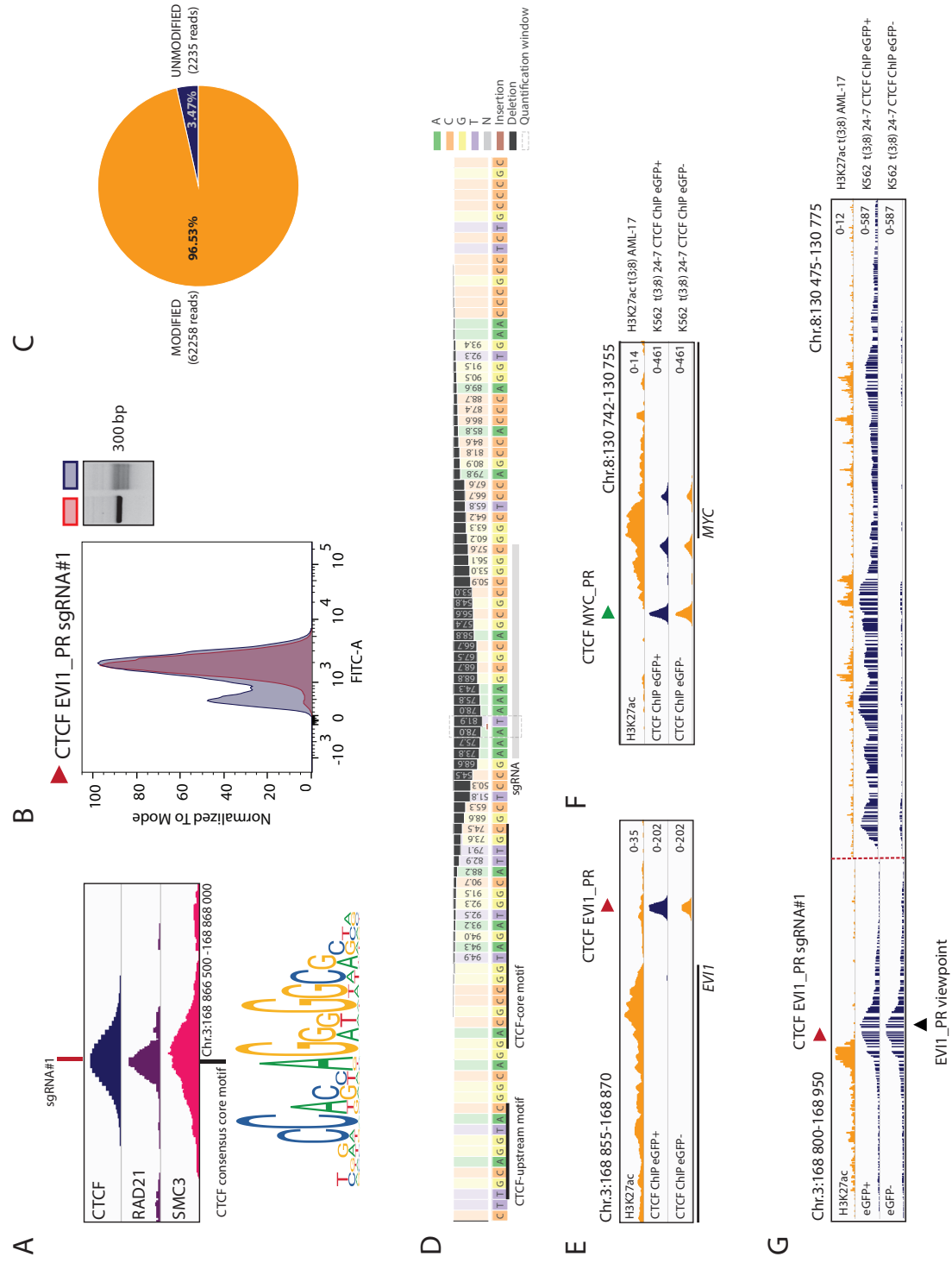ncy at the CTCF binding site upstream of the *MYC* promoter. (G) Chromatin interaction at the *MYC* SE for eGFP+ and eGFP- cells (clone 24-7), shown by 4C-seq with the *EVI1* promoter as viewpoint after targeting the CTCF motif with sgRNA#1.
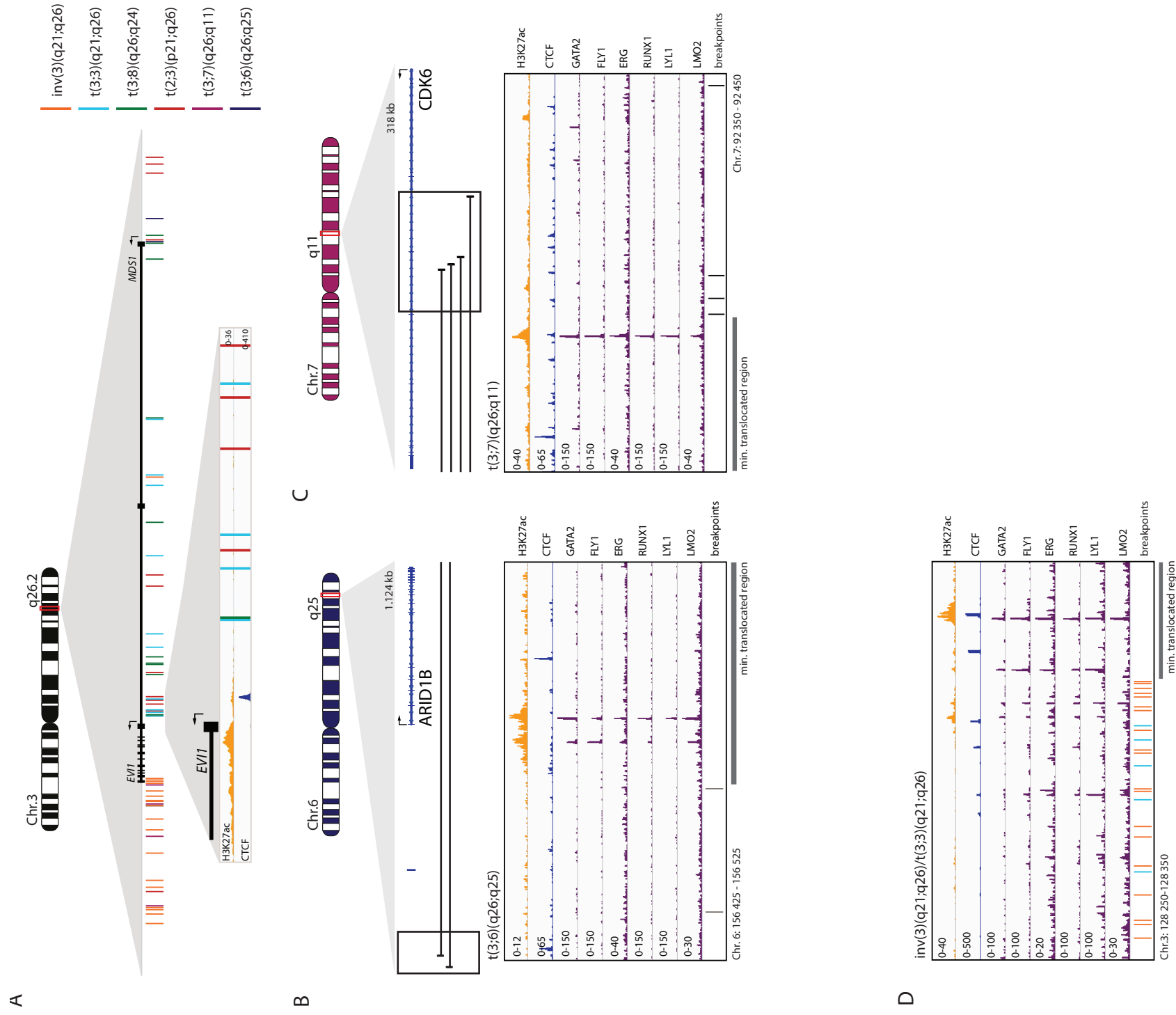
Figure 7. CTCF enhancer-docking site upstream of the *EVI1* promoter is preserved in 3q26-rearranged AML

(A) Schematic depiction of Chr.3 with a zoom-in on the *EVI1* locus, indicating the exact breakpoints (detected by 3q-seq) of 3q26-rearranged AMLs as vertical lines. In the lowest zoom-in panel the *EVI1* promoter with a CTCF binding site upstream marked respectively by H3K27ac (t(3;8) AML-17, orange) and CTCF (K562 t(3;8) clone 24-7, blue) ChIP-seq. (B) Schematic overview of Chr.6 and the locus where the breakpoints (black lines) were found by 3q-seq in t(3;6)(q26;q25) AML. The black box indicates the area of which the zoom-in is shown below. Zoom-in: putative enhancer indicated by H3K27ac (t(3;8) AML-17, orange), CTCF binding (K562 t(3;8) clone 24-7, blue) and HSPC active transcription factor recruitment (CD43+ cell (Beck et al., 2013), purple) at translocation site. The lines below indicate the exact breakpoints. The grey bar the minimal translocated region brought into close proximity of *EVI1* in that specific translocation. (C) Same as B, but here for t(3;7)(q26;q11) AMLs. (D) Same as B, but here for inv(3)/t(3;3)(q21;q26) AMLs. The exact translocated locus was previously shown to be an enhancer of *GATA2* (Gröschel et al., 2014).

DISCUSSION

In this study we investigated how *EVI1* is deregulated in AML with a translocation t(3;8) (q26;q24). Using an *EVI1-eGFP* t(3;8) model, we demonstrated that hyperactivation of *EVI1* was driven by a hijacked *MYC* SE. One enhancer element within this *MYC* SE, previously reported as enhancer module C [28], was particularly essential for *EVI1* transcription. Module C is reported to be responsible for *MYC* expression in primary leukemic cells. The high accessibility of this module and the binding of a core set of hematopoietic transcription factors drive *MYC* expression in HSPCs [28, 29]. The other reported elements in the *MYC* SE, which did not affect *EVI1* transcription in a t(3;8) setting, may well be responsible for *MYC* transcription in other tissues [28]. Module C is the only element within the *MYC* SE to which early hematopoietic regulators bind, including GATA2, FLY1, ERG, RUNX1, LMO2 and LYL1 [29]. Since those factors also bind to other enhancers that recurrently translocate to *EVI1* in t(2;3)(p21;q26), t(3;7)(q26;q24), t(3;6)(q26;q11) or inv(3)/t(3;3)(3q26;3q21) AML, we argue that *EVI1* expression is driven by a common mechanism. This is in line with our previous published data on a variety of atypical 3q26-rearranged AMLs [20]. The loci donating their enhancer to *EVI1* harbor genes that are normally expressed in early HSPCs, e.g. *MYC, ARID1B, CDK6, THADA* or *GATA2.* Leukemias with high *EVI1* levels are chemotherapy-resistant and exhibit a unique gene expression signature comparable to that of CD34+ HSPCs [30]. This suggests that the cell of origin transformed in these leukemias is a very primitive hematopoietic progenitor cell.

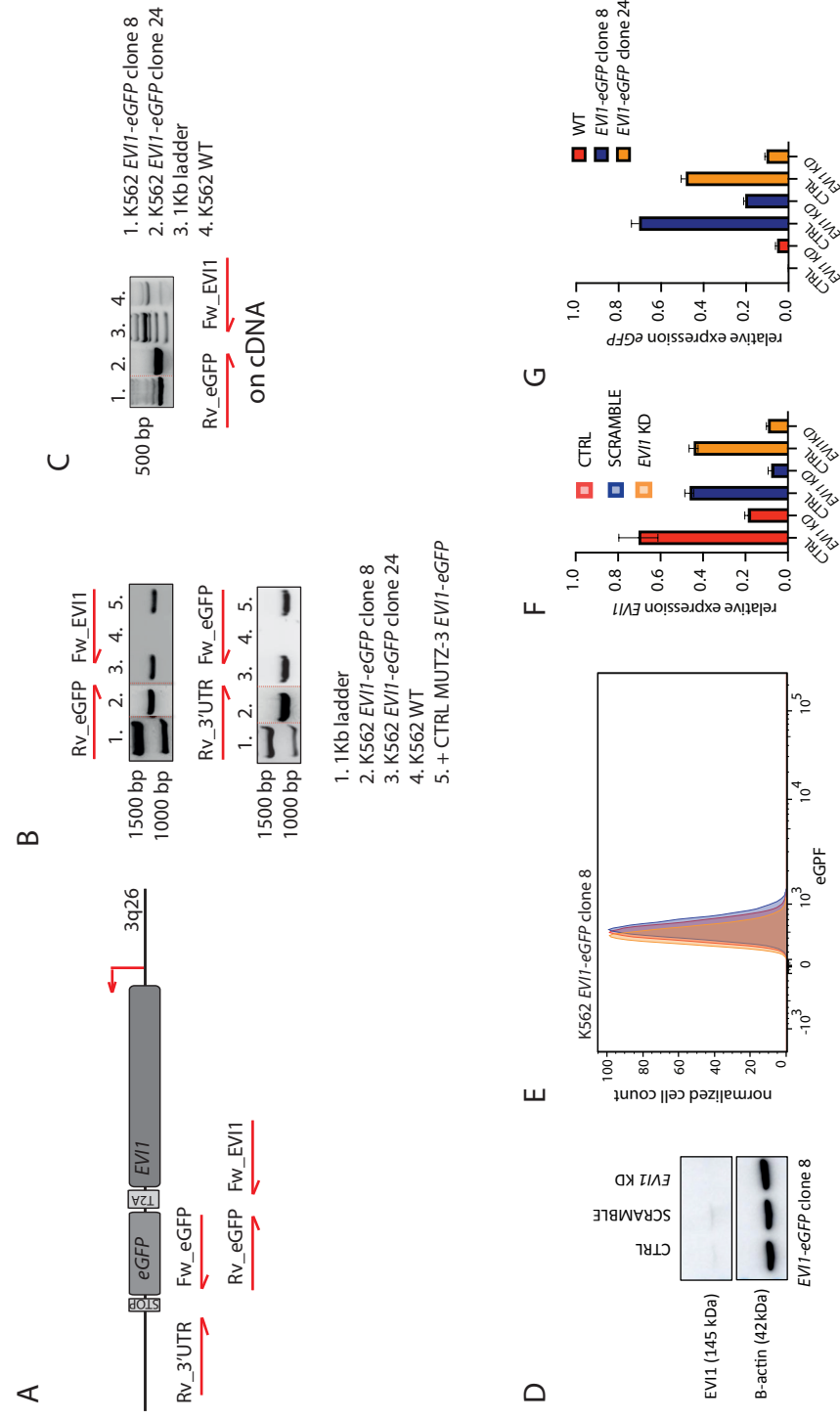The high-resolution 4C-seq data generated using our t(3;8) model revealed interaction of the *EVI1* promoter with the *MYC* SE, with multiple interaction zones associated with different enhancer modules indicative of a highly organized SE. Accordingly, Huang et al. defined the *MYC* SE as hierarchically organized. A hierarchical SE contains an enhancer element, referred to as hub enhancer, which is responsible for structural organization of the SE and is distinctly associated with CTCF and cohesin binding [7]. Module C was characterized as a hub enhancer within the *MYC* SE in K562 cells [7]. Interestingly, the deletion of module C, while leaving CTCF binding sites intact, not only affected *EVI1* expression but also disrupted *MYC* SE-*EVI1* promoter interaction. This suggested that transcription factors and co-activators occupying this location play a role in enhancer-promoter interaction, either independently or in cooperation with CTCF. Analogous to CTCF, YY1 contributes to DNA-looping, but preferentially occupies interacting enhancers and promoters [8]. Although there is no indication that YY1 binds directly to enhancer element C, we did find YY1 binding flanking this element (Figure S7). In embryonic stem cells (ESC), pluripotency factors e.g. OCT4, NANOG and SOX2 recruit the mediator complex and stabilize the cohesin complex in order to facilitate cell type specific non-CTCF mediated enhancer-promoter looping [31]. In HSPCs a subunit of the mediator complex, MED12, co-localizes with key hematopoietic transcription factors, interacting with additional transcriptional co-activators to maintain enhancer activity [32]. We hypothesize that in t(3;8) and other 3q26-rearranged AMLs, enhancer-promoter interaction is facilitated by CTCF and cohesin, which is further stabilized by recruitment of co-factors by hematopoietic regulators.

All CTCF binding motifs in the *MYC* SE are oriented in a 'reverse' fashion, allowing a CTCF/cohesin complex to be formed with the 'forward' CTCF binding site 2.6 kb upstream of the *EVI1* transcriptional start site (TSS). In all 3q26-rearranged AMLs this upstream CTCF binding site was preserved with respect to *EVI1*. Interestingly, a CTCF binding site upstream of the *MYC* promoter has been reported to function as a docking site for enhancers driving *MYC* expression [6]. Our findings point to a highly similar mechanism of transcriptional activation of *EVI1* in 3q26-rearranged AML. CTCF binding at this site proved to be absolutely critical for enhancer-promoter interaction and consequently indispensable for enhancer driven *EVI1* transcription.

Leukemias with 3q26 rearrangements depend on *EVI1*: interfering with EVI1 causes growth inhibition, differentiation and ultimately death of leukemic cells [15, 33]. Our data demonstrate mechanistic similarities between the distinct enhancer-driven *EVI1+* leukemias, suggesting that a therapy for one subtype may be effective for all these AMLs. The *EVI1-eGFP* t(3;8) model is a valuable tool for compound screens to identify inhibitors of *EVI1* transcription that could constitute a promising treatment for these refractory leukemias. As enhancer-driven transcription is not limited to leukemia, this model can also be used to study (super-) enhancer biology and transcriptional regulation in a broader context.

Figure S1. Generation of cell model with eGFP reporter for *EVI1* expression

(A Schematic overview of PCR primers located in the *eGFP-T2A-EVI1* insert. (B) PCRs of genomic DNA of K562 *EVI1-eGFP* clones to verify correct genotype. (C) PCR on cDNA to verify *EVI1-eGFP* transcript in the same clones. (D Western blot showing very low EVI1 protein levels in the controls as expected, and absent EVI1 protein levels upon *EVI1* knockdown by shRNA. (E) Flow cytometry plot showing eGFP reduction upon *EVI1* knockdown. (F) *EVI1* expression levels relative to *PBGD* by qPCR upon *EVI1* knockdown in two different K562 *EVI1-eGFP* clones (clones 8 and 24, P<0.0001) and WT K562 (P<0.0001). (G) *eGFP* expression levels relative to *PBGD* by qPCR upon *EVI1* knockdown in two different K562 *EVI1-eGFP* clones (clones 8 and 24, P<0.0001) and WT K562 (not significant).

A

Chr.3:168 918 020

Chr.8:130 487 250

t(3;8) clone 8-3

t(3;8) clone 8-4

t(3;8) clone 24-7

t(3;8) clone 24-12

B

Chr.3

Chr.8

3q26

EVI1

8q24

MYC

3 color MECOM

MYC

centromere 8

breakpoint

C

EVI1-eGFP t(3;8) clone 8-3

3 color MECOM

D

EVI1-eGFP t(3;8) clone 8-3

MYC

E

EVI1-eGFP t(3;8) clone 8-3

centromere 8

3 color MECOM

F

EVI1 (145 kDa)

B-actin (42 kDa)

EVI1-eGFP t(3;8) clone 8-4

CTRL   SCRAMBLE   EVI1 KD

G

EVI1-eGFP t(3;8) clone 8-4

CTRL
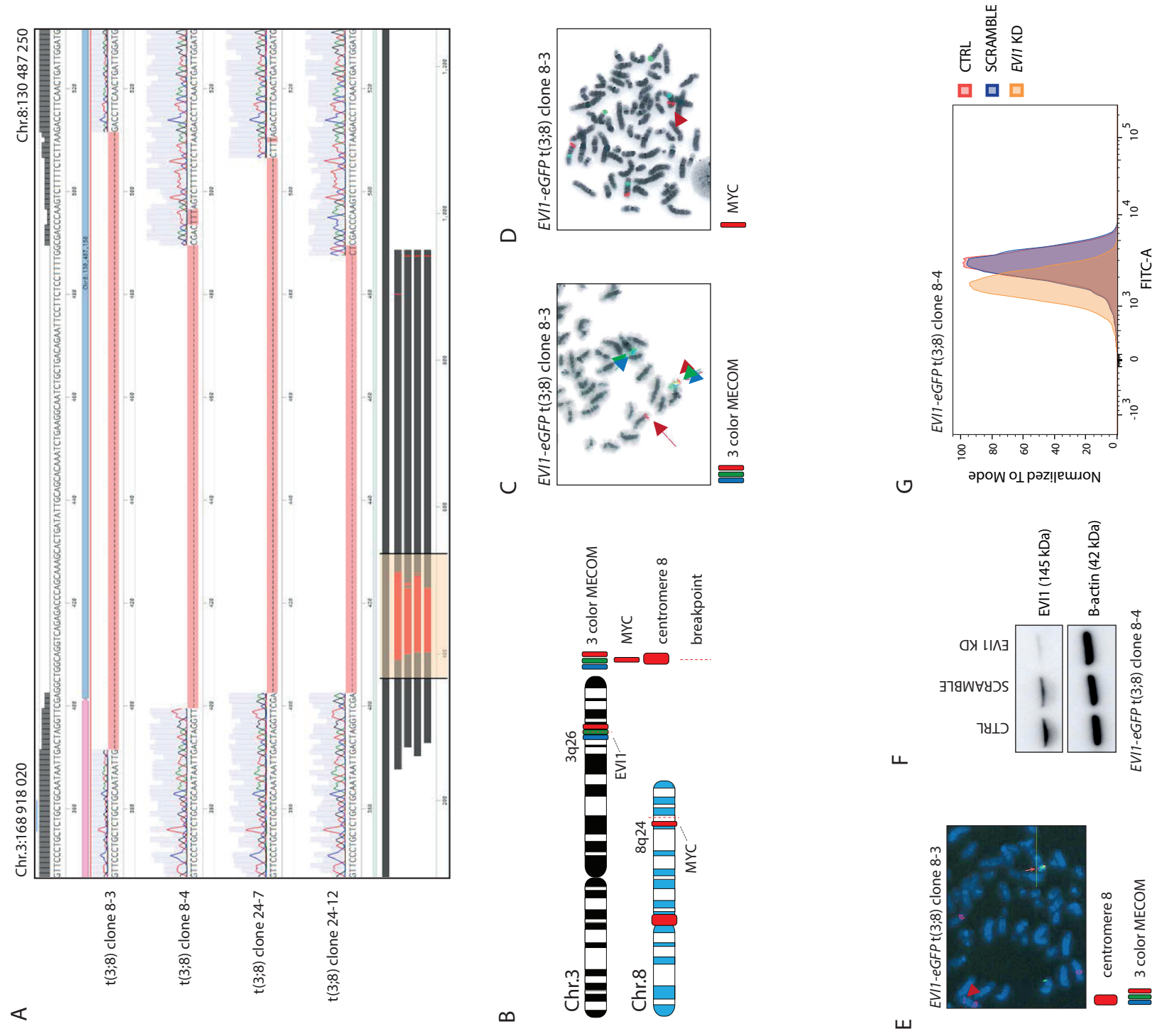SCRAMBLE
EVI1 KD

Normalized To Mode

FITC-A

Figure S2. A t(3;8) cell model recapitulates *EVI1* overexpression in human AML

(A) Sanger sequencing (amplicon covering the breakpoint as shown in Figure 2G), to validate the generation of a t(3;8). Nucleotide sequencing data of all four K562 *EVI1-eGFP* t(3;8) clones is shown. Using the online tool BLAT (Kent, 2002) the left part of the sequence maps back to 3q26.2 (pink) and the right part maps to 8q24 (blue). About 100bp were deleted on the Chr.8 side of the breakpoint in the generation of the translocation (depicted below in red). **(B)** Schematic overview of diagnostic FISH experiments performed to validate the presence of t(3;8) in the four clones. Fluorescent FISH pictures of K562 t(3;8) clone 8-3 are in C-E are shown. **(C)** Detection of t(3;8) by FISH. *MECOM* was split and the red signal was separated from the blue and green, indicating a translocation of the distal part of the q arm of chromosome 3. **(D)** Fluorescent images obtained by FISH to detect t(3;8). Three red signals represent the *MYC* gene on chromosomes 8 (3 copies of Chr.8 in K562), demonstrating that *MYC* gene is unaffected. The longer tip of Chr.8 near the *MYC* signal (arrow) is in line with a rearrangement at this chromosome. **(E)** Fluorescent images obtained by FISH to detect t(3;8). The three Chr.8s can be visualized with the bigger red signal at the centromeres; the red part of the separated *MECOM* probe is located at the q-arm of one of the Chr.8s (arrow). Together the FISH images demonstrate that part of *MECOM* had been translocated to Chr.8, forming a t(3;8)(q26;q24). **(F)** Western blot showing high EVI1 protein levels for t(3;8) clone 8-4, and a reduction of EVI1 levels upon *EVI1* knockdown using a shRNA. **(G)** Flow cytometry plot showing eGFP reduction upon *EVI1* knockdown.
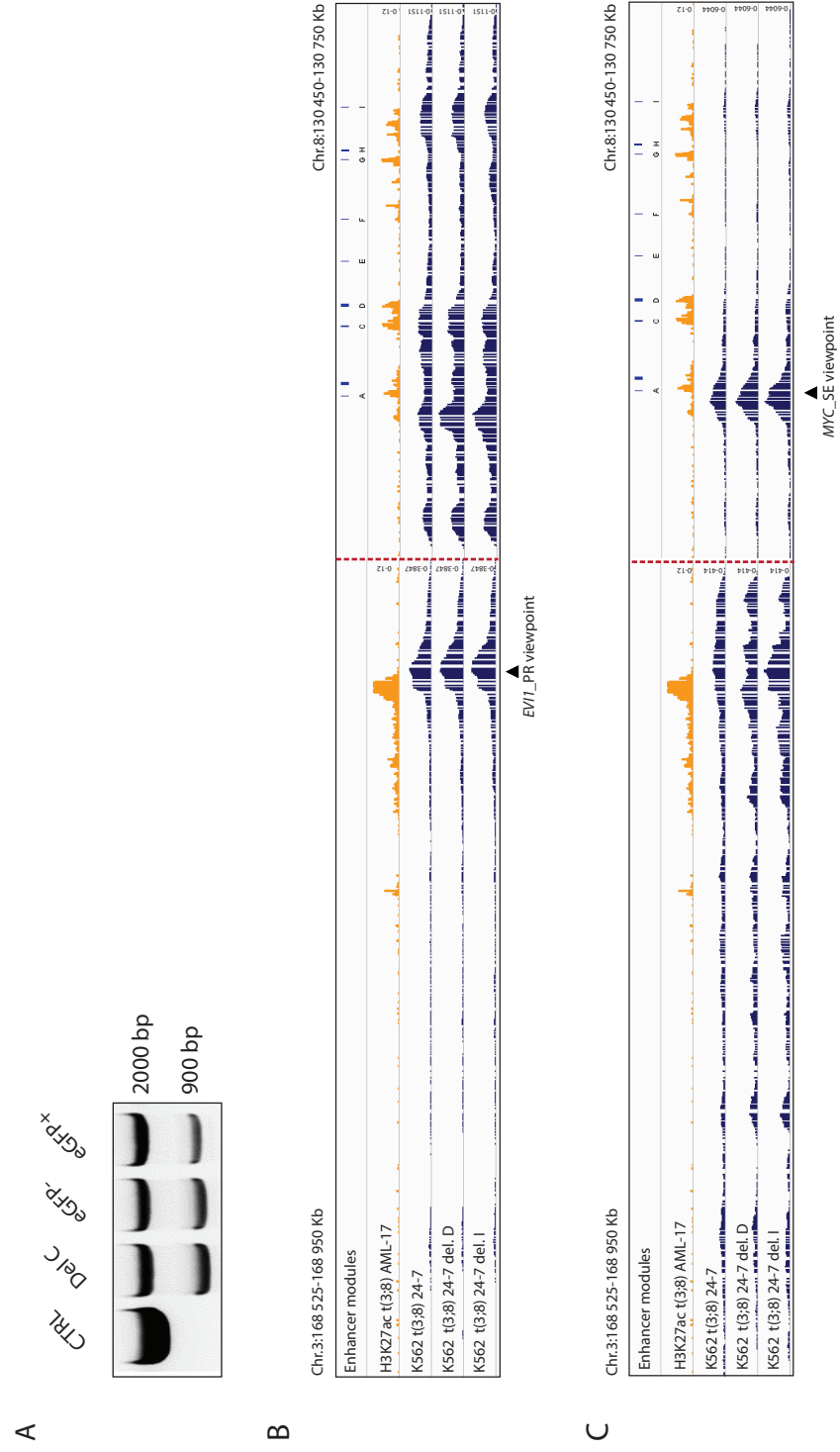
Figure S3. One critical enhancer element in the *MYC* SE drives *EVI1* transcription

(A) PCR showing the deletion of enhancer element C present in the bulk cells, eGFP- and eGFP+ fractions, but not in the controls. The deletion observed by the smaller PCR band in the eGFP+ fraction can be explained by deletions in the non-translocated Chr.8 allele not influencing *EVI1* expression, and thus not eGFP. Nevertheless, clearly more cells with the deletion are found in the eGFP- fraction. (B) No significant difference in *MYC* expression levels relative to *PBGD* by qPCR between the parental K562 *EVI1-eGFP* clones and K562 *EVI1-eGFP* t(3;8) clones. (C) 4C-seq data (t(3;8) clone 24-7) with the *EVI1* promoter as viewpoint (black triangle). No structural changes were observed following the deletions of *MYC* SE modules D or I (Figure 4B). (D) 4C-seq data (t(3;8) clone 24-7) with viewpoint in the *MYC* SE (black triangle) showing the same as D. No structural changes upon deletion of *MYC* SE modules D or I were observed (Figure 4B).
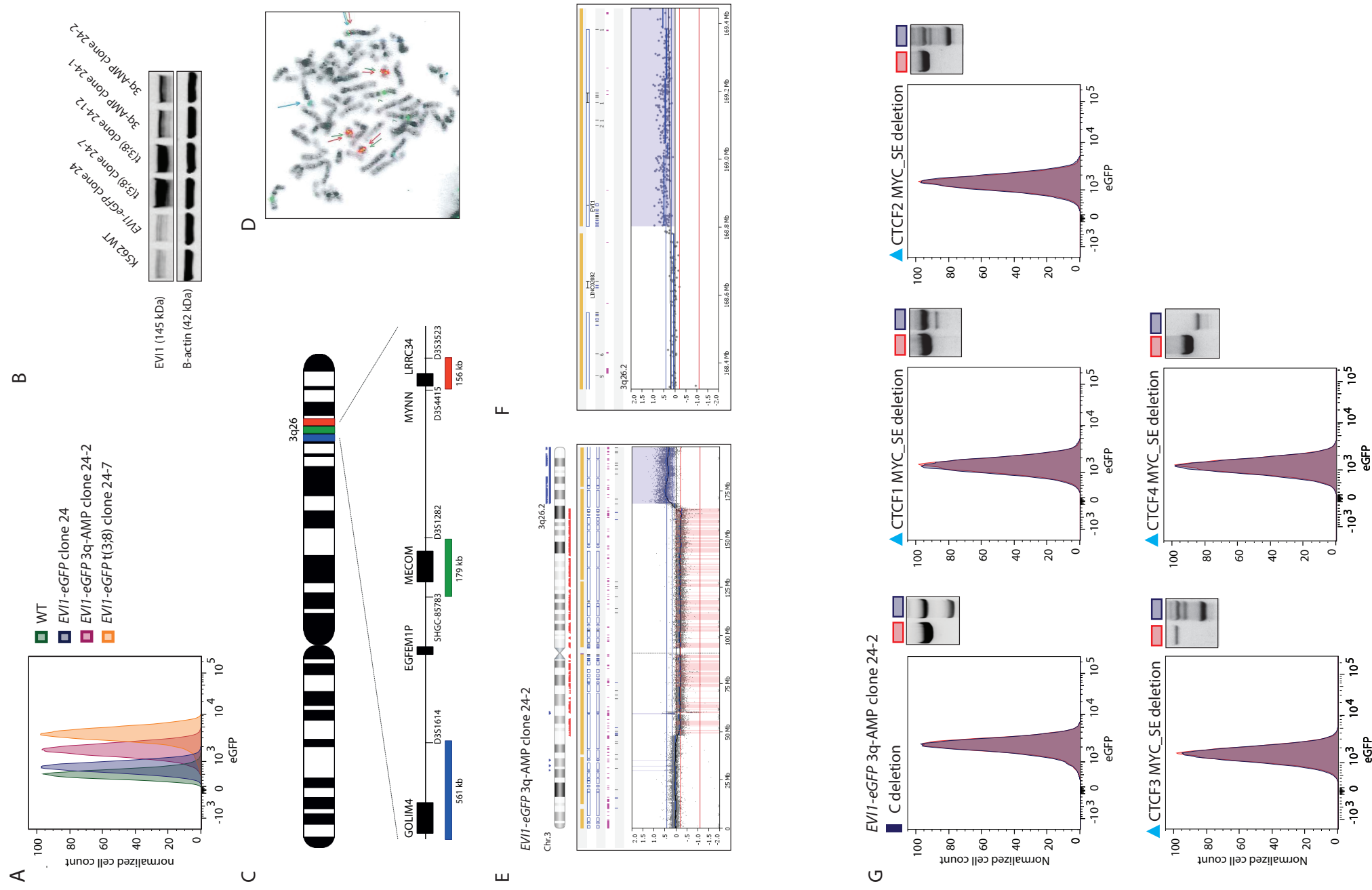
Figure S4. To control experiments a *eGFP-EVI1* reporter cell model with 3q/*MECOM* amplification was used

(A) Flow cytometry plot comparing eGFP levels of *EVI1-eGFP* K562 with 3q/*MECOM* amplification (*EVI1*-AMP) clone 24-2 to WT, parental (*EVI1-eGFP*) and t(3;8) clone 24-7 K562 cells. **(B)** Western blot comparing EVI1 protein levels of *EVI1*-AMP clones 24-1 and 24-2 to WT, parental (*EVI1-eGFP*) and t(3;8) clones 24-7 and 24-12. **(C)** Schematic overview of *MECOM* (*EVI1*) tri-color FISH probes. **(D)** *MECOM* (*EVI1*) FISH for clone 24-2 illustrating the *EVI1* amplification. **(E)** Overview of Chr.3 SNP array data showing high copy number for the q arm of Chr.3 starting from 3q26.2 in clone 24-2. **(F)** SNP array data: zoom-in on the 3q26.2 locus of clone 24-2 indicating the breakpoint/start of the amplification. The amplification includes exactly the complete *EVI1* (*MECOM*) locus. We estimated clone 24-2 has 4 copies of this part of the 3q arm including *EVI1*, resulting in elevated *EVI1* expression and protein levels. **(G)** Clone 24-2 was used as a control clone with high *EVI1* expression but without a t(3;8). All genomic deletions in the *MYC* SE (Figure 4A and 5A) made in this study were also successfully performed in clone 24-2 as shown here by PCR. Flow analysis showed that none of the deletions produced changes in *EVI1* expression in this clone.

A

K562 *EVI1-eGFP* t(3;8) clone 8-4

▼ CTCF2 MYC_SE deletion

Chr.3:168 625-168 950 Kb   Chr.8:130 475-130 775 Kb

H3K27ac   0-12   0-12

eGFP+   0-173   0-3492

eGFP-   0-173   0-3492

▲ MYC_SE viewpoint

B

K562 *EVI1-eGFP* t(3;8) clone 24-7

▲ CTCF EVI1_PR deletion

Chr.3:168 625-168 950 Kb   Chr.8:130 475-130 775 Kb

H3K27ac   0-12   0-12

eGFP+   0-208   0-3928

eGFP-   0-208   0-3928

▲ MYC_SE viewpoint

C

K562 *EVI1-eGFP* t(3;8) clone 24-7

▲ CTCF EVI1_PR sgRNA#1

Chr.3:168 625-168 950 Kb   Chr.8:130 475-130 775 Kb

H3K27ac   0-12   0-12

eGFP+   0-181   0-4008

eGFP-   0-181   0-4008

▲ MYC_SE viewpoint

Figure S5. CTCF binding site upstream of the *EVI1* promoter hijacks a *MYC* SE in t(3;8) AML.
(A) 4C-seq data (t(3;8) clone 8-4), with the *MYC* SE as viewpoint, of cells with a CTCF2 *MYC* SE deletion and sorted on eGFP. In the eGFP- fraction, a loss of interaction with the *EVI1* promoter was observed compared to the eGFP+ cells. **(B)** 4C-seq data (t(3;8) clone 24-7), with the *MYC* SE as viewpoint, of cells with a CTCF *EVI1* promoter deletion and sorted on eGFP. In the eGFP- fraction, a loss of interaction with the *EVI1* promoter was observed compared to the eGFP+ cells. **(C)** Comparison of chromatin interaction at the *EVI1* promoter for eGFP+ and eGFP- cells, shown by 4C-seq (t(3;8) clone 24-7) with the *MYC* SE as viewpoint after targeting the CTCF motif with sgRNA#1 as presented in Figure 6D.
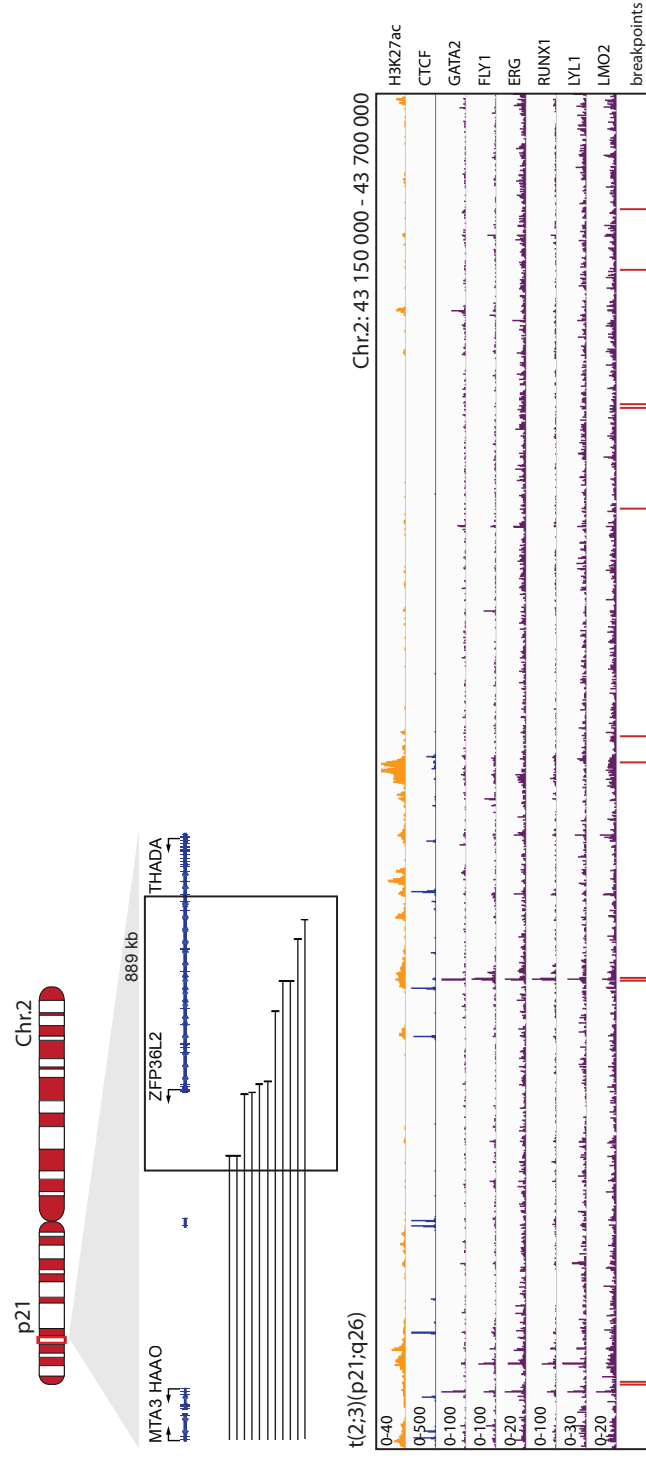
Figure S6. Translocated region in t(2;3)(p21;q26) AML contains strong regulatory elements

Translocated locus in t(2;3)(p21;q26) AML showing the presence of the gene *THADA*, the black box indicating the zoom-in shown below. Zoom-in: ChIP-seq for H3K27ac (t(3;8) patient AML-17, orange) indicating putative enhancer regions, CTCF binding and HSPC-active transcription factor recruitment. The red lines below indicate the exact breakpoint of the t(2;3) AMLs (detected by 3q-seq). We predict that the translocated region of each of these cases could be unique, but in all cases contains a strong regulatory locus.
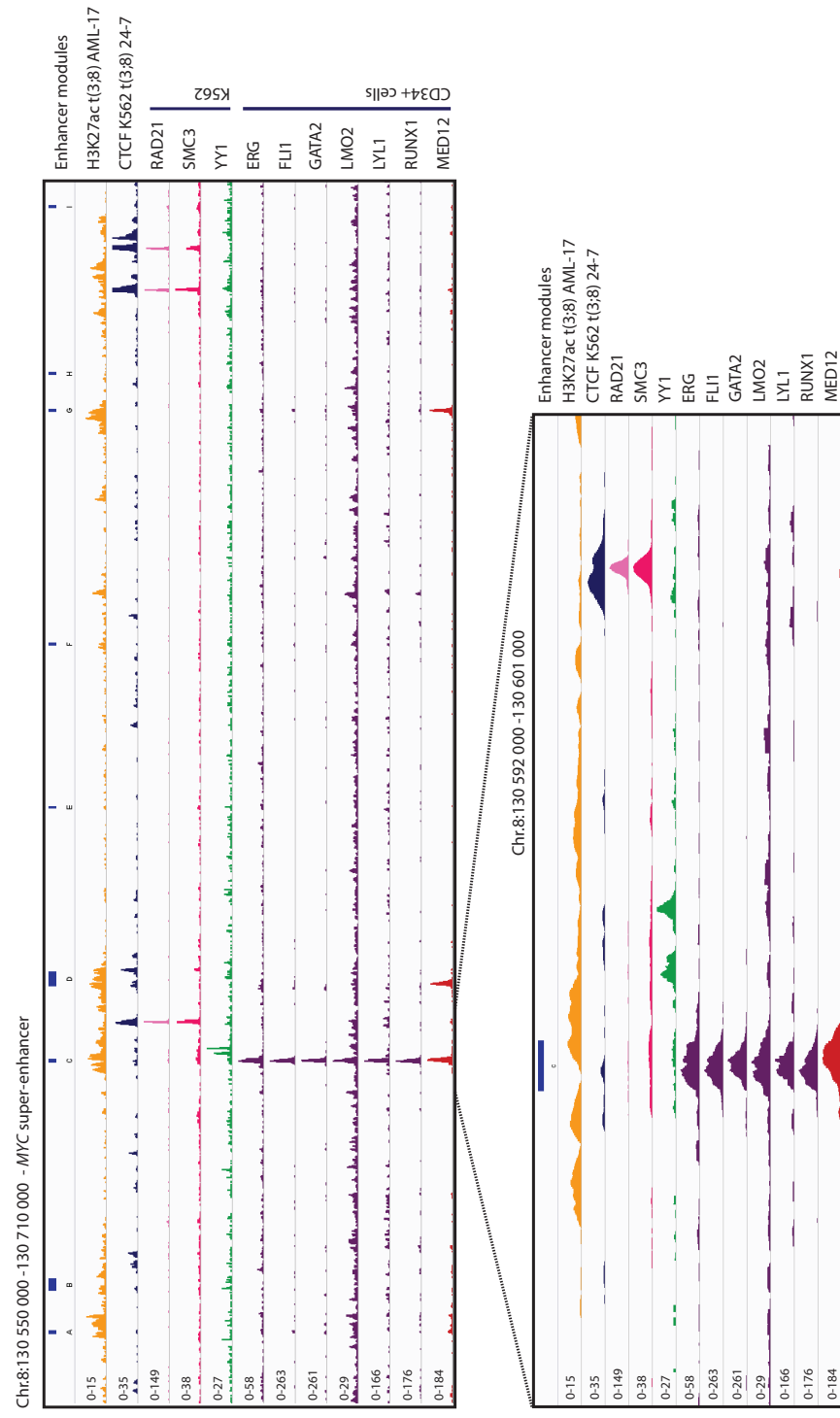
Figure S7. Transcription factor and transcriptional co-factor occupation of the *MYC* SE

ChIP-seq data is shown, in the upper panel the *MYC* SE and the enhancer elements are illustrated by H3K27ac (orange, t(3;8) AML-17 patient). In the lower panel a zoom-in on enhancer element C is shown. Data were retrieved RAD21, SMC2, YY1 (ENCODE (Davis et al., 2018; Encode Project Consortium, 2012)), ERG, FLI1, GATA2, LMO2, LYL1, RUNX1(Beck et al., 2013), MED12 (Aranda-Orgilles et al., 2016).

MATERIAL AND METHODS

Patient material

AML patient samples were collected either from the Erasmus MC Hematology department biobank (Rotterdam, The Netherlands) or from the MLL Munich Leukemia Laboratory biobank (Munich, Germany). Leukemic blast cells were purified from bone marrow or blood by standard diagnostic procedures. All patients provided written informed consent in accordance with the Declaration of Helsinki.

Generation of *EVI1* expression cell model

The plasmids to clone T2A-eGFP in frame with *EVI1* were designed and described by my colleagues as follows [33]. The repair template was generated using Gibson Assembly (NEB). Both homology arms were PCR amplified from MUTZ3 genomic DNA using Q5 polymerase (NEB). The first homology arm consists of a part of the intron and last exon of *EVI1* minus the STOP, the second homology arm consists of part of the 3'UTR with the PAM sequence of sgRNA omitted. The *T2A-eGFP* was PCR amplified from dCAS9-VP64_2A_GFP. All fragments were cloned using the Gibson assembly into the PUC19 backbone. The sgRNA sequence AGCCACGTATGACGTTATCA was cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9. Cells were nucleofected with pX330 vector containing the sgRNA and Cas9 and the repair template using the NEON transfection system (Thermo Fisher) with buffer R and program 3 (1350 V, 10 ms, 4 pulses). GFP$^+$ cells were sorted using a FACS AriaIII (BD Biosciences), and after two rounds of enrichment for cells expressing eGFP+, these cells were single cell sorted and tested for proper integration. Subsequently, clones were named K562 *EVI1-eGFP*; multiple clones were obtained, but in this study only clone 8 and 24 were used for further experiments.

Generation of a t(3;8)(q24;q26) model

K562 *EVI1-eGFP* clones (clone 8 and clone 24) were used as parental clones to generate the t(3;8)(q24;q26) clones. Based on the breakpoints (Chr.3:168.917.999 - Chr.8:130.487.191) of primary AML sample (#HF-80), sgRNAs were designed (using ChopChop V3 [34], Table S1) and mixed with purified Cas9 (IDT) to make ribonucleoproteins (RNPs). The NEON transfection system (Thermo Fisher) was used to get the RNPs into the K562 *EVI1-eGFP* clones. Three days after transfection the eGFP+ cells were sorted using the FACS AriaIII, and this enrichment process was repeated twice before eGFP+ single cells were sorted to produce single cell clones. The clones were characterized for the designed specific t(3;8)(q24;q26) translocation by PCR (primers in Table S1), Sanger-seq, flow cytometry and FISH.

Cytogenetics: karyotype and FISH

Diagnostic cytogenetics for all samples was performed by each of the institutes mentioned above. For this study, samples were selected based on t(3;8)(q26;q24) rearrangements detected by karyotyping and/or *MECOM* interphase fluorescence *in situ* hybridization (FISH). FISH and classic metaphase karyotyping were performed and reported according to standard protocols based on the International System of Human Cytogenetics Nomenclature (ISCN) 2017 [35]. For both patient samples and K562 clones *MECOM* FISH was performed according to the manufacturer's protocol, using the *MECOM* t(3;3); inv(3)(3q26) triple color probe (blue, green, red, Cytocell, LPH-036). For the characterization of the K562 *EVI1-eGFP* t(3;8) clones the *MECOM* FISH was combined with: CEP8 (cen.8, blue), IGH (14q32, green), C-MYC(8q24, orange) (Vysis, 04N10-020) and C8 (Vysis, SpO, 07J22-008).

Targeted chromosomal region 3q21.1-3q26.2 DNA sequencing (3q-seq)

3q-seq and data analysis have been performed as previously described [15].

RNA isolation, quantitative PCR (qPCR) and RNA sequencing

RNA isolation, qPCR an RNA-seq and analysis of this data have been performed as described in our previous publication [20]. For qPCR data one-way ANOVA (PRISM) was performed to indicate level of significant differences between clones or conditions. For qPCR data of cells directly after FACS no statistical test could be performed due to the limited number of cells (Figure 4 C-D).

Cell lines and culture

K562 cells were cultured in RPMI 1640 + L-glutamine (Hyclone SH30027.LS), 10% fetal calf serum (FCS, Gibco) and 50 U/mL penicillin and 50 μg/mL streptomycin (Gibco 15140-163). Cells were incubated at 37°C and 5% CO_2 and passaged every 3-4 days to 100.000 cells/ml.

Genome editing

CRISPR-Cas9 technology was used to make mutations or deletions in the regions described in the results section. All primer sequences to generate sgRNAs can be found in Table S1 and were ordered from IDT. By *in vitro* transcription sgRNAs were produced as described above for the generation of the t(3;8). In short: the constant and specific oligos were annealed and filled in 20 min 12°C by T4 polymerase (NEB, M0203S), sgRNAs were produced by *in vitro* transcription using HIScribe T7 High-Yield RNA Synthesis kit (NEB, E2050S) 3-4h, 37°C, DNA was eliminated by Turbo DNase (Thermo Fisher, AM2238), 15min, 37°C. The sgRNas were concentrated and purified using RNA clean & concentrator -25 (Zymo Research, R1017). The concentration of sgRNAs was estimated using Qubit RNA BR assay (Invitrogen, Q10210). Ribonucleoproteins (RNPs) were made by mixing purified Cas9 protein (IDT, Alt-R S.p. Cas9 Nuclease V3, 1081058) with the sgRNAs, 20-30 min at RT.

For all K562 clones the NEON buffer T and same settings (1350V, 10ms, 4 pulses) were used. After a minimum of 72 hrs post nucleofection DNA or RNA was extracted (DNA Quick

extract, Epicenter or Qiagen Allprep DNA/RNA mini, #80204) or cells were harvested for further analysis by respectively PCR, qPCR or flow cytometry analysis/FACS sorting.

Flow cytometry and sorting (FACS)

Flow cytometric analysis or cell sorting was performed using the FACS Canto or the FACS Aria flow cytometer (BD Biosciences). Cells were gated on viability and single cells using FSC/SSC, eGFP intensity levels were measured using the FITC channel. Data was analyzed using FlowJo.

PCR and primers

For all PCRs used to detect translocations, point mutations or deletions; Q5 High-Fidelity DNA polymerase was used following the manufactures protocol (NEB, #M0491). Sanger sequencing was performed using the BigDye direct Sanger sequencing kit (Thermo Fisher) by the principles first described in literature [36]. Primer sequences can be found in Table S1.

Amplicon sequencing

To check mutations after targeting with CRISPR-Cas9 we performed amplicon sequencing using the Illumina PCR-based custom amplicon sequencing method using the TruSeq Custom Amplicon index kit (Illumina). The first PCR was performed using Q5 polymerase (NEB), the second nested PCR with KAPA HiFi HotStart Ready mix (Roche). Samples were sequenced paired-end (2x 250bp) on a MiSeq (Illumina). Reads were trimmed with trimgalore [37] to remove low-quality bases and adapters, and subsequently aligned to the human reference genome build hg19 with BBMap [38] allowing for 1000 bp indels. Mutations introduced by genome editing were analyzed and visualized using CRISPResso2 [39].

Western Blotting

Cells were lysed using the NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific) following the manufacturer' protocol and nuclear extract was used for Western Blotting of EVI1 (#2265 Cell Signalling). As loading control an antibody against B-Actin (clone AC15, A5441, Sigma) was used. The Odyssey infrared imaging system (Li-Cor) was used for visualization of the protein levels.

4C sequencing

Chromosome Conformation Capture Sequencing (4C-seq) sample preparation was performed as previously described (van de Werken et al., 2012). In short, genomic regions that are spatially proximal in the cell nucleus were fixated by formaldehyde-induced crosslinks. The DNA was fragmented with DpnII as a primary restriction enzyme, Csp6I as a secondary 4 bp-cutter. To identify and quantify fragments that were ligated to the genomic region of interest, a two-step PCR was performed as previously described (Krijger et al., 2020).

The first PCR step was an inverse PCR with viewpoint-specific primers that are listed in Table S1. In the second PCR step, universal primers were used that contain the Illumina adapters. The amplicons were analyzed using next generation sequencing on the IIlumina NovaSeq6000 platform. Data analysis was performed as described previously by our group [15]. For visualization, 4C-seq data were normalized to reads per million (RPM) and a running mean was applied for smoothing. In all figures, the tracks were displayed on IGV using group auto-scale to compare relevant samples.

ChIP sequencing

Chromatin immunoprecipitation sequencing (ChIP-seq) experiments were performed as previously described [15, 40, 41].

Immunoprecipitation of crosslinked chromatin was performed with antibodies directed against H3K27Ac (Diagenode C15410196), H3K9Ac (Diagenode C15410004), H3K4me3 (Diagenode C15410003) or CTCF (Cell Signalling, 2899S). Crosslinks were reversed overnight at 65°C in the presence of proteinase K (New England Biolabs). De-crosslinked material was purified using a QIAGEN PCR Purification Kit. The purified DNA was processed according to the Nextflex ChIP Sample Preparation Protocol (Perkin Elmer) or the Microplex library preparation kit V2 (Diagnode C05010013) and sequenced on the Illumina NovaSeq6000 platform. Reads were aligned to the human genome build 19 (hg19). Processing and analysis of the data were done as described previously [42, 43]. RAD21 (YALE) and SMC3 (ab9263) ChIP-seq tracks in K562 cells were downloaded from ENCODE [44, 45]. In all figures displaying ChIP-seq data the y-axis shows normalized RPKM, when comparing samples group auto-scale was used on IGV to show differences.

SNP array

DNA was isolated from K562 cells using the AllPrep DNA/RNA mini kit (Qiagen, #80204). All SNP arrays were performed at the Erasmus MC Department of Clinical Genetics (Rotterdam, The Netherlands) and analyzed as previously described [20, 46, 47].

Accession numbers

The ChIP-seq, 3q-seq, 4C-seq and RNA-seq data are available at the EGA repository under the accession numbers EGAS00001004808.

Author contribution

S.O., L.S and R.D. conceived of and designed the study. B.W., C.H. and T.H. provided study materials or patient samples. H.B.B supervised all cytogenetic and FISH characterization. C.E., S.v.H., M.H., A.A.V., L.S., and S.O. performed experiments provided data analysis and interpretation. M.V. operated the FACS sorter for a significant part of the experiments. E.B. organized all NGS and was involved in data interpretation. R.M-L. was responsible for all

bioinformatics data processing and analysis of this study. S.O., R.M-L., L.S. and R.D. wrote the manuscript.

Declaration of interest

T.H. and C.H. are employees of and have equity ownership in MLL Munich Leukemia Laboratory. The remaining authors declare no competing financial interests.

Acknowledgements

REFERENCES

1. Stadhouders, R., G.J. Filion, and T. Graf, *Transcription factors and 3D genome conformation in cell-fate decisions.* Nature, 2019. **569**(7756): p. 345-354.

2. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.

3. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.

4. Dowen, J.M., et al., *Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes.* Cell, 2014. **159**(2): p. 374-387.

5. Zuin, J., et al., *Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.* Proc Natl Acad Sci U S A, 2014. **111**(3): p. 996-1001.

6. Schuijers, J., et al., *Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism.* Cell Rep, 2018. **23**(2): p. 349-360.

7. Huang, J., et al., *Dissecting super-enhancer hierarchy based on chromatin interactions.* Nat Commun, 2018. **9**(1): p. 943.

8. Weintraub, A.S., et al., *YY1 Is a Structural Regulator of Enhancer-Promoter Loops.* Cell, 2017. **171**(7): p. 1573-1588 e28.

9. Bulger, M. and M. Groudine, *Functional and mechanistic diversity of distal transcription enhancers.* Cell, 2011. **144**(3): p. 327-39.

10. Spitz, F., *Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles.* Semin Cell Dev Biol, 2016. **57**: p. 57-67.

11. Muerdter, F. and A. Stark, *Gene Regulation: Activation through Space.* Curr Biol, 2016. **26**(19): p. R895-R898.

12. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

13. Mitelman, F., B. Johansson, and F. Mertens, *The impact of translocations and gene fusions on cancer causation.* Nat Rev Cancer, 2007. **7**(4): p. 233-45.

14. Yamazaki, H., et al., *A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression.* Cancer Cell, 2014. **25**(4): p. 415-27.

15. Groschel, S., et al., *A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia.* Cell, 2014. **157**(2): p. 369-381.

16. Lugthart, S., et al., *High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated.* Blood, 2008. **111**(8): p. 4329-4337.

17. Gröschel, S., et al., *High EVI1 Expression Predicts Outcome in Younger Adult Patients With Acute Myeloid Leukemia and Is Associated With Distinct Cytogenetic Abnormalities.* Journal of Clinical Oncology, 2010. **28**(12): p. 2101-2107.

18. Lugthart, S., et al., *Clinical, Molecular, and Prognostic Significance of WHO Type inv(3)(q21q26.2)/t(3;3) (q21;q26.2) and Various Other 3q Abnormalities in Acute Myeloid Leukemia.* Journal of Clinical Oncology, 2010. **28**(24): p. 3890-3898.

19. Barjesteh van Waalwijk van Doorn-Khosrovani, S., et al., *High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients.* Blood, 2003. **101**(3): p. 837-845.

20. Ottema, S., et al., *Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia.* Blood, 2020. **136**(2): p. 224-234.

21. Lin, P., et al., *Translocation (3;8)(q26;q24): a recurrent chromosomal abnormality in myelodysplastic syndrome and acute myeloid leukemia.* Cancer Genetics and Cytogenetics, 2006. **166**(1): p. 82-85.

22. Lennon, P.A., et al., *Aberrant EVI1 expression in acute myeloid leukemias associated with the t(3;8)(q26;q24).* Cancer Genetics and Cytogenetics, 2007. **177**(1): p. 37-42.

23. De Braekeleer, M., et al., *Breakpoint heterogeneity in (2;3)(p15–23;q26) translocations involving EVI1 in myeloid hemopathies.* Blood Cells, Molecules, and Diseases, 2015. **54**(2): p. 160-163.

24. Trubia, M., et al., *Characterization of a recurrent translocation t(2;3)(p15–22;q26) occurring in acute myeloid leukaemia.* Leukemia, 2006. **20**(1): p. 48-54.

25. Storlazzi, C.T., et al., *A novel chromosomal translocation t(3;7)(q26;q21) in myeloid leukemia resulting in overexpression of EVI1.* Annals of Hematology, 2004. **83**(2): p. 78-83.

26. Nucifora, G., L. Laricchia-Robbio, and V. Senyuk, *EVI1 and hematopoietic disorders: History and perspectives.* Gene, 2006. **368**: p. 1-11.

27. Tang, G., et al., *t(3;8)(q26.2;q24) Often Leads to MECOM/MYC Rearrangement and Is Commonly Associated with Therapy-Related Myeloid Neoplasms and/or Disease Progression.* J Mol Diagn, 2019. **21**(2): p. 343-351.

28. Bahr, C., et al., *A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies.* Nature, 2018. **553**(7689): p. 515-520.

29. Beck, D., et al., *Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes.* Blood, 2013. **122**(14): p. e12-22.

30. Valk, P.J.M., et al., *Prognostically Useful Gene-Expression Profiles in Acute Myeloid Leukemia.* New England Journal of Medicine, 2004. **350**(16): p. 1617-1628.

31. Denholtz, M. and K. Plath, *Pluripotency in 3D: genome organization in pluripotent cells.* Curr Opin Cell Biol, 2012. **24**(6): p. 793-801.

32. Aranda-Orgilles, B., et al., *MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis.* Cell Stem Cell, 2016. **19**(6): p. 784-799.

33. Smeenk, L., *An essential role for MYB in driving oncogenic EVI1 expression in enhancer-rearranged leukemias.* manuscript submitted for publication, 2020.

34. Labun, K., et al., *CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing.* Nucleic Acids Research, 2019. **47**(W1): p. W171-W174.

35. International Standing Committee on Human Cytogenomic, N., et al., *ISCN : an international system for human cytogenomic nomenclature (2016).* 2016.

36. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-5467.

37. Krueger, F., *Trim Galore.* 2012.

38. Bushnell, B., *BBMap short-read aligner, and other bioinformatics tools.* 2016.

39. Clement, K., et al., *CRISPResso2 provides accurate and rapid genome editing sequence analysis.* Nat Biotechnol, 2019. **37**(3): p. 224-226.

40. Bindels, E.M., et al., *EVI1 is critical for the pathogenesis of a subset of MLL-AF9-rearranged AMLs.* Blood, 2012. **119**(24): p. 5838-49.

41. Lovén, J., et al., *Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers.* Cell, 2013. **153**(2): p. 320-334.

42. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome biology, 2008. **9**(9): p. R137-R137.

43. Feng, J., et al., *Identifying ChIP-seq enrichment using MACS.* Nature Protocols, 2012. **7**(9): p. 1728-1740.

44. Davis, C.A., et al., *The Encyclopedia of DNA elements (ENCODE): data portal update.* Nucleic Acids Res, 2018. **46**(D1): p. D794-D801.

45. Encode Project Consortium, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

46. Srebniak, M., et al., *Application of SNP array for rapid prenatal diagnosis: implementation, genetic counselling and diagnostic flow.* European Journal Of Human Genetics, 2011. **19**: p. 1230.

47. Srebniak, M.I., et al., *Prenatal SNP array testing in 1000 fetuses with ultrasound anomalies: causative, unexpected and susceptibility CNVs.* European Journal Of Human Genetics, 2015. **24**: p. 645.

6

General discussion

GENERAL DISCUSSION

To keep a healthy balance between cell proliferation and differentiation, gene expression is highly regulated. Over the years, knowledge of gene control mechanisms has been gained and we particularly learned much from cases where errors in the genome had occurred which altered gene expression and led to disease. These mistakes can appear at various levels of gene control and in many different types of cells, leading to countless types and sub-types of diseases. In this thesis a subtype of leukemia was studied in which the control over the gene *EVI1* is severely affected. Comprehension of gene control and the loss of gene control mechanisms could be a step closer to improved treatment options for these specific patients. Besides that, mechanisms of gene regulation are not necessarily specific for a disease type, the models generated and knowledge gained in our studies may well be of general use.

Different hijacked enhancers, similar mechanisms of gene control

Many different chromosomal aberrations involving the *MECOM*/3q26 locus leading to *EVI1* overexpression have been identified. These include the most frequently found rearrangement: inv(3)/t(3;3) studied in Chapter 3 and 4, other recurrent translocations such as t(3;8) defined in Chapter 5, as well as the more unique reorganisations of this locus described in Chapter 2. Within the regions translocated to *EVI1* we found cases with strong regulatory elements present that were normally associated with genes expressed in early myeloid cells. In inv(3)/t(3;3) and t(3;8) AML, highly active enhancers originally belonging to the genes *GATA2* or *MYC* drive aberrant *EVI1* expression. For many other 3q26-rearranged AML cases we found evidence that likewise enhancers active in early hematopoiesis had been translocated to *EVI1*. Many of these enhancers are considered to be super-enhancers, characterized by large regions of open chromatin, determined by H3K27acetylation and binding of key hematopoietic transcription factors (e.g. GATA2, RUNX1, ERG) and transcriptional co-regulators like the histone acetylase p300 (Chapter 3 and 5). Correspondingly, the unbiased enhancer-screen using the inv(3) EVI1-eGFP model demonstrated in Chapter 3, showed mutated transcription factor binding motifs among the top scoring hits as being essential for *EVI1* transcription. We found that mutations in motifs known to bind factors such as RUNX, GATA2 or ERG indeed caused decreased *EVI1* levels. Knockdown or genome editing experiments to deplete each of the factors either alone or in combination would further help to uncover which of these factors are essential for *EVI1* transcriptional regulation. As an example, in an additional screen in which we depleted transcription/epigenetic factors using the same inv(3) EVI1-eGFP reporter system we observed that GATA2 itself was a very strong activator of *EVI1* transcription (unpublished observation L. Smeenk). Targeting *GATA2* in the t(3;8) EVI1-eGFP model (unpublished observation), demonstrated that GATA2 is also essential for *EVI1* transcriptional

regulation by the translocated *MYC* super-enhancer. The finding that these enhancers are interchangeable also suggests that factors we identify as being essential for *EVI1* activation in one of the 3q26-rearranged AMLs could also be critical for *EVI1* transcription in other. For instance, we found that MYB is essential for *EVI1* activation in inv(3)/t(3;3) AML by binding to the translocated *GATA2* enhancer (Chapter 3). Similarly, we found a MYB binding motif at the *MYC* super-enhancer in t(3;8) AML. When this motif was mutated using CRISPR-Cas9 in our t(3;8) EVI1-eGFP model (Chapter 5), *EVI1* transcription was disturbed (unpublished observation). Gaining information on the mechanisms of *EVI1* activation and transcription factor involvement in one 3q26-rearranged AML subtype, may provide insight into the role of these factors in other 3q26-rearranged AMLs. Further exploration of the models that we have generated and targeting a combination of transcription factors will provide further clarity about which factors together drive the expression of *EVI1* in these AMLs.

GATA2 haploinsufficiency in 3q26-rearranged AML

Besides activation of *EVI1*, loss of expression of the gene donating its enhancer to *EVI1* might be important for leukemic transformation. This has particularly been suggested for *GATA2* in inv(3)/t(3;3) AML [1, 2]. The question is whether the loss of *GATA2* expression at one allele as a consequence of *GATA2* enhancer hijacking by *EVI1,* has indeed an additive effect on *EVI1* driven leukemic transformation. Biological evidence for a cooperating activity of mono-allelic *GATA2* with *EVI1* overexpression in leukemia development was provided by a study from the group of Yamamoto in an *in vivo* mouse model [3]. They showed that overexpression of *EVI1* in bone marrow cells that were *GATA2* heterozygous were more aggressive in leukemia development then bone marrow cells from wild type animals that overexpressed the *EVI1* gene. If loss of one *GATA2* allele is of importance for *EVI1* transcriptional regulation, one would expect that *GATA2* expression is also affected in AML in other *EVI1* overexpressing 3q26-rearranged leukemias. Indeed, in our study presented in Chapter 2, the hypothesis that allele-specific expression of *GATA2* has an additive value to EVI1 driven leukemic transformation was further supported [4]. In multiple atypical 3q26-rearranged AMLs, we found copy number losses in *GATA2* or in regulatory elements of the gene, consequently leading to loss of *GATA2* expression at one allele. In other recurrent 3q26 rearranged AMLs we found some patients samples with skewed *GATA2* expression as well (e.g. in a t(2;3)(q22;q26), t(3;6)(q26;q25), t(3;21)(q26;q22) and inv(3)(p24;q26); personal communication with L. Smeenk. Investigating more cases might reveal more frequent allele-specific *GATA2* expression in these types of AML. Although loss of one GATA2 allele is a common theme in 3q26 rearranged AMLs, *GATA2* is at the same time essential for abnormal *EVI1* expression. We evidently found a loss of *EVI1* expression in both the inv(3) as the t(3;8) EVI1-eGFP model when *GATA2* was knocked out using CRISPR-Cas9 technology (preliminary data, not shown). Clearly a tight balance of *GATA2* expression appears important for the full transformation by *EVI1*. Decreased *GATA2* levels are either important for *EVI1* activation or

function, or it provides an important condition for the myeloid progenitor cells to develop leukemia in general. If the latter is the case, *GATA2* haploinsufficiency may also occur in other types of AML. In a recent study, Mulet-Lazaro demonstrated that mono-allelic expression of *GATA2* indeed occurs in more than 60% of all AML patients [5].

Besides *GATA2*, it is also possible that other genes, particularly the ones that donate their enhancer to *EVI1* in 3q26 rearranged AMLs, add to leukemic transformation. We observed skewed allelic expression of genes at the translocated region for some 3q26-rearranged samples. For example, we detected allelic expression of *CDK6* in AML with a translocation t(3;6)(q26;q25) and of *SATB1* in an inv(3)(P24;q26) AML (personal communication L. Smeenk). Furthermore, we observed allele specific expression of *PROM1* in t(3;4)(q26;p15) AML and *CD164* in ins(6;3)(q26;q21;q26) AML (Chapter 2) [4]. Whether the skewed-allelic expression of these genes has an additive effect on leukemic transformation remains unanswered for now and requires more samples to be analyzed and models to be built. Based on this all, we conclude that in particular allelic expression of *GATA2* has a proven additive effect on the development of AML, in particular in leukemias with a 3q26 rearrangement and consequently *EVI1* expressing. Why *GATA2* haploinsufficiency is so important in many AMLs is unclear.

Super-enhancer specificity: MYB drives EVI1 but not GATA2 in inv(3) AML

In AML with inv(3)/t(3;3), a distal enhancer of *GATA2* was translocated to *EVI1*. Upon translocation, the *GATA2* enhancer activity has been amplified and the enhancer turned into a super-enhancer as shown by an increased H3K27acetylation [1]. That the translocated enhancer differed from the non-translocated enhancer of *GATA2* was further exploit in Chapter 3. The region of open H3K27acetylated chromatin was enlarged and transcription factor occupancy at the translocated enhancer was significantly increased, as it was shown in particular for the transcriptional regulator MYB (Chapter 3). The sequences of events leading to the hyperactivity of the translocated enhancer remain unknown. Is the enhancer of *GATA2* more activated upon translocation and would that lead to stronger transcription factor binding? Or does the interaction of the *GATA2* enhancer with the *EVI1* promoter lead to hyper-activation of the enhancer and sequentially *EVI1* expression enhancement? Another question to be addressed is whether there is a role for the breakpoint itself in the hyperactivity of the enhancer (next paragraph)? An alternative explanation for the increased activity could be that within the *GATA2* locus suppressor elements lay which are absent in the *EVI1* locus. Using the distinct models that we have now generated, including an inv(3) EVI1-eGFP/GATA2-mCherry reporter, we may be able to study the mechanisms of *EVI1* and *GATA2* transcription and uncover the role of the enhancer in normal and abnormal gene regulation. How MYB particularly activates *EVI1* and not *GATA2* using the same enhancer may be further studied using these models.

In the case of a t(3;8), we do not observe a transformation of a typical enhancer into a

super-enhancer as we observed for the *GATA2* enhancer in AML with inv(3)/t(3;3). In t(3;8) AML the translocated enhancer was already a strong super-enhancer originally belonging to the *MYC* gene as shown in Chapter 5. In fact, in many other 3q26-rearranged AMLs we found that *EVI1* hijacked already existing super-enhancers like in the case of a t(3;8) (Chapter 2) [4]. To investigate if the translocated *MYC* super-enhancer behaves like the *GATA2* enhancer it would be interesting to examine the role of MYB on *EVI1* expression in t(3;8) AML. MYB ChIP-seq experiments should be performed on primary AML patient cells harbouring a t(3;8) for which we may be able to distinguish between the translocated and non-translocated allele using a specific SNP. Motif analysis revealed a MYB binding motif in the enhancer module of the *MYC*-SE that was shown to be essential for *EVI1* activation (unpublished observation). By means of ChIP-seq experiments we obtained evidence that MYB indeed binds at the *MYC* super-enhancer in our cell line models (data not shown). Preliminary data showed that targeting this MYB motif affected *EVI1* expression in the *EVI1*-eGFP t(3;8) model generated in Chapter 5. It is possible that MYB binds both the translocated as the non-translocated allele in the case of a t(3;8), since the *MYC* super-enhancer was already a super-enhancer before being translocated. Both ways, MYB could be essential for *EVI1* expression in t(3;8) AML and possibly also in other 3q26 rearranged AMLs. Whether MYB specifically drives *EVI1* expression and not *MYC* via this super-enhancer binding site will be further investigated.

Retro-elements: activating capacity or a genomic vulnerability?

Aberrant transcription of *EVI1* is often driven by chromosomal aberrations in AML patients. Our group has exactly identified many chromosomal breakpoints and the question arose whether these patient-specific breakpoints add to the deregulation of *EVI1*. As presented in Chapter 4, RNA read-trough, a high enrichment of retro-elements and CpG demethylation of those rearranged elements were particularly observed near inv(3)/t(3;3) breakpoints. As RNA read-trough at enhancers (referred to as eRNA) has been reported to have a functional role in transcription [6], the observation of transcription of the translocated enhancer and the breakpoint region in these inv(3)/t(3;3) AMLs may indicate a role for the breakpoint sequences in *EVI1* transcriptional enhancement. The suggestion that ancient retro-elements might play a role in *EVI1* gene activation, was also supported by the fact that *EVI1* was first identified as an ectropic viral insertion site in mouse myeloid leukemias, in which the gene is activated by inserted retroviral LTR sequences [7]. In fact, patients with X-linked chronic granulomatous disease that underwent gene therapy developed AML that was caused by *EVI1* overexpression as the result of retroviral insertion as well [8]. Together this made us wonder whether the translocated and de-methylated retro-elements could be of importance for *EVI1* expression in 3q26-rearranged AMLs. However, with the experiments described in Chapter 4, no evidence was obtained for an additional regulatory function or activating signal for *EVI1* transcription for the nucleotide sequences at these breakpoint locations.

Since the involvement of retro-elements has been described in the formation of complex chromosomal aberrations through retro-transposition across different cancer entities [9], it can well be that retro-elements form a genomic vulnerability rather than being involved in transcriptional enhancement itself. The location where breakpoints occur can be explained by the possibility that these chromosomal locations are more vulnerable for DNA damage. This DNA damage can be caused by the high enrichment of retro-elements, or simply due to the open and consequently less protected nature of the chromatin at these locations [10]. Alternatively, if activation of *EVI1* is only efficient when an already active enhancer is translocated to its vicinity, chromosomal breakpoints are only found at loci where active enhancers reside. Although the outcome of the experiments in Chapter 4 do not point to a role of the breakpoints in *EVI1* activation, it would still be interesting to find out whether the demethylation that we found of the retro-elements has any role in the opening of the chromatin at the *GATA2* super enhancer that was translocated to *EVI1*.

Enhancer-promoter interaction: the role of transcription factor binding?

In Chapter 5 we discussed the structural organisation of super-enhancers. In this Chapter we distinguished hierarchical from non-hierarchical organized super-enhancers [11]. Within hierarchical super-enhancers, one enhancer module (hub-enhancer) has more chromosomal interactions than the other modules within the same super-enhancer. This hub-enhancer is also distinctly associated with CTCF and cohesin binding [11]. In Chapter 5 we demonstrate that the translocated *MYC* super-enhancer is hierarchically organised and its hub-enhancer is the only module essential for *EVI1* transcription in t(3;8) AML. This module can be distinguished from the other modules by the binding of key hematopoietic transcription factors, known as the hematopoietic heptad factors [12] (Chapter 2, 3 and 5). The deletion of this module not only caused the loss of *EVI1* transcription, but also disrupted the interaction of the *EVI1* promoter with the *MYC* super-enhancer. We hypothesize that transcription factor binding at the enhancer (or possibly at the promoter as well) can add to enhancer-promoter looping and/or stability. To test this, transcription factor DNA-binding motifs should be mutated within this hub-enhancer. ChIP-seq experiments should be done to reveal loss of binding of the corresponding transcription factors to this locus, followed by 4C-seq experiments to examine enhancer-promoter interactions. Another approach to study enhancer-promoter interaction, organisation and dynamics which is currently under development is nanoFISH. This technology uses super-resolution microscopy to study molecular interactions with a 5kb resolution (e.g. enhancer-promoter interaction) in single cells. This has as a major advantage that we can study promoter to enhancer interaction in primary leukemias at the single cell level (collaboration with the group of Prof. Dr. Hartmann – University of Munich). This hopefully will allow us to validate our findings in primary leukemia cells with 3q26-rearrangements of which only limited material is available.

CTCF dependent enhancer docking: the key mechanism of enhancer hijacking?

In Chapter 5 we show that a CTCF binding site 2.6 kb upstream op the transcriptional start site (TSS) of *EVI1* facilitates enhancer hijacking in t(3;8) AML. In all 3q26-rearranged AMLs this upstream CTCF binding site was preserved with respect to *EVI1*. Interestingly, a CTCF binding site upstream of the *MYC* TSS has been reported to function as a docking site for enhancers driving *MYC* expression [13]. Our findings suggest for a highly similar mechanism of transcriptional activation of *EVI1* in all 3q26-rearranged AML. We demonstrated that targeting CTCF binding at this site in our t(3;8) model, caused loss of *EVI1* transcription and abolishes enhancer-promoter interaction. We hypothesize that this CTCF enhancer-docking site is not limited to t(3;8) AMLs. To test this, we targeted the same CTCF binding site in our EVI1-eGFP inv(3) model system. Experiments are still on going, but the targeting CTCF binding at this site did affect EVI1 levels (data not shown). These preliminary results make us believe that also in inv(3)/t(3;3) AML, this CTCF enhancer-docking site facilitates the *GATA2* enhancer hijacking. CTCF ChIP-seq and 4C-seq experiments to address this question are in preparation. One may wonder if this enhancer-docking site is only essential for the enhancer-to-promoter looping, or whether it adds to the hyper-activation of the enhancer and consequently enhanced *EVI1* expression in 3q26-rearranged leukemias. It could be that such a CTCF/cohesin mediated enhancer-promoter interaction is very stable; possibly attracting a high density of transcriptional (co-) regulators or activators leading to decreased fluctuation of the promoter activity [14].

Chromosomal breakpoints in the 3q26 locus described in this thesis were often found relatively close to each other, i.e. near *EVI1* and always with the CTCF bound enhancer-docking site preserved with respect to the gene. It has been reported that translocation hotspots strongly associate with TSS in the genome [15, 16] and that breakpoints more frequently occur in regions enriched for genes with CTCF/cohesin binding sites nearby [10]. These observations, together with our data argue for an 'active' role in the hijacking of enhancers or even the formation of translocations for these type of CTCF binding sites found upstream of TSS of genes. With the models that we have generated and the fast evolving genome editing tools, it should be possible to answer these questions experimentally in the future.

Therapeutic targets

Leukemias with 3q26-rearrangements depend on EVI1: interfering with *EVI1* expression causes growth inhibition, differentiation and ultimately death of leukemic cells [1, 17]. As discussed in this thesis, both characteristic and mechanistic similarities between the distinct enhancer-driven *EVI1*+ leukemia's are found, suggesting that a therapy for one subtype may be effective for all these AMLs. The *EVI1-eGFP* t(3;8) model (Chapter 5) but also *EVI1-eGFP* inv(3) (Chapter 3) model can be valuable tools for compound screens to identify inhibitors of *EVI1* transcription that could constitute a promising treatment for these refractory leukemias. Since

3q26-rearranged leukemia's described in this thesis are driven by strong enhancers, marked by H3K27ac, P300 and transcription factor binding, a rational for the use of compounds like P300/HAT inhibitors is in place. Promising preliminary results have been obtained with our t(3;8) model, using a potent P300/HAT inhibitor that interferes with *EVI1* expression (e.g. A-485 [18]). Since these types of inhibitors affect P300 function by blocking H3K18 and H3K27acetylation, the effect will not be limited to *EVI1* expression. Intentions are to find a concentration window or combination of compounds to affect *EVI1* oncogenic expression and ultimately push cells towards terminal differentiation and at the same time limit the effects on healthy cells. Furthermore, a compound screen (>6000 compounds) is being set up using our *EVI1-GFP* models. The sensitivity of those systems will allow us to identify new but possibly also already available drugs able to interfere with *EVI1* expression.

The t(3;8) model is a useful first line tool to study direct effects on *EVI1* expression, since this cell line does not depend on *EVI1* expression for survival (Chapter 5). This makes follow-up experiments, like ChIP-seq and 4C-seq, to analyse direct mechanistic effects of genomic targeting or compounds possible. Later on in the process of characterising the biological effect and efficiency of compounds, for example cell differentiation or viability, *EVI1* dependent model systems like the inv(3) model (Chapter 3) should be used. If promising drugs will be found, their efficacy can be tested in primary AMLs with 3q26-rearrangements *in vitro* and if possible ultimately in xeno-transplant models *in vivo*.

REFERENCES

1. Groschel, S., et al., *A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia.* Cell, 2014. **157**(2): p. 369-381.

2. Yamazaki, H., et al., *A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression.* Cancer Cell, 2014. **25**(4): p. 415-27.

3. Katayama, S., et al., *GATA2 haploinsufficiency accelerates EVI1-driven leukemogenesis.* Blood, 2017. **130**(7): p. 908-919.

4. Ottema, S., et al., *Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia.* Blood, 2020. **136**(2): p. 224-234.

5. Mulet-Lazaro, R., *Allele-specific expression of GATA2 due to epigenetic dysregulation in CEBPA double mutant AML.* Manuscript submitted for publication, 2020.

6. Lam, M.T., et al., *Enhancer RNAs and regulated transcriptional programs.* Trends Biochem Sci, 2014. **39**(4): p. 170-82.

7. Mucenski, M.L., et al., *Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors.* Molecular and cellular biology, 1988. **8**(1): p. 301-308.

8. Ott, M.G., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.* Nature Medicine, 2006. **12**(4): p. 401-409.

9. Rodriguez-Martin, B., et al., *Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition.* Nature Genetics, 2020. **52**(3): p. 306-319.

10. Lin, C.-Y., et al., *Translocation Breakpoints Preferentially Occur in Euchromatin and Acrocentric Chromosomes.* Cancers, 2018. **10**(1): p. 13.

11. Huang, J., et al., *Dissecting super-enhancer hierarchy based on chromatin interactions.* Nat Commun, 2018. **9**(1): p. 943.

12. Beck, D., et al., *Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes.* Blood, 2013. **122**(14): p. e12-22.

13. Schuijers, J., et al., *Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism.* Cell Rep, 2018. **23**(2): p. 349-360.

14. Ren, G., et al., *CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression.* Molecular Cell, 2017. **67**(6): p. 1049-1058.e6.

15. Chiarle, R., et al., *Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells.* Cell, 2011. **147**(1): p. 107-19.

16. Klein, I.A., et al., *Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes.* Cell, 2011. **147**(1): p. 95-106.

17. Smeenk, L., *An essential role for MYB in driving oncogenic EVI1 expression in enhancer-rearranged leukemias.* manuscript submitted for publication, 2020.

18. Lasko, L.M., et al., *Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours.* Nature, 2017. **550**(7674): p. 128-132.

A

ENGLISH SUMMARY

The **first chapter** is an introduction to this thesis. Haematopoiesis, leukemic transformation and the role of the oncogene *EVI1* in both processes are introduced in this section, as well as 3D chromatin interactions, gene regulation, the role of transcription factors and several experimental methods to study these processes. With the studies presented in the chapters that follow we aimed to gain knowledge about the mechanism of oncogenic *EVI1* activation in AML with inv(3)/t(3;3) or other 3q26-rearrangements with the goal to interfere with *EVI1* expression and ultimately treat these leukemic patients.

In the **second chapter** a study is reported that involved a unique cohort of atypical 3q26/*MECOM* rearranged leukemia. We demonstrated high levels of the *EVI1* oncogene, but no or very low levels of *MDS1-EVI1* similar to what typically has been observed in *EVI1*+ inv(3)/t(3;3) leukemia. Experimental approaches as FISH, 3q-captured DNA-sequencing and SNP-array have been used to identify rearrangements as well as copy number changes in the *MECOM* locus and map unique breakpoints in AML patient samples. We demonstrate that translocated loci frequently involve strong regulatory elements, such as super-enhancers. Our data suggest that these enhancers originally belong to genes active in myeloid development. In addition, allele specific *GATA2* expression was observed in over half of the leukaemia cases in this study. The molecular defects identified in this group of AMLs recapitulate the main mechanism of leukemic transformation in inv(3)/t(3;3) AML and we conclude that atypical 3q26/*MECOM* and inv(3)/t(3;3) should be categorized as a single entity of 3q26-rearranged AML.

In the **third chapter,** the *GATA2* enhancer driven *EVI1* oncogene expression in inv(3)/t(3;3) AML has been studied using an CRISPR-Cas9 based enhancer scan. This method allowed us to identify essential regions within the translocated enhancer that regulates *EVI1* expression. A single 1kb enhancer module, present in a 18kb minimal translocated super-enhancer, was found to be indispensable for *EVI1* transcription. In this locus, a highly conserved transcription factor binding motifs were identified to be among the top scoring sites in this screen. In particular a MYB binding motif and consequently the protein MYB was identified as a unique player in this process, by being preferably bound to the rearranged *GATA2* enhancer driving *EVI1* transcription, but not affecting *GATA2* expression. It was shown that interfering with the MYB binding motif at this locus, or interfering with MYB activity using a specific MYB/p300 interfering peptide, led to *EVI1* deregulation, differentiation of the cells and ultimately cell death. The findings in this chapter suggest the possibility of selectively interfering with oncogenic *EVI1* expression by targeting MYB functioning in inv(3)/t(3;3) and possibly other 3q26-rearranged leukemia, leaving normal hematopoietic stem and progenitor cells, in which GATA2 is a crucial hematopoietic regulator, unaffected.

The **fourth chapter** discusses the possible involvement of retro-elements in the activation of *EVI1* in inv(3)/t(3;3) AML, by a functional genomics analysis of an AML patient cohort and cell lines. To dissect the putative role of retro-elements in *EVI1* transcriptional activation, CRISPR-Cas9 technology based *in vitro* studies were performed using a human cell line model where an *eGFP* reporter was cloned 3' of *EVI1.* Though retro-elements were found highly enriched at inv(3)/t(3;3) breakpoint sites, no added function in transcriptional activation of *EVI1* could be shown for these ancient elements. We conclude in this chapter that retro-elements represent a genomic vulnerability, but do not provide any additional regulatory signal to activate *EVI1*.

The studies in the **fifth chapter** demonstrate that the hyper-activation of *EVI1* in t(3;8) AML is caused by a distant *MYC* super-enhancer. These findings have been generated by the use of 3q-captured DNA-sequencing, ChIP-sequencing and chromatin conformation captures on primary patient samples. The generation of a patient-based t(3;8)(q26;q24) model, using CRISPR-Cas9 technology, provided a unique working tool to study enhancer organisation, enhancer-promoter interaction and *EVI1* transcriptional regulation. A single enhancer module within the *MYC* super-enhancer was identified to be particularly essential in driving *EVI1* expression and interestingly also seemed to be involved in enhancer-promoter looping. This enhancer module was found to recruit a key set of transcription factors that have been described as a heptad of early hematopoietic regulators, such as GATA2, RUNX1 and ERG. To our knowledge, these factors have never been shown to be involved in chromatin interactions. After applying 3q-captured DNA-sequencing on other 3q26-rearranged AML samples, similar transcription factor-recruiting enhancer hubs were shown to be translocated to the *EVI1* locus. Besides a role for transcription factors in chromatin interactions, CTCF binding site within the *MYC* SE and upstream of the *EVI1* promoter were found to be involved in enhancer organisation and promoter interaction. Targeting the CTCF binding site upstream of the *EVI1* promoter strongly affected enhancer interaction and *EVI1* transcription. This CTCF binding site was preserved in the all 3q26-rearranged leukemias addressed in this chapter, including inv(3)/t(3;3) AMLs. Our findings point to a highly similar mechanism of *EVI1* transcriptional activation in all *EVI1*+ 3q26-rearranged AMLs. We hypothesise that in each of these AMLs, the CTCF binding site found near the *EVI1* promoter is essential for interaction with a hijacked enhancer active in early hematopoiesis to ultimately drive *EVI1* expression. We suggest that our model can be used as a first line tool for compound screens to interfere with enhancer driven oncogenic transcription, but can also be used to study super-enhancer organisation and enhancer-promoter interaction in general.

Finally, in **chapter 6** the content of this thesis is summarized, the findings are discussed and put in perspective**.** Furthermore, propositions for the usage of our model systems are done and suggestions are made for follow-up experiments.

NEDERLANDSE SAMENVATTING

Het **eerste hoofdstuk** is een inleiding op de studies in dit proefschrift. Hematopoëse, de transformatie tot leukemie en de rol van het oncogen *EVI1* in deze beide processen worden in dit deel geïntroduceerd. 3D-chromatine-interacties, genregulatie, de rol van transcriptiefactoren en verschillende experimentele methoden om deze processen te bestuderen worden ook in dit hoofdstuk uitgelegd. Mede dankzij de studies beschreven in dit proefschrift begrijpen wij het mechanisme van oncogene *EVI1* expressie beter. Het uiteindelijk doel van dit onderzoek is om in de toekomst te kunnen ingrijpen op de expressie van *EVI1* en daarmee specifiek deze leukemiepatiënten beter te kunnen behandelen.

Het **tweede hoofdstuk** beschrijft een studie waarin een uniek cohort van verschillende atypische 3q26/*MECOM* herschikte leukemieën wordt onderzocht. We tonen hoge expressie levels van *EVI1* aan, in deze atypische 3q26/*MECOM* AML patiënten, maar geen of zeer lage *MDS1-EVI1* expressie. Deze observatie is vergelijkbaar met wat wordt waargenomen in *EVI1*+ inv(3)/t(3;3) leukemieën. Experimentele benaderingen zoals FISH, 3q-cature DNA-sequencing en SNP-array zijn gebruikt om afwijkingen en het verlies of winst van DNA in de 3q26/*MECOM*-locus te identificeren. Verder werden unieke breekpunten en translocatie partner-loci bij deze patiënten in kaart gebracht. We hebben aangetoond dat deze loci vaak sterke regulerende elementen bevatten, die kunnen worden gekarakteriseerd als super-enhancers. Onze gegevens suggereren dat deze enhancers origineel toebehoren aan genen die actief zijn in vroege hematopoëtische voorloper cellen. Bovendien werd allel-specifieke *GATA2* expressie waargenomen in meer dan de helft van de gevallen van leukemie in deze studie. We concluderen dat in deze groep van AML het belangrijkste leukemische mechanisme van ontsporing is de aberrante expressie van *EVI1* door een 3q26 re-locatie. We concludeerden dat zowel atypische 3q26/*MECOM* en inv(3)/t(3;3) AML patienten gecategoriseerd kunnen worden als een specifieke leukemie met een *EVI1-* rearrangering.

In het **derde hoofdstuk** is de *GATA2*-enhancer, die de oncogene expressie van *EVI1* in inv(3)/t(3;3) AML aanstuurt, ontleed met behulp van een op CRISPR-Cas9 gebaseerd enhancer screen. Met deze methode konden we essentiële motieven binnen de getransloceerde superenhancer identificeren die belangrijk zijn voor de expressie van *EVI1*. In het bijzonder werd een MYB-binding motief en daarmee het eiwit MYB eiwit geïdentificeerd als een unieke factor in dit proces. MYB bleek bij voorkeur gebonden te zijn aan de *GATA2*-enhancer op het herschikte allel, waardoor de *EVI1* expressie werd verhoogd. *GATA2* expressie zelf bleek niet te worden beïnvloed door MYB. Mutaties van het MYB-bindingsmotief op deze locus, of het verstoren van de MYB-functie met behulp van een peptide, leidde tot *EVI1* deregulatie, verlies van proliferatie, verhoogde myeloide differentiatie en uiteindelijk tot celdood. De gegevens in dit hoofdstuk suggereren dat het mogelijk is om selectief te interfereren met

oncogene *EVI1* expressie door de binding en de functie van het eiwit MYB te beïnvloeden. Dit zou een manier kunnen zijn om inv(3)/t(3;3) en mogelijk andere 3q26-herschikte leukemie te behandelen zonder dat normale hematopoëtische stam- en voorlopercellen, waar *GATA2* expressie van belang is, geraakt worden.

Het **vierde hoofdstuk** bespreekt de mogelijke betrokkenheid van retro-elementen bij de activering van *EVI1* in inv(3)/t(3;3) AML door middel van een functionele genomics analyse van een cohort van AML patiënten en cellijnen. Om de rol van deze retro-elementen in *EVI1* transcriptie activering te bestuderen, werden *in vitro* studies met behulp van CRISPR-Cas9 uitgevoerd op een menselijk cellijnmodel waarin een *eGFP*-reporter 3 'van *EVI1* werd ingebracht. Hoewel retro-elementen sterk verrijkt bleken te zijn op breekpunten gevonden in inv(3)/t(3;3) leukemie, lijken deze elementen geen meerwaarde te hebben bij de transcriptionele activatie van *EVI1*. We concludeerden in dit hoofdstuk dat er een verhoogde genomische kwetsbaarheid is op plekken waar we retro-elementen vinden, met als gevolg chromosomale afwijkingen zoals bij AML met 3q26 afwijkingen. Alhoewel het zo zou kunnen zijn dat deze retro-elementen betrokken zijn bij de formatie van nieuw gevormde TADs, hebben we geen bewijs gevonden dat ze een directe rol hebben bij transcriptionele regulatie.

Het **vijfde hoofdstuk** demonstreert de activering van *EVI1* door een ver gelegen *MYC* super-enhancer doormiddel van 3q-captured DNA-sequencing, ChIP-sequencing en chromatine conformation captures op primair patiënten materiaal. Het genereren van een op een patiënt gebaseerde t(3;8)(q26;q24) model met behulp van CRISPR-Cas9-technologie leverde een uniek model op. Met dit model kon de organisatie van enhancers, enhancer-promotor-interacties en *EVI1*-transcriptionele regulatie bestudeerd worden. Een enkele enhancer module binnen de *MYC* super-enhancer werd geïdentificeerd als bijzonder essentieel voor het aansturen van *EVI1*-expressie en leek ook betrokken te zijn bij enhancer-promoter interactie. Deze enhancer module bleek een belangrijke set transcriptiefactoren te binden, deze factoren zijn eerder beschreven als een zevental van vroege hematopoëtische regulatoren. Voor zover wij weten, is het nooit aangetoond dat deze factoren ook betrokken zijn bij chromatine-interacties. Ook in andere 3q26-herschikte leukemieën zijn met behulp van 3q-captured DNA-sequencing vergelijkbare transcriptiefactor bindende enhancer-hubs geïdentificeerd die verhuizen naar *EVI1*. Naast een rol van transcriptiefactoren in chromatine-interacties, bleken ook CTCF-bindingsplaatsen binnen de *MYC* SE en 5' van de *EVI1* promotor betrokken te zijn bij enhancer-organisatie en enhancer-promotor interactie. Het verstoren van de CTCF-bindingsplaats 5' van de *EVI1* promotor beïnvloedde de enhancer-interactie en *EVI1* transcriptie sterk. Deze CTCF bindingsplaats werd behouden in alle 3q26 herschikkingen die in dit hoofdstuk worden behandeld, inclusief die van inv(3)/t(3;3) leukemie. Onze bevindingen wijzen op een sterk vergelijkbaar mechanisme van *EVI1* transcriptionele

activering. Het lijkt er op dat alle *EVI1*+ 3q26-herschikte leukemieën deze promotor CTCF-bindingsplaats gebruiken om een enhancer te kapen die actief is in de vroege hematopoëse om zo *EVI1* expressie te verhogen en leukemogenese te induceren. We suggereren dat ons t(3;8) model kan worden gebruikt als een eerstelijns systeem om medicijnen te testen die interfereren met enhancer-gestuurde oncogene transcriptie. Verder zou het model gebruikt kunnen worden om super-enhancer organisatie en enhancer-promotor interactie in het algemeen beter te bestuderen.

Ten slotte worden er in **hoofdstuk 6** de bevindingen in dit proefschrift besproken en in perspectief geplaatst. Verder worden er suggesties gedaan voor het gebruik van de beschreven model systemen en worden vervolg experimenten voorgesteld.

LIST OF MOST IMPORTANT ABBREVIATIONS

4C-seq	Circularized Chromatin Conformation Capture by sequencing
ChIP-seq	Chromatin immune precipitation by sequencing
3q-seq	targeted sequencing of the long arm (q) of chromosome 3
NGS	Next generation sequencing
FACS	Fluorescent activated cell sorting
PCR	Polymerase chain reaction
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
HDR	homology-directed repair
AML	Acute Myeloid Leukemia
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem and progenitor cell
TAD	Topological associated domains
TSS	Transcriptional start site
TF	Transcription factor
SE	super-enhancer
RE	Retro-elements
SINE	short interspersed element
LINE	long interspersed element
LTR	long terminal repeat
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
MECOM	MDS1 and EVI1 complex locus (gene)
EVI1	Ecotropic viral integration site 1 (gene)
MDS1	Myelodysplasia syndrome 1
GFP	green fluorescent protein
eGFP	enhanced green fluorescent protein
MYC	MYC proto-oncogene
GATA2	GATA Binding Protein 2
SMC3	Structural maintenance of chromosomes protein 3
RAD21	RAD21 cohesin complex component
CTCF	CCCTC binding factor
H3K27ac	Histone 3 lysine 27 acetylation
H3K4me3	Histone 3 lysine 4 tri-methylation
H3K9ac	Histone 3 lysine 9 acetylation

CURRICULUM VITAE

Sophie Ottema was born in The Netherlands on the 22th of October 1989. After secondary school, she successfully followed a Bachelor in Life, Science & Technology, a joint study of the University of Leiden and the Technical University of Delft (2008-2012). During this time she did her internship in the department of Genetics at the Leiden University Medical Centre (LUMC) supervised by Maartje Huijbrechts, working on a project studying Myasthenia Gravis. At this time she was involved in the set-up of an *in vitro* assay to stimulate post-synaptic neuromuscular junctions. After completion she continued with her research studies by following the Masters in Molecular Medicine at the Erasmus University in Rotterdam. During these two years she interned at the department of Cell Biology in the lab of dr. Niels Galjart studying the role of CLIPs and CLASPs in microtubule dynamics (2012-2014). After graduation she received an EMBO short-term fellowship with which she continued her research in the lab of Helder Maiato at the IBMC institute – Porto, Portugal. This lab has expertise in chromatin instability and dynamics; here she studied the role of CLASP2 in cell division, using flow cytometry and immune fluorescent approaches specifically in mouse embryonic stem cells (2014). The year after she made her choice to precede a career in academia by accepting a function as a PhD candidate in the group of Prof. dr. Pieter Sonneveld at the Hematology department of the Erasmus University Medical Centre (Erasmus MC) in Rotterdam. The lab focuses their research on Multiple Myeloma and the project was on finding a marker to predict patient outcome using an experimental drug targeting the protein CRM1 (2015-2016). After about one year the project initiated in the Myeloma group felt too translational and it was decided in consultation with her promotor Prof. dr. Pieter Sonneveld it would be more fitting to change research to a more fundamental group and topic. From then onwards, Sophie started her new PhD program in the group of Prof. dr. Ruud Delwel, department of Hematology Erasmus MC were she studied chromosomal aberrations leading to transcriptional deregulation of the oncogene *EVI1* in acute myeloid leukemia (2016-2020, this thesis). Her career at the Erasmus MC will be continued at the department of Clinical Genetics, where she will be trained as laboratorial specialist in the Clinical Genetic diagnostics starting from February 2021.

LIST OF PUBLICATIONS

Yu N, Signorile L, Basu S, Ottema S, Lebbink JHG, Leslie K, Smal I, Dekkers D, Demmers J, Galjart N. Isolation of Functional Tubulin Dimers and of Tubulin-Associated Proteins from Mammalian Cells. Curr Biol. 2016 Jul 11;26(13):1728-1736. doi: 10.1016/j.cub.2016.04.069. Epub 2016 Jun 9. PMID: 27291054.

Ottema S, Mulet-Lazaro R, Beverloo HB, Erpelinck C, van Herk S, van der Helm R, Havermans M, Grob T, Valk PJM, Bindels E, Haferlach T, Haferlach C, Smeenk L, Delwel R. Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia. Blood. 2020 Jul 9;136(2):224-234. doi: 10.1182/blood.2019003701. PMID: 32219447.

Mika J, Ottema S, Kiehlmeier S, Kruse S, Smeenk L, Müller J, Schweiggert S, Herrmann C, Sanders M, Delwel R, Gröschel S. Derepression of retroelements in acute myeloid leukemia with 3q aberrations. Haematologica. 2021 Mar 25. doi: 10.3324/haematol.2020.277400. Epub ahead of print. PMID: 33764001.

Smeenk L, Ottema S, Mulet-Lazaro R, Ebert A, Havermans M, Arricibita Varea A, Fellner M, Pastoors D, van Herk S, Erpelinck-Verschueren C, Grob T, Hoogenboezem RM, Kavelaars FG, Matson DR, Bresnick EH, Bindels EM, Kentsis A, Zuber J, Delwel R. Selective requirement of MYB for oncogenic hyperactivation of a translocated enhancer in leukemia. Cancer Discov. 2021 May 12:candisc.1793.2020. doi: 10.1158/2159-8290.CD-20-1793. Epub ahead of print. PMID: 33980539.

Ottema S, Mulet-Lazaro R, Erpelinck-Verschueren C, van Herk S, Havermans M, Arricibita Varea A, Vermeulen M, Beverloo HB, Gröschel S, Haferlach T, Haferlach C, Valk PJM, Bindels E, Smeenk L, Delwel R. MYC super-enhancer hijacking by EVI1 in leukemia is CTCF dependent. Accepted for publication, July 2021 by Nature Communications.

PHD PORTOFOLIO

Name: Sophie Ottema
PhD period: may 2015- dec 2020
Erasmus MC department: Hematology

Courses and workshops

|	
Academic Integrity	2015
Basic course on R	2015
Photoshop and Illustrator (Basic)	2015
Molmed day	2017
Workshop 'Molecular aspects of Hematological malignancies'	2015-2019

Scientific meetings Department of Hematology

Workdiscussions (weekly)	2015-2020
Erasmus Hematology lectures (Monthly)	2015-2020
PhD lunch with invited speaker	2015-2019
Friday Floor meetings (weekly)	2015-2020
Journal club (bi-weekly)	2015-2020

National and international conferences

European Hematology Association, Madrid, ESP	2017
Cancer genomic meeting, Oncode, Amsterdam, NL	2017
American Society of Hematology, San Diego, US (poster)	2018
Epigenetics and epitranscriptomics, NKI, Amsterdam, NL	2019
European Hematology Association, Amsterdam, NL (oral)	2019

Supervising and organisation activities

Supervising Bachelor internship	
Esther Blommert, Nano Biology, 6 months	2017-2018
Teun Been, Laboratory techniques, 9 months	2018-2019
Organization and supervision PhD lunch with invited speakers	2017-2018
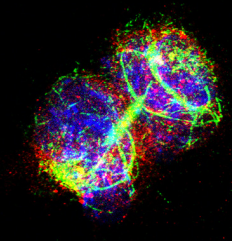
Two mouse embryonic stem cells wrapping up a successful cell division
- quite some organisation in a picture that looks rather chaotic