Altered enhancer-promoter interaction leads to *MNX1* expression in pediatric acute myeloid leukemia with t(7;12)(q36;p13)

Dieter Weichenhan[1,*], Anna Riedel[1,*], Etienne Sollier[1,*], Umut H. Toprak[2,*], Joschka Hey[1], Kersten Breuer[1], Justyna A. Wierzbinska[1], Aurore Touzart[1], Pavlo Lutsik[1], Marion Bähr[1], Anders Östlund[3], Tina Nilsson[4], Susanna Jacobsson[4], Marcel Edler[4], [4] Ahmed Waraky[3,4], Yvonne Lisa Behrens[5], Gudrun Göhring[5], Brigitte Schlegelberger[5], Clemens Steinek[6], Hartmann Harz[6], Heinrich Leonhardt[6], Anna Dolnik[7], Dirk Reinhard[8], Lars Bullinger[7], Lars Palmqvist[3,4], Daniel B. Lipka[9,10, #], Christoph Plass[1,10 #]

[1]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[3]Department of Laboratory Medicine, University of Gothenburg, Sweden
[4]Department of Clinical Chemistry, Sahlgrenska University Hospital, Gothenburg, Sweden.
[5]Department of Human Genetics, Hannover Medical School, Hannover, Germany
[6]Faculty of Biology, Ludwig-Maximilians-Universität in Munich, 82152 Planegg-Martinsried, Germany.
[7]Department of Hematology, Oncology, and Tumor Immunology, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-University Berlin, Berlin, Germany
[8]Department of Pediatric Oncology, University of Duisburg-Essen, Essen, Germany.
[9]Section of Translational Cancer Epigenomics, Division of Translational Medical Oncology, German Cancer Research Center (DKFZ), National Center for Tumor Diseases (NCT) Heidelberg, Heidelberg, Germany
[10]German Cancer Research Consortium (DKTK), 69120 Heidelberg, Germany

*Co-first author, [#]Co-senior author

Corresponding author:

Christoph Plass
Division of Cancer Epigenomics
German Cancer Research Center (DKFZ)
INF 280; 69120 Heidelberg; Germany
Phone: +49-6221-42-3301,
Fax: +49-6221-42-3359
e-mail: c.plass@dkfz.de

Abstract:

Acute myeloid leukemia (AML) with the t(7;12)(q36;p13) translocation occurs only in very young children and has a poor outcome. The expected oncofusion between breakpoint partners (*MNX1* and *ETV6*) has only been reported in a subset of cases. However, a universal feature is the strong transcript and protein expression of *MNX1*, a homeobox transcription factor that is normally not expressed in hematopoietic cells. Here, we map the translocation breakpoints on chromosomes 7 and 12 in affected patients to a region downstream of *MNX1* and either introns 1 or 2 of *ETV6*. The frequency of *MNX1* overexpression in pediatric AML (n=1556, own and published data) is 2.4% and occurs predominantly in t(7;12)(q36;p13) AML. Chromatin interaction assays in a t(7;12)(q36;p13) iPSC cell line model unravel an enhancer-hijacking event that explains *MNX1* overexpression in hematopoietic cells. Our data suggest that enhancer-hijacking is a more common and overlooked mechanism for structural rearrangement-mediated gene activation in AML.

Introduction:

Acute myeloid leukemia (AML) has been successfully investigated in the past using cytogenetic analysis. This led to the discovery of numerous recurrent chromosomal translocations [e.g. t(8;21)(q22; q22) or t(15;17)(q22;q12)] generating oncofusion proteins (e.g. RUNX1::RUNX1T1 or PML::RARA, respectively) that drive leukemogenesis. For many years, these translocations have served as diagnostic and prognostic markers, and affected patients can now be treated with specific targeted therapies (e.g. retinoic acid and arsenic trioxide in t(15;17) cases [1]. In 1998, two publications reported the translocation t(7;12)(q36;p13) in AML of infants [2,3]. AML with t(7;12)(q36;p13) occurs in children <18 months and is not seen in adult AML. A recent meta-analysis of the Nordic Society for Pediatric Hematology and Oncology (NOPHO-AML) determined that t(7;12)(q36;p13) AML constituted 4.3% of all children with AML under the age of 2 and found a 3-year event-free survival of 24% (literature-based data) and 43% (NOPHO-AML data) [4]. Cytogenetically, t(7;12)(q36;p13) AML is often associated with the occurrence of trisomy 19 [4,5], but no other recurrent aberrations have been described.

Reported breakpoints in t(7;12)(q36;p13) AML have mainly been evaluated by fluorescent in situ hybridization (FISH) analysis. The breakpoints on chromosome 12 (chr12) are located within intron 1 or 2 of *ETS variant transcription factor 6* or *ETV6*. On chr7 a heterogeneous breakpoint cluster region downstream to M*otor neuron and pancreas homeobox 1* (*MNX1*) and within the 3' end of *Nucleolar protein with MIF4G domain 1* (*NOM1*) has been described [6]. Beverloo et al. reported a fusion transcript of *MNX1* exon 1 to *ETV6* exon 2 or exon 3 [7], a result confirmed by several subsequent studies for some, but not all, t(7;12)(q36;p13) AML cases [4–6]. These *MNX1*::*ETV6* fusion transcripts could only be detected in about 50% of the cases [8], and an oncofusion protein has never been detected in t(7;12)(q36;p13) AML. This observation challenges the prevailing dogma for AML as well as other hematopoietic neoplasms that chromosomal translocations lead to the creation of oncofusion proteins driving leukemogenesis.

All AML cases with t(7;12)(q36;p13) have high expression of *MNX1* in common [5], suggesting a, yet unknown, mechanism of *MNX1* activation. Consistent with the activation of a silenced gene locus, a translocation of the *MNX1* locus from the nuclear periphery to the internal nucleus was seen, an observation that is in line with the idea that condensed and silent chromatin is located in the nuclear periphery [6]. Also, the interactions of *ETV6* downstream elements with the *MNX1* locus have been postulated as possible mechanisms for *MNX1* activation [6,9]. The first clue for the existence of possible aberrant promoter-enhancer interactions leading to *MNX1* activation came from our investigations of the GDM-1 AML cell

line which harbors a t(6;7)(q23;q36) translocation. In GDM-1, the *MNX1* promoter interacts with an enhancer element from the *MYB* locus on chr6q23 [10].

Recently, Nilsson et al. reported the introduction of a translocation between chr7q36 and chr12p13 modeling the one found in t(7;12)(q36;p13) AML into the human induced pluripotent stem cell (iPSC) line, ChiPSC22^{WT} [9]. The derivative line, ChiPSC22$^{t(7;12)}$, can be differentiated into hematopoietic stem and progenitor cells (HSPC), and as such expresses *MNX1*, suggesting that hematopoietic enhancers play a role in *MNX1* activation. Enhancer-hijacking has initially been described as a mechanism for oncogene activation in AML with inv(3)/t(3;3)(q21q26) AML where activation of *EVI1*, an isoform encoded from the *MDS and EVI1 complex locus* (*MECOM*), results from the repositioning of a *GATA2* enhancer [11–13]. Enhancer-hijacking is also implicated in acute leukemia of ambiguous lineage in which translocated hematopoietic enhancers from different chromosomes are involved in activating *BCL11B* [14].

Here we provide a detailed description of the molecular alterations found in six t(7;12)(q36;p13) AML patients and dissect the molecular mechanism leading to *MNX1* activation through the use of CRISPR-engineered ChiPSC22$^{t(7;12)}$ iPSCs and HSPCs. We identify an enhancer-hijacking event activating the *MNX1* promoter via hematopoietic enhancers from the *ETV6* locus and validate this event in the iPSC/HSPC system. Our data suggest that enhancer-hijacking may be a more widespread, but so far largely unappreciated, mechanism for gene activation in AML with cytogenetic abnormalities.

Results

Whole genome sequencing of t(7;12)(q36;p13) AML

To precisely map structural rearrangements, somatic copy number alterations (SCNAs) and genetic mutations in t(7;12)(q36;p13) AML, we performed whole genome sequencing (WGS) of four t(7;12)(q36;p13) AML cases (subsequently named: T1, T2, T3 and T4) from Germany and Sweden (**Fig. 1a** and **Supplementary Table 1**). Matched normal germline tissue was not available for these patients. We complemented our data with published WGS data from two samples with t(7;12)(q36;p13) from the TARGET-AML cohort [15]. WGS verified the presence of t(7;12)(q36;p13) as a reciprocal balanced translocation in all samples (**Fig. 1a**). The breakpoint on chr12 located in five samples in intron 1 of *ETV6* and in one sample in *ETV6* intron 2. On chr7, all breakpoints located downstream of *MNX1*, in four cases either within *NOM1*, which is the gene next to *MNX1*, or in two cases between *MNX1* and *NOM1* (**Fig. 1b**). In none of these cases the creation of an oncofusion gene between *MNX1* and *ETV6* is supported by the observed translocation breakpoints leaving the main *MNX1* variant (RefSeq: NM_005515) unaffected by the genomic rearrangements (**Fig. 1c**). Accompanying cytogenetic data (**Supplementary Table 1**) revealed trisomy 19 in all cases, a result confirmed by SCNA analysis for all cases except for T1, for which SCNA analysis identified a trisomy 22 but no trisomy 19 (**Fig. 1a**). A genome-wide analysis for gene mutations was not possible due to the lack of germline material, however mutations for common leukemia genes were tested (see methods for a list of genes tested) and identified mutations in *NOTCH2* (p.A2319V), *NF1* (p.P2310fs) and *PTPN11* (p.D61Y and p.E69K) (**Fig. 1a**).

Comparison of single nucleotide polymorphisms (SNPs) in the coding regions of *ETV6* indicated that the translocation on chr12 in T1, T2 and T3 (for T4 no RNA-Seq data is available) leads to transcriptional inactivation of *ETV6* on one allele as evident in monoallelic expression of *ETV6* (data not shown). In line with this, a previously described alternative exon 1B [16], located upstream of exon 3 does not show significant expression (data not shown). Thus, the translocation not only leads to activation of *MNX1* but also to a reduced and monoallelic expression of the transcriptional repressor *ETV6*.

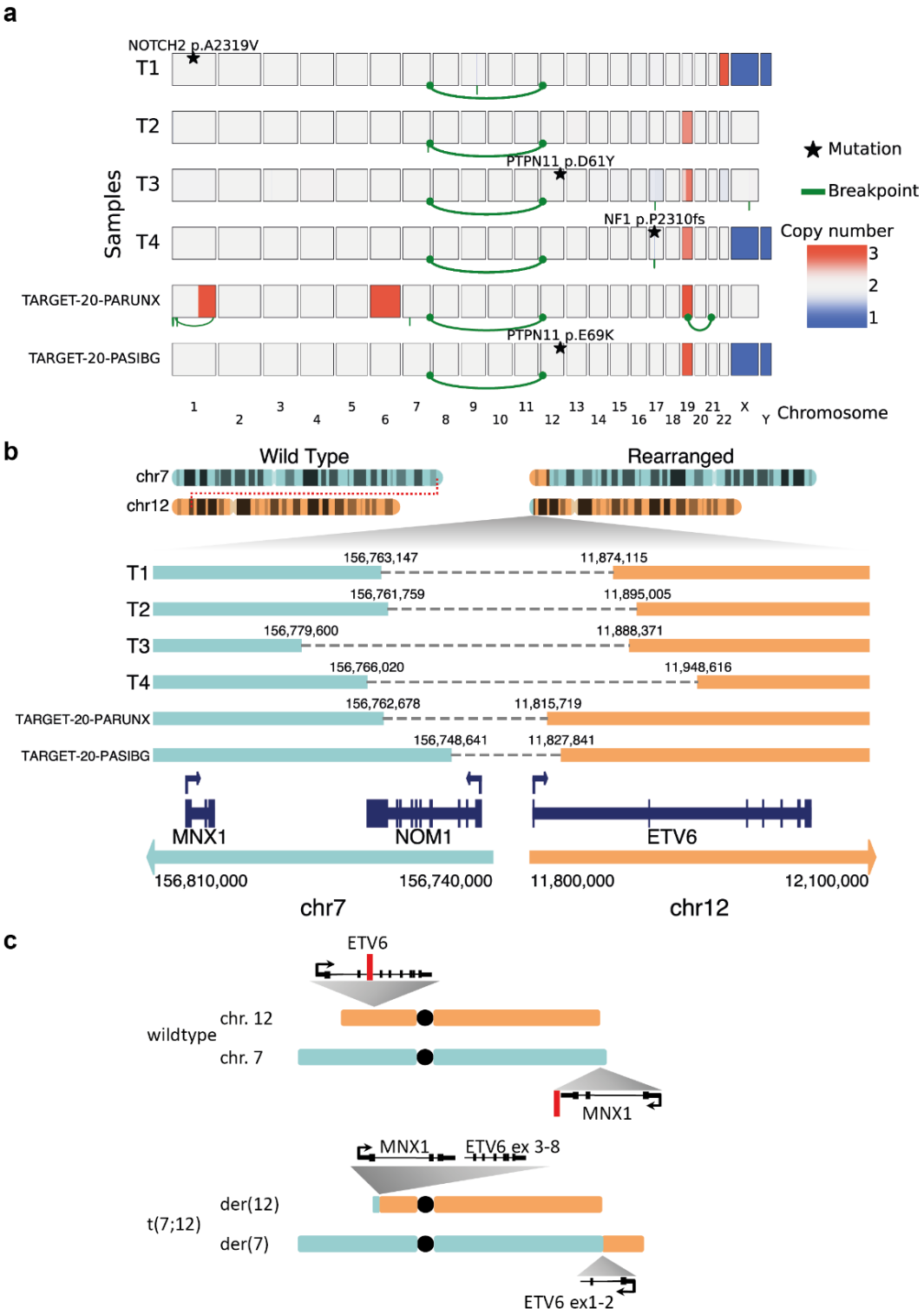

Fig. 1. Whole genome sequence analysis of t(7;12)(q36;p13) AML. a, Copy numbers (blue: loss; red: gain), structural rearrangements (green bow connecting 2 chromosomes) and mutations in known AML driver genes for six t(7;12)(q36;p13) AML samples based on whole genome sequencing. Samples T1, T2, T3 and T4 were profiled in this study, while TARGET-20-PARUNX and TARGET-20-PASIBG are from the TARGET-AML cohort [15]. **b,** Sketch of the rearranged chromosomes 7 and 12, and zoom-in on the region around the breakpoints. Breakpoints coordinates correspond to the GRCh37/hg19 reference genome. **c,** Schematic overview of chr7 (turquoise), chr12 (orange) and derivative chromosomes der(12) and der(7) resulting from the reciprocal t(7;12) translocation involving *MNX1* on chr7 and *ETV6* on chr12. Red lines indicate positions of break/fusion points.

MNX1 is highly expressed in all t(7;12)(q36;p13) AML and is associated with a characteristic gene expression signature

MNX1 is normally not expressed in the hematopoietic lineage, but previous reports indicated high *MNX*1 expression in all analyzed t(7;12)(q36;p13) AML [5,6,17]. Quantitative RT-PCR of t(7;12)(q36;p13) AML cases T1-T4 showed high *MNX1* expression as compared to most AML samples without t(7;12) (data not shown). To further validate this aberrant *MNX1* expression, we determined the frequency with which pediatric AML shows overexpression of *MNX1*. We used two pediatric AML cohorts with available gene expression data: Balgobind et al. 2011 (237 samples profiled with Affymetrix array) [18] and TARGET-AML (1319 samples profiled with RNA-Seq) [15]. For samples profiled with Affymetrix array, we used the 3-sigma rule to identify samples with *MNX1* overexpression, and for samples profiled with RNAseq, we considered an expression >0.5TPM to be overexpressed (**Fig. 2**). We detected *MNX1* overexpression in 7/237 (2.9%) samples in the Balgobind cohort, and in 31/1319 (2.3%; including resample for samples TARGET-20-PARUNX and TARGET-21-PASVJS) samples in the TARGET-AML cohort. All samples with a translocation t(7;12) had *MNX1* overexpression. Some samples had *MNX1* overexpression without genomic rearrangements close to *MNX1* reported in the karyotyping information, which indicates that there might be alternative mechanisms leading to *MNX1* activation. All t(7;12)(q36;p13) samples were diagnosed younger than 2 years old, however, most *MNX1*-overexpressing samples without t(7;12) were diagnosed at an older age (**Supplementary Table 1**).

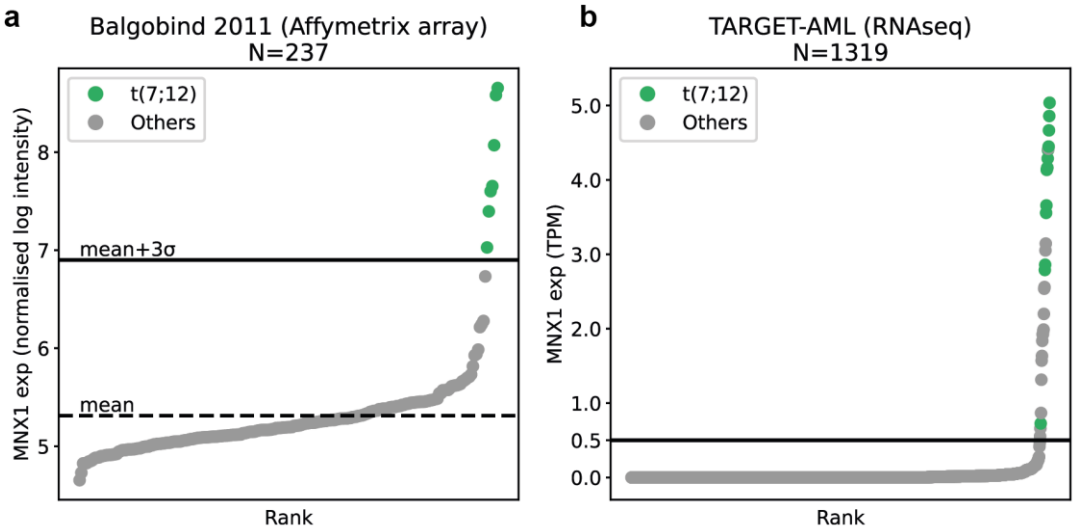

Fig. 2. *MNX1* expression in pediatric AML with t(7;12)(q36;p13) translocation. *MNX1* expression in two different pediatric AML cohorts: **a,** Balgobind et al. 2011 [18], 237 samples profiled with Affymetrix arrays. The mean expression level is shown with a dashed line, and the mean plus three standard deviations is shown with a horizontal line. **b,** TARGET-AML [15], 1319 samples profiled with RNAseq (transcripts per million, TPM). A cut-off of 0.5 TPM was used. Samples with cytogenetically detected t(7;12)(q36;p13) translocation are shown in green and other samples in grey.

Balgobind et al. described a characteristic gene signature for t(7;12)(q36;p13) AML consisting of genes with a most significantly discriminating expression as compared to other cytogenetic subgroups in pediatric AML [18]. The majority of these genes are either consistently downregulated in t(7;12)(q36;p13) AML (e.g. *TP53BP2*) or upregulated together with *MNX1* (*EDIL3*, *LIN28B*, *BAMBI*, *MAF*, *FAM171B*, *AGR2*, *CRISP3, KRT72* and *MMP9*). For each of the two cohorts (Balgobind 2011 and TARGET-AML), we performed a differential expression analysis between the t(7;12)(q36;p13) cases and the others (**Supplementary Table 2**). The lists of up- and downregulated genes that we identified include the genes identified by Balgobind et al. [18], which are indeed consistently deregulated in t(7;12)(q36;p13) across several cohorts, as well as other genes not reported before, and, therefore, provides a more complete picture of the deregulated gene expression in t(7;12)(q36;p13) AML. The samples with *MNX1* expression but without genomic rearrangement close to *MXN1* did not express this typical gene signature (**Extended Data Fig. 1**). The breakpoints on chr12 locate in *ETV6* and, hence, led to a corrupted *ETV6* allele. In line with this, we found in all cohorts a negative log-fold change for *ETV6* expression in the t(7;12)(q36;p13) samples compared to the others (data not shown).

A t(7;12)(q36;p13) cell line model exhibits MNX1 protein expression and chromatin interactions between the *MNX1* and *ETV6* regions

Previously, Nilsson et al. engineered an induced pluripotent stem cell (iPSC) line harboring a balanced translocation t(7;12)(q36;p13) (ChiPSC22$^{t(7;12)}$) with a breakpoint in *ETV6* intron 2 and a breakpoint 13 to 26 kb downstream of *MNX1* in the candidate common breakpoint region [8,9]. Upon differentiation of ChiPSC22$^{t(7;12)}$ cells to hematopoietic stem and progenitor cells (HSPC), *MNX1* became activated, while it remained silent in the differentiated parental line ChiPSC22^{WT}. We confirmed this result at the protein level by Western blot in three ChiPSC22$^{t(7;12)}$ sublines, 14D7, 23G8 and 24C7: the MNX1 protein was only expressed in the ChiPSC22$^{t(7;12)}$ HSPCs but neither in the ChiPSC22$^{t(7;12)}$ iPSCs nor in ChiPSC22^{WT} iPSCs or HSPCs (**Fig. 3a**). To examine, if hematopoietic-specific enhancers become relocated into the *MNX1* vicinity by the t(7;12)(q36;p13) translocation and may, thus, activate *MNX1* in ChiPSC22$^{t(7;12)}$ HSPCs, we applied circular chromosome conformation capture (4C) with iPSCs and HSPCs of ChiPSC22^{WT} and ChiPSC22$^{t(7;12)}$ using as viewpoint a region 3 kb upstream of *MNX1*. We found interaction between the *MNX1* and the *ETV6* regions, but only in HSPCs of ChiPSC22$^{t(7;12)}$, about 120-150 kb distant to the breakpoint in *ETV6*, but neither in undifferentiated iPSCs nor in ChiPSC22^{WT} (**Fig. 3b**). To substantiate the 4C data, we used a second *MNX1* viewpoint located in exon 1 of *MNX1*. Again, a strong interaction between *MNX1* and the vicinity of *ETV6*, about 400 kb distant to the *ETV6* breakpoint and located in

the *LRP6* gene, could only be observed in HSPCs of ChiPSC22$^{t(7;12)}$ but neither in iPSCs nor in ChiPSC22^{WT} (**Extended Data Fig. 2a**).

We further profiled the iPSCs and HSPCs of both ChiPSC22^{WT} and ChiPSC22$^{t(7;12)}$ lines with Hi-C. No interactions were seen between chr7 and chr12 in the ChiPSC22^{WT} (**Extended Data Fig. 2b**). However, we observed interactions between the *MNX1* promoter and sites vicinal to *ETV6* on chr12 in both iPSCs and HSPCs of ChiPSC22$^{t(7;12)}$ **(Fig. 3c)**. This observation could be explained by either the presence of two distinct enhancers or by recently reported states of poised enhancers, which enable gene activation with the support of additional stimuli upon differentiation [19–21].

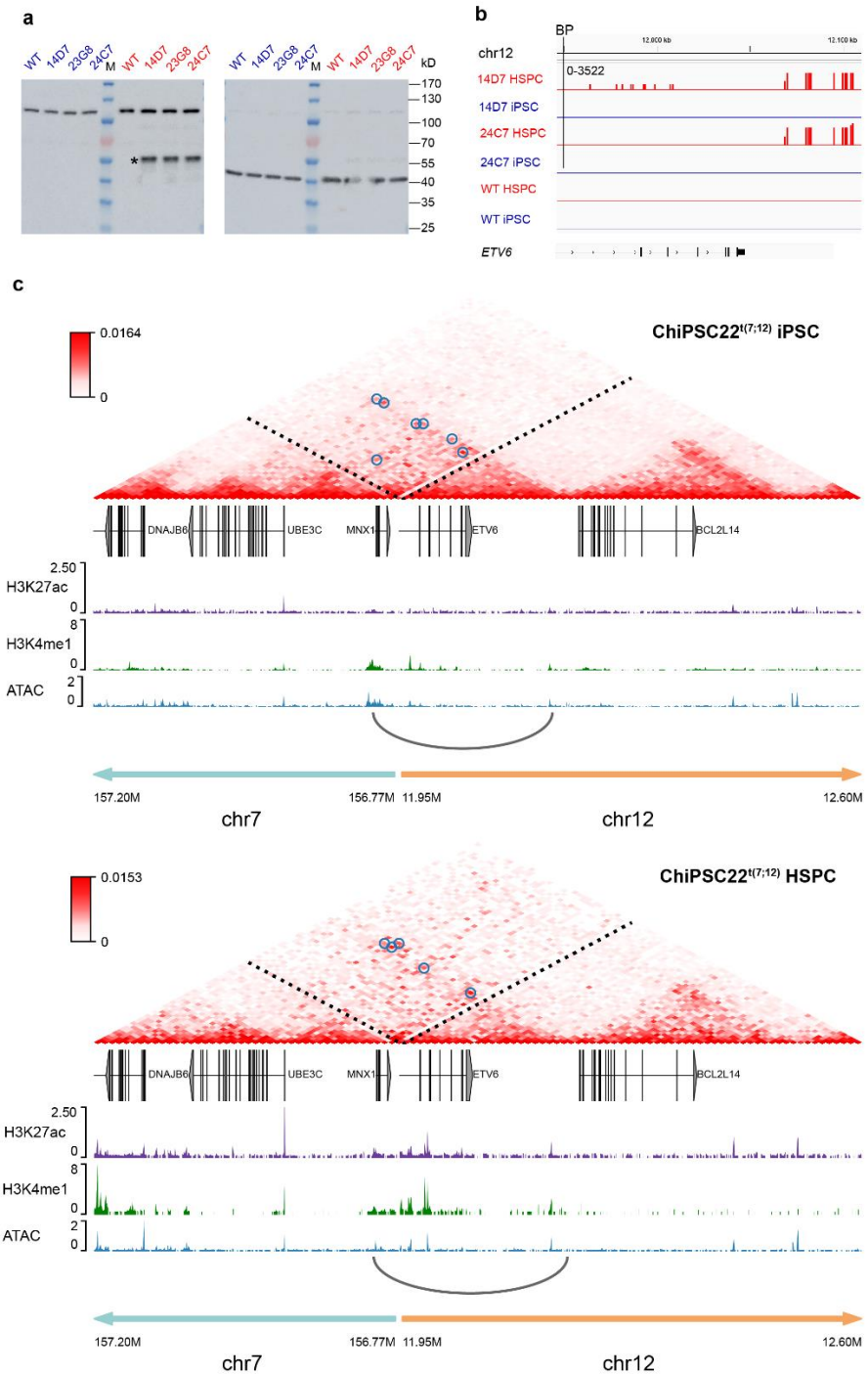

Fig. 3. MNX1 protein expression and chromatin interaction of the *MNX1* gene with the *ETV6* region in ChiPSC22$^{t(7;12)}$ HSPCs. a, Western blot with an MNX1 antibody (left) and iPSC (blue) and HSPC (red) protein extracts from ChiPSC22^{WT} and ChiPSC22$^{t(7;12)}$ sublines 14D7, 23G8 and 24C7. The MNX1 protein (asterisk) is only detected in HSPCs of ChiPSC22$^{t(7;12)}$ sublines 14D7, 23G8 and 24C7. The common band at about 120 kD results from an unknown protein cross-reacting with the MNX1 antibody. To demonstrate loading of equal protein amounts, the unstripped blot was re-incubated with a horseradish-peroxidase coupled antibody against β-actin (right). **b,** Circular chromosome conformation capture (4C) indicates specific interaction between the *MNX1* (viewpoint about 3 kb upstream of *MNX1*) and the *ETV6* region in HSPCs of ChiPSC22$^{t(7;12)}$ sublines 14D7 and 24C7. The range, 0-3522, of read numbers which are indicated by vertical bars was given by sample 24C7 HSPC and set to same for all samples. **c,** Chromatin interactions analyzed by Hi-C seq in the genomic region flanking the translocation breakpoint in the ChiPSC22$^{t(7;12)}$ subline 24C7, either as iPSCs (left) or HSPCs (right). Below, gene locations and peaks of H3K27ac, H3K4me1 and open chromatin are shown to indicate locations of putative enhancers.

Identification of hematopoietic enhancers in *ETV6* and its vicinity

To further support the enhancer-hijacking hypothesis, we searched for hematopoietic enhancers located up to 1 Mbp distant from the breakpoint in *ETV6*. Applying antibody-guided chromatin tagmentation (ACT-seq), we mapped peaks of enhancer marks, i.e. acetylated lysine 27 of histone 3 (H3K27ac) and mono-methylated lysine 4 of histone 3 (H3K4me1) in both iPSCs and HSPCs from ChiPSC22^{wt} and ChiPSC22$^{t(7;12)}$. To facilitate the identification of hematopoietic enhancers in ChiPSC22$^{t(7;12)}$ and ChiPSC22^{WT} HSPCs, we applied MACS2-peak calling using the corresponding iPSC data as internal reference. In addition, we used public ChIPseq-datasets for the same enhancer marks and the histone acetyltransferase P300 generated from CD34+ cells and the chronic myeloid leukemia (CML)-derived cell line MOLM-1. In the 1 Mbp region of interest, we identified eleven H3K27ac peaks which largely coincided with the P300 and H3K27ac peaks from MOLM-1 and CD34+ cells, respectively, as well as with the open chromatin peaks (**Fig. 4a**). We mapped four H3K4me1 peaks which were found as well in MOLM-1 and CD34+ cells and located at the same position as four of the H3K27ac peaks (**Fig. 4b**). In conclusion, we identified four strong enhancer candidates in *ETV6* and its close vicinity, which may drive *MNX1* activation in t(7;12)(q36;p13) AML. The enhancer located downstream of ETV6 was also identified in the Hi-C data as a strong interaction point with the *MNX1* promoter (**Fig. 3c**).

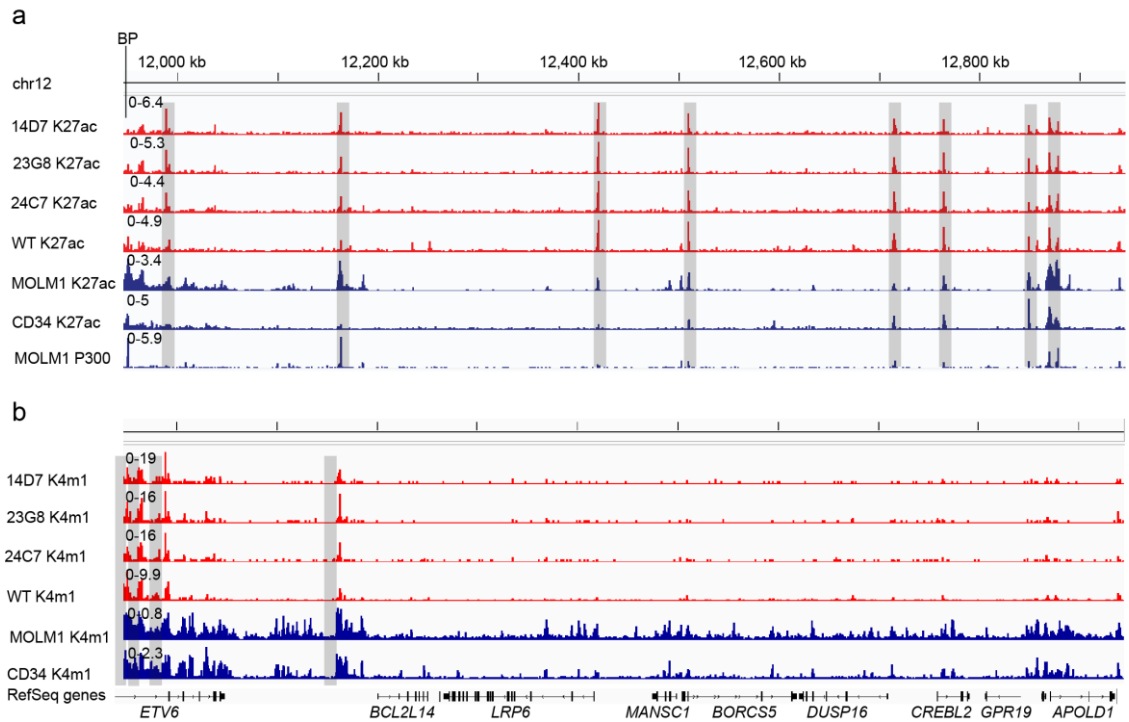

Fig. 4. Enhancer mark profiles 1 Mbp distal to the breakpoint (BP) position in *ETV6* of ChiPSC22$^{t(7;12)}$ sublines 14D7, 23G8 and 24C7 and of ChiPSC22^{WT}. a, HSPC-specific H3K27ac and **b,** HSPC-specific H3K4me1 profiles. Publicly available respective profiles of CD34+ and MOLM-1 (also for enhancer mark P300) are shown as well. Relevant common peak positions in a and b are highlighted by a gray shading.

Putative enhancer region 1 consists of two distinct enhancers

Since active promoters and enhancers reside in open chromatin, we profiled accessible chromatin by the assay for transposase-accessible chromatin (ATAC) in the two AML patient samples AML-T1 and -T2 and in both iPSCs and HSPCs of our cell line model. We found eleven consistent open chromatin sites common to the patient samples and to the ChiPSC22$^{t(7;12)}$ and ChiPSC22^{WT} cells (**Fig. 5a**). Four of those overlapped with the enhancers mapped in ChiPSC22$^{t(7;12)}$ **(Fig. 4a, 4b**). Enhancer region 1 showed interaction in Hi-C data in both iPSC and HSPC ChiPSC22$^{t(7;12)}$. To clarify if these are identical interactions, we zoomed into the enhancer region to better resolve ATAC (**Fig. 5b**), H3K27ac (**Fig. 5c)** and H3K4me1 (**Fig. 5d**) peaks. The ATACseq data clearly shows the presence of two open chromatin regions (putative enhancer region 1a and 1b). Putative enhancer 1a was detected in iPSC cells and putative enhancer 1b was only present upon differentiation of ChiPSC22$^{t(7;12)}$ into HSPCs. The interaction of putative enhancer 1a with the *MNX1* promoter seen in ChiPSC22$^{t(7;12)}$ iPSCs represents an interaction of an active enhancer (presence of H3K4me1 and H3K27ac marks) with the *MNX1* promoter, however, does not result in activation of *MNX1*. The function of this interaction remains to be elucidated, since no other gene on the translocated part of chr7 in the immediate vicinity (e.g. *MNX1-AS1*, *UBE3C*, *DNAJB6*) shows expression in iPSCs and nor in HSPCs. However, the interaction of putative enhancer 1b with the *MNX1* promoter is newly established in ChiPSC22$^{t(7;12)}$ HSPCs and leads to the activation of *MNX1*.

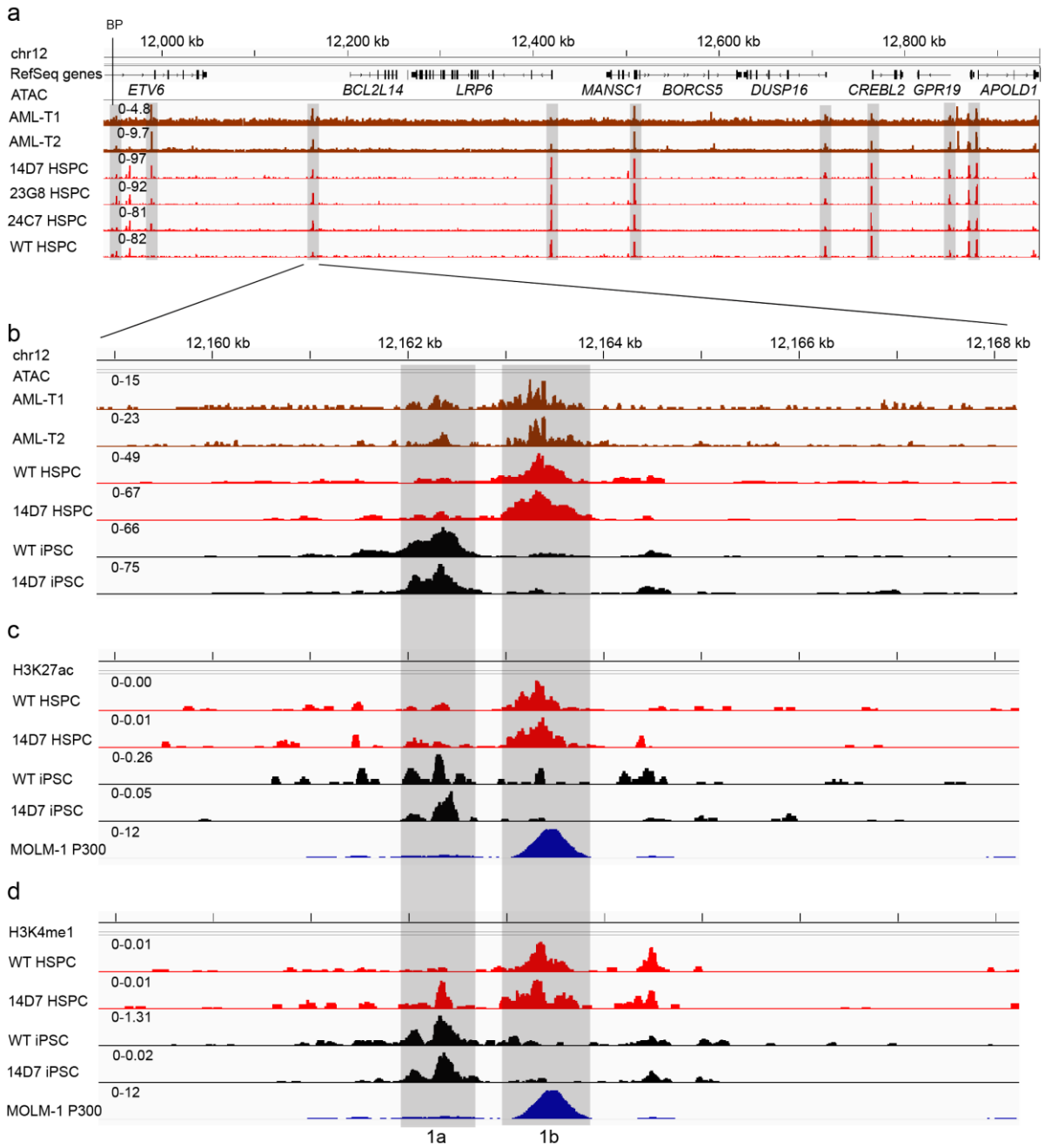

Fig. 5 Open chromatin and enhancer mark profiles in the *ETV6* region of patient and cell line samples. a, Open chromatin profiles of AML patients AML-T1 and -T2 and of ChiPSC22$^{t(7;12)}$ sublines 14D7, 23G8 and 24C7 and of ChiPSC22^{WT} HSPCs 1 MB distal to the breakpoint (BP) in *ETV6*. **b,** High resolution open chromatin profiles in the enhancer region 1 of AML-T1, AML-T2 and of iPSCs and HSPCs of ChiPSC22$^{t(7;12)}$ and ChiPSC22^{WT}**. c** and **d,** High resolution H3K27ac and H3K4me1, respectively, profiles in the enhancer region 1 of iPSCs and HSPCs of ChiPSC22$^{t(7;12)}$ and ChiPSC22^{WT}**.** A publicly available P300 profile of MOLM-1 is shown as well. Relevant common peak positions are highlighted by a gray shading. Enhancer region 1 is resolved into enhancer 1a, active in iPSCs, and enhancer 1b, active in HSPCs.

Hematopoietic enhancers in the *ETV6* region interact with the *MNX1* promoter in differentiated ChiPSC22$^{t(7;12)}$ cells

As a further layer of experimental evidence for *de novo* promoter-enhancer interactions in t(7;12)(q36;p13) AML, we examined *MNX1* expression levels in three monoclonal ChiPSC22$^{t(7;12)}$ sublines carrying a CRISPR-Cas9-engineered 212 kb deletion which removed all four strong enhancer candidates on the translocated allele (ChiPSC22$^{t(7;12)\Delta En}$, **Fig. 6a, Extended Data Fig. 3**). In all three ChiPSC22$^{t(7;12)\Delta En}$ lines, HSPC-specific *MNX1* expression was abrogated (**Fig. 6b**). This supports the hypothesis that *MNX1* activation in ChiPSC22$^{t(7;12)}$ is the result of interactions between the *MNX1* promoter and one or multiple enhancers in the vicinity of *ETV6.* We also observed a significant downregulation of genes that are upregulated together with *MNX1* in t(7;12)(q36;p13), such as *AGR2*, *MMP9*, *MAF* and *CRISP3*, suggesting that they are regulated by *MNX1*.

We, furthermore, performed 2-color FISH targeting the *MNX1* promoter and a 5 kb region around the enhancer candidate 1 (probe set in **Supplementary Table 7**) located most distant to the breakpoint and overlapping with the region that showed interaction in Hi-C. We observed a decreased 3D distance between the targeted regions in ChiPSC22$^{t(7;12)}$ HSPCs as compared to the corresponding iPSCs (median HSPC: 0.53 µm, median iPSCs: 1.06 µm, **Fig. 6c**). This suggests a reinforced contact between the enhancer candidate in the *ETV6* region and the *MNX1* promoter upon differentiation to HSPCs.

In conclusion, we provide several layers of experimental evidence for an enhancer-hijacking event rather than the creation of an oncofusion protein in pediatric AML with t(7;12)(q36;p13). Using an *in vitro* iPSC/HSPC cell system, we demonstrate that one or several enhancer elements in the *ETV6* region interact with the *MNX1* promoter to regulate *MNX1* expression.

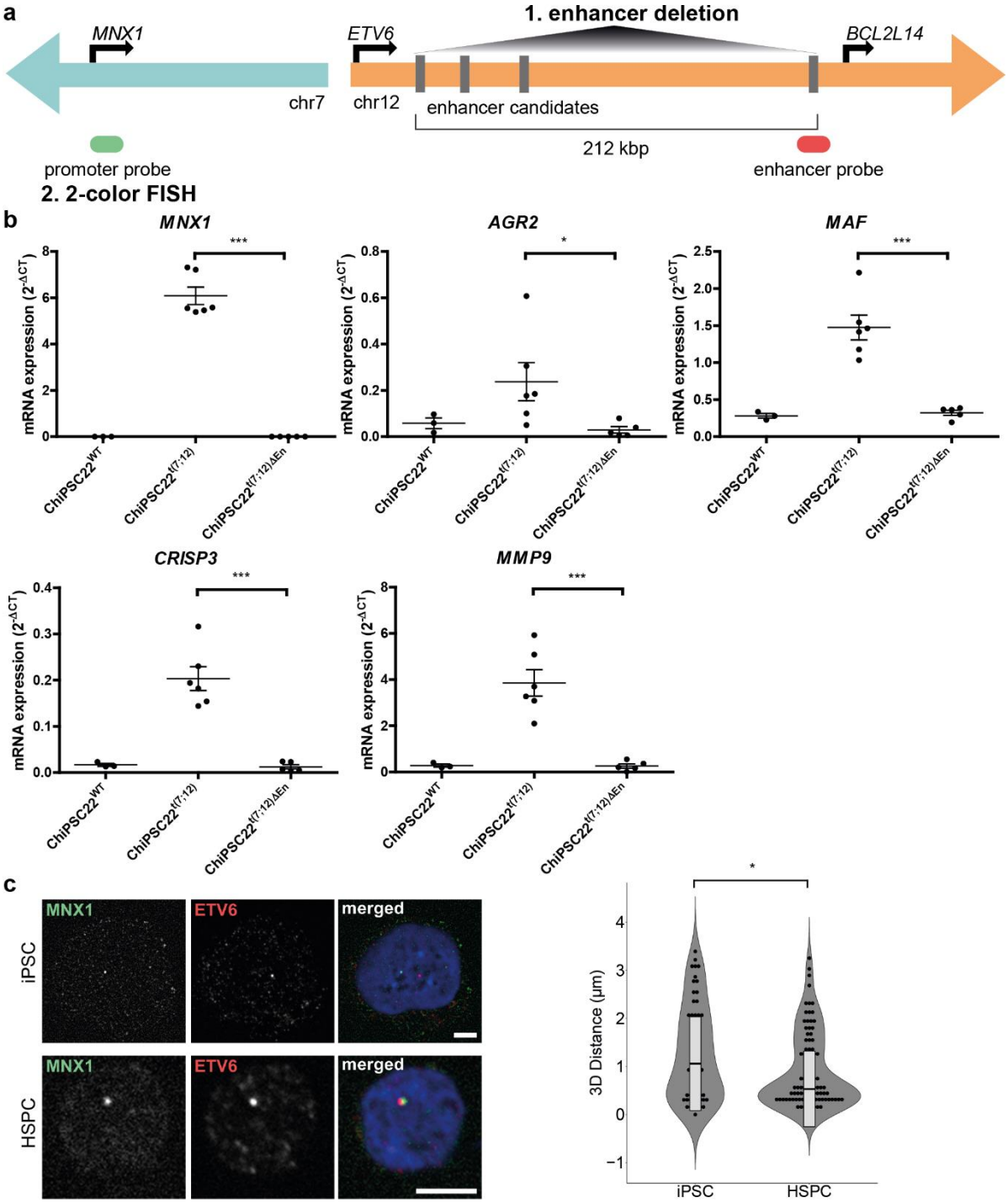

Fig. 6 Molecular validation of enhancer-promoter interaction in ChiPSC22$^{t(7;12)}$ upon differentiation. a, Scheme of experiments performed to validate the interaction between the *MNX1* promoter and enhancers distal to the breakpoint including deleting the enhancer region and spatial proximity probing. **b,** Gene expression in ChiPSC22^{WT} (n=3), ChiPSC22$^{t(7;12)}$ (n=6, from 3 independent cell lines) and ChiPSC22$^{t(7;12)\Delta En}$ (n=5, from 3 independent cell lines) measured via qRT-PCR and shown as $2^{-\Delta Ct}$ vs. *GUSB as* endogenous reference. Unpaired, two-tailed t-test with $*P < 0.05$ and $***P < 0.0001$. **c,** Increased proximity between *MNX1* and *ETV6* in ChiPSC22$^{t(7;12)}$ derived HSPCs compared to iPSCs. Left: 2-color FISH targeting a 5 kb region at *MNX1* promoter and putative enhancer 1 in *ETV6* in iPSCs and HSPCs. Representative single planes are depicted. Note that the iPSC representative image was scaled down. Scale bars, 5 µm. Right: 3D distances between *MNX1* and *ETV6* upon differentiation. Distances under 3.5 µm are displayed. Black horizontal lines within boxes represent median values. The limits of the boxes indicate the upper and lower quartiles. iPSC: n = 64 , HSPC: n = 90, across 3 independent replicates. $*P < 0.05$, Wilcoxon rank sum test.

Discussion

In this manuscript, we describe enhancer-hijacking and activation of *MNX1* as a novel molecular mechanism resulting from a translocation between chr7 and 12 [t(7;12)(q36;p13)] in pediatric AML. Our study shifts the focus from a putative *MNX1::ETV6* oncofusion transcript to the activation of *MNX1* as the unifying putative leukemia-driving event. Overexpression of *MNX1* is accompanied by mono-allelic inactivation of *ETV6* on the translocated chromosome, putatively resulting in haploinsufficiency. In addition, the majority of published t(7;12)(q36;p13) AML cases show trisomy 19 as a recurrent secondary event, suggesting that activation of a gene on chr19 is an additional component in the leukemogenesis of this AML subtype [22]. Our observation has important implications for the diagnosis of this subgroup of patients, as well as novel therapeutic approaches. As shown in the expression reanalysis, all t(7;12)(q36;p13) AML demonstrate overexpression of *MNX1*. Only few AML cases without t(7;12)(q36;p13) show *MNX1* overexpression, but while t(7;12)(q36;p13) AML is diagnosed at very young age (<20 months), *MNX1* overexpression in the absence of this translocation occur predominantly at a later age, suggesting different, yet, unexplained molecular pathways converging in *MNX1* expression. Considering that *MNX1* is not expressed in the normal hematopoietic system, quantitative *MNX1* expression analysis could be used as a diagnostic marker for this subgroup of pediatric AML.

A few examples of enhancer-hijacking have been reported in the literature for myeloid leukemia in inv(3;3)/t(3;3) AML [11] and atypical 3q26/MECOM rearranged AML [12]. Enhancer-hijacking in the AML cell line, GDM-1, was recently reported as a result of a t(6;7)(q23;q36) and leads to activation of *MNX1* through the interaction of the *MNX1* promoter with an enhancer in the *MYB* locus [10]. Enhancer-hijacking events resulting in the activation of proto-oncogenes have been described also in other human malignancies including translocations resulting in the activation of oncogenes *MYC*, *BCL2* or *CCND1* in B-cell lymphoma [23–25] or rearrangements in medulloblastoma [26]. Systematic approaches using genetic and transcriptomic data for the discovery of structural variant-driven proto-oncogene activation have recently been reported by the 'Pan-Cancer Analysis of Whole Genomes' (PCAWG) Consortium, identifying an unexpected number of such enhancer-hijacking events in cancer genomes [27], suggesting that many of these events have been overlooked. Gene activation through promoter-enhancer interaction within the nucleus seems to be regulated within structural domains of megabase scale, so-called "topologically associating domains" (TADs) [28]. TADs create compartments in the form of chromatin loops, in which gene expression patterns are tightly regulated. These structures are demarcated by boundary elements including DNA methylation-sensitive binding sites for the CCCTC-binding factor (CTCF) and the cohesion complex [29]. Loss or rearrangements of these boundaries lead to an aberrant

interaction of enhancers with novel promoter elements. Subtype-specific 3D genomic alterations were recently discovered in AML leading to enhancer-promoter loops as well as enhancer-silencer loops [30]. We demonstrated that undifferentiated ChiPSC22$^{t(7;12)}$ cells already establish enhancer-promoter loops (putative enhancer 1a with MNX1 promoter) within the *MNX1*/*ETV6* region on the translocated chromosome, but this event alone does not lead to *MNX1* activation, despite active H3K27ac marks. The molecular consequences of this interaction are currently unclear since no other gene in the immediate vicinity (e.g. *MNX1-AS1*, *UBE3C*, *DNAJB6*) shows expression in iPSCs and also not in HSPCs. One possibility is that putative enhancer 1a represents a silencer element interacting with the *MNX1* promoter [31]. Upon differentiation, interaction with 1a is lost and a new interaction between the putative enhancer 1b, just 1 kb afar, and the *MNX1* promoter is established. This interaction is associated with the *MNX1* expression.

Initial evidence for an oncogenic role of MNX1 in leukemogenesis comes from a study by Nagel *et al*. characterizing the *MNX1*-overexpressing cell line GDM-1 [32]. In this study, the authors used antisense oligonucleotides against *MNX1* to reduce its expression. This led to a reduction of cell viability and cell adhesion. Ingenhag et al. demonstrated that *in vitro* overexpression of *MNX1* in HT1080 and NIH3T3 cells leads to premature, oncogene-induced senescence mediated by the induction of p53-signalling [33]. *In vivo*, ectopic *MNX1* expression in murine hematopoietic stem and progenitor cells resulted in strong differentiation arrest and accumulation at the megakaryocyte/erythrocyte progenitor stage [33]. Overall, this phenotype is in line with reports on t(7;12)(q36;p13) AML blast cells which are less differentiated (FAB subtype M0 or M2) and demonstrate expression of the stem cell markers CD34 & CD117 [5,34]. Waraky et al. used retroviral transduction of *MNX1*-expressing constructs into murine fetal HSPCs and were able to induce AML [35]. A possible link to leukemogenesis was described with the observation that MNX1 activation resulted in reduced H3K4me1/2/3 and H3K27me3 levels providing increased chromatin accessibility [35].

Our study challenges the dogma in AML that all reciprocal translocations lead to oncofusion proteins as an overestimated molecular mechanism in AML for gene activation. Future studies unraveling the molecular defects of t(7;12)(q36;p13) AML should focus on the targets of homeobox transcription factor *MNX1* rather than the oncofusion, as already initiated in some reports [17,33,36]. However, also the role of genes on chromosome 19 (e.g. *DNA methyltransferase 1*, *DNMT1* or *RNA polymerase II transcriptional elongation factor*, *ELL*), and effects of *ETV6* haploinsufficiency should be considered as co-factors in the leukemogenic process. ETV6 is a strong transcriptional repressor, and haploinsufficiency could result in reactivation of its target genes. Recent work highlighted that ETV6 binds GGAA microsatellite enhancers and prevents them from being activated through histone

acetylation [37]. Upon *ETV6* loss, the ETS transcription factor ERG can bind these sites and activate the target genes.

Methods:

Samples and cell line

Pediatric leukemia samples T1, T2 and T3 were obtained at diagnosis after informed consent of patients and/or their legal guardians in accordance with the institution's ethical review board (Essen, Hannover MHH, No. 2899). T4 came from a pediatric AML cohort in Gothenburg, Sweden, and collected after informed consent was obtained from the legal guardians in accordance with the local ethical review board (Gothenburg). The human iPSC line, ChiPSC22 (Cellartis/Takara Bio Europe AB, Gothenburg, Sweden), was maintained in the feeder-free DEF-CS culture system (Cellartis/Takara Bio Europe) under standard conditions. Before differentiation, the cells were transferred to Matrigel (BD Biosciences, Stockholm, Sweden) and mTeSR1 medium (StemCell Technologies Inc., Grenoble, France) for 2-3 passages. Cell line ChiPSC22 was authenticated, and all cell lines were regularly tested for mycoplasma contamination using a commercial test kit (VenorGeM Classic, Minerva Biolabs, cat.no. 11-1050, Berlin, Germany).

Differentiation to hematopoietic cells

The iPSC lines were cultured in mTeSR-1 medium (StemCell Technologies Inc.) on Matrigel (Corning Inc.) for 3-4 passages as clusters using Gentle Cell Dissociation Reagent (Stemcell Technologies Inc.). The STEMdiff Hematopoietic kit (Stemcell Technologies Inc.) was used for the differentiations according to the protocol. On day 12 of the differentiation, the hematopoietic cells were harvested and analyzed by epigenomics profiling (ATAC-seq, ACT-seq, 4C-seq). For mRNA expression analysis, the cells were recovered overnight in StemSpan SFEM II + CC100 (Stemcell Technologies Inc.) and RNA was isolated from the CD34+CD43+CD45+ cell pool, which was sorted on a FACS Aria (BD Biosciences).

Deletion of the enhancer region

For the deletion of the putative enhancer region in ChiPSC22$^{t(7;12)}$, we chose a sequence of 212 kbp containing the four strongest enhancer candidates (chr12: 11,798,090-12,011,665; GRCh37/hg19).

CRISPR/Cas9 editing was done as previously described [9]. In brief, CRISPR RNAs were designed and synthesized using the Alt-R Custom Cas9 crRNA Design Tool (Integrated DNA Technologies, IDT). GuideRNAs (gRNA) were generated by annealing the crRNAs with ATTO^{TM}550-labeled tracrRNA (from IDT). A gRNA:Cas9 ribonucleoprotein (RNP) complex was formed by mixing gRNA and Cas9 in a 1.2:1 (60 pmol:52 pmol per transfection). To join the two ends by homology-directed repair, a 150 nt single-stranded deoxynucleotide (ssODN)

was designed and synthesized by IDT, with sequence homology to 75 nt on each side of the double-strand break. We chose three cell lines for transfection: WT ChiPSC22^{WT} and two sublines ChiPSC22$^{t(7;12)}$ (14D7, 24C7) from Nilsson et al. [9]. Nucleofection was done using the 4D-Nucleofector from Lonza with the 16-well nucleocuvette and program CA-137. Per transfection, a 20 µL mix containing 3x10^5 cells in nucleofector solution P3 (Lonza), 2 µL RNP complex and 200 pmol ssODN was prepared. After the pulse, 75 µl DEF-CS medium (Cellartis, Takara Bio Europe) was added and the cell suspension was transferred to a COAT-1-coated (Cellartis, Takara Bio Europe) 24-well plate with DEF-CS. After 48 h, the ATTO550^{TM}-positive cells were sorted by a FACS Aria (BD Biosciences) and transferred to COAT-1-coated 96-well plates. Limiting dilutions with 0.5 cells/well were seeded in COAT-1-coated 96-well plates. Lysates were taken using the Extraction buffer (Takara Bio Europe)and the presence of the deletion on the translocated allele as well as the wildtype allele were validated using PCR (Terra PCR Direct Polymerase Mix, Takara Bio Europe, **Supplementary Table 6**). From ChiPSC22$^{t(7;12)}$ 14D7 cell line 2304B4 was generated and from ChiPSC22$^{t(7;12)}$ 24C7 the cell lines 2305B10 and 2305C9 were generated.

Whole genome sequencing

Genomic DNA was isolated using the Quick DNA Miniprep kit (Zymo research, Freiburg, Germany) according to the manufacturer's protocol. Sequencing of two pediatric t(7;12)(q36;p13) AML cases without paired germline DNA was done in an Illumina HiSeq X Ten sequencer in one lane with the targeted depth of 40x and insert size 300 bp. The fastq files were aligned with BWA-MEM to the hs37d5_PhiX reference genome, which corresponds to the hg19 reference to which decoy sequences were added. SNVs were called using mutect2, and since no matched normal samples were available, we only looked for mutations in 52 known AML driver genes: ASXL1, ASXL2, BCOR, CEBPA, CEBPG, CREBBP, DNMT3A, DNMT3B, ETV6, EZH2, FLT3, GATA2, IDH1, IDH2, JAK2, JARID2, KAT6A, KDM3B, KDM6A, KIT, KMT2A, KMT2C, KMT2D, KMT2E, KRAS, MED12, NCOR1, NCOR2, NF1, NOTCH1, NOTCH2, NPM1, NRAS, NSD1, PHF6, PTPN11, RB1, RUNX1, SF3B1, SMARCA2, SMARCA4, SMC1A, SMC3, SRSF2, STAG2, SUZ12, TET1, TET2, TP53, U2AF1, WT1 and ZRSR2. SVs and SCNAs were called using the Hartwig Medical Foundation (HMF) pipeline (https://github.com/hartwigmedical/hmftools), which consists of identifying SVs with GRIDSS2 [38] and allele-specific SCNAs with COBALT and AMBER, and integrating SVs and SCNAs using PURPLE. Since matched normal samples were not available, we used the HMF tools in tumor-only mode, and filtered out the putative germline SVs using a large panel of normals provided by the HMF, as well as removing SVs smaller than 20kb or resulting in small insertions. For the two samples of the TARGET-AML dataset

(dbGaP accession: phs000465.v22.p8), we downloaded the processed data using the globus platform.

RNA isolation and sequencing

Total RNA was isolated using the RNeasy Plus Mini kit (Qiagen, cat.no. 74136, Hilden, Germany).
RNA was sequenced on NOVASEQ 6000 with 100 bp paired-end. The fastq files were processed using the nf-core [39] RNAseq v3.9 pipeline, with the alignment performed using STAR [40] and the quantification performed with Salmon [41].

Protein extraction and Western blot

Washed cells were 30 min lysed on ice followed by 5 min at 97°C in 62.5 mM Tris-HCl pH 6.8, 2% sodium dodecyl sulfate (SDS), 10% glycerol, 1 mM DTT, 1 mM NaVO4, 5 mM NaF supplemented with cOmplete protease inhibitor (Roche, 11697498001), Benzonase (Santa Cruz, sc-202391) and PhosStop (Roche, 4906837001). The suspension was centrifuged for 5 min in a benchtop centrifuge, and the protein concentration of the supernatant was either determined with the Qubit protein assay kit (ThermoFisher Scientific, Q33212) or the Bicinchoninic Acid kit (Sigma-Aldrich, B9643-1L-KC and C2284-25ML-KC). Thirty µg protein in 4x TruPage LDS sample buffer (Sigma, PCG3009-10ML) were run per lane of a 4-20% SDS-PAGE gradient (BioRad, 4561096) and subsequently transferred to a polyvinylidene difluoride (PVDF) membrane (Milipore, IPVH00010). The membrane was probed with a primary antibody against MNX1 followed by incubation with a secondary antibody coupled to horseradish peroxidase (**Supplementary Table 3**). Protein signals were visualized by membrane incubation with the Novex ECL HRP Chemiluminescent Substrate Reagent Kit (Invitrogen, WP20005) and an Amersham Imager 680 (GE Healthcare). For loading quantity control, the membrane was re-probed with a horseradish peroxidase-coupled antibody against β-actin (**Supplementary Table 3**).

RT-PCR

Total RNA was isolated using the RNeasy Plus Mini kit (Qiagen, Solna, Sweden) and cDNA synthesized with SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen/Life Technologies). Quantitative RT-PCR was performed using TaqMan Universal Master Mix II with UNG and TaqMan Gene expression assays (ThermoFisher Scientific, Applied Biosystems) for MNX1 (Hs00907365_m1), AGR2 (Hs00356521_m1), MMP9 (Hs00957562_m1), CRISP3 (Hs00195988_m1), MAF (Hs04185012_s1) and, as reference gene, GUSB (Hs99999908_m1). The reactions were performed on the ABI 7900HT Sequence Detection System (Applied Biosystems/Life Technologies) with GUSB as the

reference gene. Gene expression was calculated using the ΔCt method and is presented as $2^{-\Delta Ct}$.

Expression screens

We downloaded CEL files for the Balgobind et al. cohort from GEO (GSE17855) and normalized them using the affy R package. We downloaded processed RNAseq expression data from the TARGET-AML cohort (both TARGET-NCI and TARGET-FHCRC) from https://target-data.nci.nih.gov/Public/AML/mRNA-seq/L3/expression/BCCA/. For the microarray data we assumed that the log-expression values are distributed normally, and used the 3-sigma rule to identify the samples expressing MNX1. Since MNX1 is not expressed in normal hematopoietic cells, in RNAseq we considered an expression >0.5 TPM for MNX1 as overexpression.

Circular chromosome conformation capture (4C)

From each cell type, about two million cells were used for circular chromosome conformation capture (4C) essentially according to [43]. Two rounds of restriction digestion/T4 DNA ligation were applied, using *Hin*dIII in combination with *Dpn*II and *Nla*III in combination with *Csp*6I. In a first PCR step, second ligation products, inverse primers (**Supplementary Table 4**) and Q5 high fidelity enzyme (New England Biolabs, Frankfurt a.M., order no. M0491) were used with reaction conditions 98°C for 30 sec, 10 cycles with 98°C for 15 sec, 63°C or 61°C, depending on the viewpoint, for 20 sec with 0.5°C touch-down per cycle, 72°C for 2 min, then 25 cycles with 98°C for 15 sec, 58°C or 56°C, depending on the viewpoint, for 20 sec, 72°C for 2 min, finally followed by 72°C for 1 min. PCR products were HighPrep bead-purified (Biozym, Hessisch Oldendorf, order no. 220002), and their DNA concentrations were analyzed with the Qubit dsDNA HS Assay (Thermo Fisher, Karlsruhe, order no. Q32854). Amplification of sequencing libraries was monitored under real-time conditions using about 3 ng DNA from the previous PCR according to [44] in 25 µl reaction volumes with Kapa 2G Robust Hot Start ReadyMix (Merck, Darmstadt, order no. KK5702), 95°C, 3 min (initial melting) and 95°C, 20 sec, 62°C, 15 sec, 72°C, 40 sec (cycling). Library DNAs were purified and quantitated as described above, and fragment sizes were determined with a TapeStation 4150 (Agilent, Waldbronn). Sequencing was done with pools of differently barcoded libraries mixed in equimolar ratios on an Illumina NextSeq 550 instrument with PE, 75 bp, mid-output mode.

Hi-C

Hi-C libraries were prepared and data analysis was performed as previously described [45] in supplementary text with the following modification: 1x10^6 cells per replicate were fixed at a

final concentration of 1% formaldehyde in RPMI 1640 medium. Digestion was performed using the DpnII restriction enzyme. Two to three Hi-C library replicates of each sample were sequenced at depth of 240 million reads per replicate. The fastq files were processed using the nf-core/hic v.2.1.0 pipeline. NeoLoopFinder was used to identify neoloops and to plot the contact maps [30].

Antibody-guided Chromatin Tagmentation (ACT-seq)

Genome-wide targeting of histone modifications and transcription factor ETV6 was done by ACT-seq, largely as published [46]. The pA-Tn5ase protein was isolated from E. coli (C3013, New England Biolabs) transformed with plasmid pET15bpATnp (#121137, Addgene). The pA-Tn5 transposome (pA-Tn5ome) was generated by mixing pA-Tn5ase (final concentration either 1.9 µM or 3.3 µM, depending on the pA-Tn5ase preparation) and Tn5ME-A+B load adaptor mix (final concentration 3.3 µM) in complex formation buffer CB [46]. The pA-Tn5ome-antibody (pA-Tn5ome-ab) complexes were formed by mixing 1 µl pA-Tn5ome with 0.8 µl CB and 0.8 µl solution of antibody (**Supplementary Table 3**). For pA-Tn5ome-ab complex binding and tagmentation, 200,000 (for ETV6 targeting) or 50,000 (all other targets) were used. For normalization of sequence reads between biological replicates, ~4,000 permeabilized nuclei of yeast Saccharomyces cerevisiae, prepared according to [47] and incubated with pA-Tn5ome-ab complex targeting yeast H2B, were spiked into each mix of cells and pA-Tn5ome-ab complex. Tagmented DNA was purified with a MinElute kit (Qiagen, 5cat.no. 28004) and eluted with 20 µl elution buffer (EB). Sequencing libraries were generated under real-time conditions with a LightCycler 480 (Roche, Mannheim) in 50 µl reaction mixes consisting of 20 µl tagmented DNA eluate, 25 µl NEBNext High Fidelity 2X Mix (#M0541, New England Biolabs), 0.5 µl 100xSYBRGreen, 2.5 µl primer Tn5McP1n and 2.5 µl barcode primer (**Supplementary Table 5**). Reaction conditions were 72°C, 5 min (gap repair); 98°C, 30 sec (initial melting); 98°C, 10 sec, 63°C, 10 sec, 72°C, 10 sec (cycling). With a few exceptions of weakly amplifying samples, cycling was stopped when the increase of fluorescence units (FUs) was 5 or higher. Libraries were purified with HighPrep magnetic beads (#220001, Biozym) with a bead:DNA ratio of 1.4:1 and 12 µl EB. Quantity and fragment size of the libraries were determined with a Qubit dsDNA HS assay kit (#Q32854, Invitrogen) and a TapeStation 4150 with D1000 High Sensitivity Assay (#5067- 5585, Agilent), respectively. Libraries with sequencing compatible barcodes were pooled in six to ten multiplexes and sequenced on a single lane of a NextSeq 550 system (paired-end, 75 bp) with mid-output at the Genome and Proteome Core Facility of the DKFZ.

Assay for transposase-accessible chromatin by sequencing (ATAC-seq)

ATAC-seq was done essentially as described [48] using about 50,000 cells and the Nextera DNA library prep kit (Illumina, Berlin, order no. 15028212). Libraries were generated and processed as described for ACT-seq, but cycling conditions were 98°C, 10 sec, 63°C, 30 sec, 72°C, 30 sec.

4C-seq, ACT-seq and ATAC-seq data analysis

4C-seq data processing and analysis was done with the pipe4C pipeline [49] using single reads starting with a *Hin*dIII-, *Nla*III- or *Csp*6I-site containing viewpoint primer, depending on the viewpoint; the pipe4C pipeline was applied with default parameters under R3.6.2. ACT-seq and ATAC-seq data was processed with Trim Galore v. 0.4.4 (http://www.bioinformatics.babraham.ac. uk/projects/trim_galore) and Cutadapt v. 1.14 (https://doi.org/10.14806/ej.17.1.200) using the non-default parameters "--paired", "--nextera", "--length_1 35", and "--length_2 35" for adapter and quality trimming. Bowtie2 v. 2.2.6 [50] was used with the "--very-sensitive" flag and a maximum insertion length of 2500 bp to map trimmed reads against hg19. Aligned reads belonging to the same lane-multiplexed library were combined using SAMtools merge v. 1.5 [51]. PCR duplicates were removed by Picard MarkDuplicates v. 2.17.4 for ATAC-seq but not ACT-seq data. Discordant mappings and alignments with a Phred score below 20 were removed using SAMtools view. Aligned fragments smaller than 38 bp are tagmentation incompatible [52] and, hence, were discarded. The ends of reads were adjusted to represent the center of the transposition event as previously described by [53]. For ACT-seq, trimmed reads were additionally aligned against the *Saccharomyces cerevisiae* R64 reference genome followed by post-alignment filtering as described above. To derive a library-specific scaling factor, we calculated the multiplicative inverse of the number of filtered alignments against the yeast genome. Coverage tracks were generated using the bamCoverage functionality of Deeptools v. 3.1.1 [54] with the non-default parameters "--ignoreForNormalization chrM chrY chrX" and "--effectiveGenomeSize 2652783500" as well as the "--scaleRatio" option to specify the spike-in-derived scaling factor. For ATAC-seq, employing a custom script, the accessibility signal was smoothed by centering a 73 bp window on the transposition event's midpoint of each read and the resulting tag coordinates were used for all downstream analysis. The analysis procedures were implemented as fully containerized workflows using the Common Workflow Language v. 1.0 (https://doi.org/10.6084/m9.figshare.3115156.v2). ACT-seq peak calling was done applying MACS v.2.2.6 (https://pypi.org/project/MACS2/) with a q-value cutoff of 0.05 and default parameters using a wrapper script, callpeaks (https://www.encodeproject.org/chipseq/histone/) with settings narrowPeak and broadPeak for H3K27ac and H3K4me1, respectively. To facilitate visualization of hematopoietic-specific enhancers in the Integrative Genomics Viewer (IGV, version 2.11.7 [55], we generated HSPC-

specific H3K27ac and H3K4me1 bw-tracks with callpeaks using corresponding iPSC data as internal reference.

Probe Selection and Synthesis

We designed oligonucleotide probe sequences to target MNX1 (hg19: chr7:156802250-156807250, hg38: chr7:157009556-157014556) or ETV6 (hg19: chr12:12,160,800-12,165,800, hg38: chr12:12007866-12012866) in FISH assays. Selection and filtering of oligonucleotides was performed according to previous works (Brandstetter et al.). Probes were synthesized to contain 40-nt sequences that bind to the genome and 30-nt sequences that carried regularly spaced ATTO594 or ATTO647N molecules (unpublished). MNX1 and ETV6 probe sets comprised 73-75 oligonucleotides each spanning roughly 5 kb in the genome. For a list of synthesized probes used in this study see **Supplementary Table 7**.

Two-Color FISH

Two-color FISH was conducted as previously described with minor adaptations [56,57]. Undifferentiated ChiPSC22 were seeded on DEF-CS COAT-1-coated (Cellartis, Takara BioSciences) coverslips and differentiated ChiPSC22 on poly-L-lysine-coated coverslips. Cells were washed three times in PBS and fixed for 10 min using methanol-free 1x PBS/ 4% formaldehyde (Polysciences) solution. Cells were treated with 1x PBS/ 0.5% Triton X-100 for 15 min, washed in 1x PBS and treated with 0.1 M HCl (Roth) for 5 min followed by two washes with 2x SSC (Thermo Fisher Scientific) for 5 min and incubation in 1 µg/ml RNase (Thermo Fisher Scientific) for 30 min at 37°C in a wet chamber. Cells were then pre-equilibrated in 2x SSC/ 50% formamide (Sigma) (adjusted to pH 7.0) for 1 h or overnight. Coverslips were inverted onto 8 µl of hybridization solution containing 50% formamide (Sigma), 10 % dextran sulfate (Sigma), 0.1 % Tween-20 (Roth) and 250 nM labeled probes and sealed with rubber cement. Slides were placed on a heat block set to 81°C for 3 min and incubated at 37°C for 16 – 20 h. The samples were then washed twice with 2x SSC for 15 min, twice with 0.2x SSC/ 0.2% Tween-20 for 7 min at 56°C and once with 4x SSC/ 0.2% Tween-20 for 5 min at room temperature. DNA was counterstained with DAPI (100 µg/ml in 2x SSC, Sigma-Aldrich) for 5 min or DiYO™-1 (25 nM in 2x SSC, AAT Bioquest) for 20 min, followed by two more washes in 2x SSC/ 0.2% Tween-20. Coverslips were mounted on microscopic slides with Mowiol (2.5% DABCO, pH 7.0; Carl Roth), dried for 30 min and sealed with nail polish[58].

Confocal Fluorescence Microscopy and Image Analysis

3D FISH images were acquired using a Nikon TiE microscope equipped with a Yokogawa CSU-W1 spinning-disk confocal unit (50 µm pinhole size), an Andor Borealis illumination unit, Acal BFi (Gröbenzell) laser beam combiner (405 nm/488 nm/561 nm/640 nm), Andor IXON

888 Ultra EMCCD camera, and a Nikon 100×/1.45 NA oil immersion objective. The microscope was controlled by software from Nikon (NIS Elements, ver. 5.21.03). Confocal image z-stacks were acquired with a step size of 300 nm.

Distances between *MNX1* and *ETV6* were measured using a semi-automatic in-house script in Fiji (v1.54f [59]). Nuclear segmentation masks were generated by thresholding. ATTO594 and ATTO647N spots were detected in the nucleus through thresholding and filtered by their size and shape. Co-occurring signals within 25 x 25 pixels (xy) were included in the analysis. Local maxima were determined using 3D ImageJ Suite (V3.96) and the distances were calculated using 3D Euclidean distance [60]. Only distances smaller than 3.5 µm were included in the analysis.

Data availability

All sequencing data generated here are based on human reference GRCh37/hg19 and were deposited at the NCBI Gene Expression Omnibus (GEO) under the accession number:

Acknowledgments:
The authors would like to thank the Genomics and Proteomics Core Facility, the Omics IT and Data Management Core Facility of the DKFZ Heidelberg and the Center for Advanced Light Microscopy (CALM) at the Ludwig-Maximilians-University Munich for their excellent support. The work was in part supported by funds from the Helmholtz Foundation (ENHANCE Program) and German Research Foundation, SFB1074 subproject B11N (to CP and AR), Carreras Foundation (CP, ES) and FOR2674 subprojects A1, A6, A9 (CP, DBL), the Swedish Cancer Society (20 0925 PjF, CAN2017/461), the Swedish Childhood Cancer Foundation (PR2021-0025 and TJ2022-0017) and Västra Götalandsregionen (ALFGBG-431881), to LP and the SFB1064 subproject A17 (HL). Further support comes from the Helmholtz International Graduate School (AR, ES)

References:

1. de Thé, H., Pandolfi, P. P. & Chen, Z. Acute Promyelocytic Leukemia: A Paradigm for Oncoprotein-Targeted Cure. *Cancer Cell* **32**, 552–560 (2017).
2. Tosi, S. *et al.* Identification of new partner chromosomes involved in fusions with the ETV6 (TEL) gene in hematologic malignancies. *Genes Chromosomes Cancer* **21**, 223–229 (1998).
3. Wlodarska, I. *et al.* Fluorescence in situ hybridization characterization of new translocations involving TEL (ETV6) in a wide spectrum of hematologic malignancies. *Blood* **91**, 1399–1406 (1998).
4. Espersen, A. D. L. *et al.* Acute myeloid leukemia (AML) with t(7;12)(q36;p13) is associated with infancy and trisomy 19: Data from Nordic Society for Pediatric Hematology and Oncology (NOPHO-AML) and review of the literature. *Genes*

Chromosomes Cancer **57**, 359–365 (2018).
5. von Bergh, A. R. M. *et al.* High incidence of t(7;12)(q36;p13) in infant AML but not in infant ALL, with a dismal outcome and ectopic expression of HLXB9. *Genes Chromosomes Cancer* **45**, 731–739 (2006).
6. Ballabio, E. *et al.* Ectopic expression of the HLXB9 gene is associated with an altered nuclear position in t(7;12) leukaemias. *Leukemia* **23**, 1179–1182 (2009).
7. Beverloo, H. B. *et al.* Fusion of the homeobox gene HLXB9 and the ETV6 gene in infant acute myeloid leukemias with the t(7;12)(q36;p13). *Cancer Res.* **61**, 5374–5377 (2001).
8. Tosi, S. *et al.* Heterogeneity of the 7q36 breakpoints in the t(7;12) involving ETV6 in infant leukemia. *Genes Chromosomes Cancer* **38**, 191–200 (2003).
9. Nilsson, T. *et al.* An induced pluripotent stem cell t(7;12)(q36;p13) acute myeloid leukemia model shows high expression of MNX1 and a block in differentiation of the erythroid and megakaryocytic lineages. *Int. J. Cancer* **151**, 770–782 (2022).
10. Weichenhan, D. *et al.* Translocation t(6;7) in AML-M4 cell line GDM-1 results in MNX1 activation through enhancer-hijacking. *Leukemia* **37**, 1147–1150 (2023).
11. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369–381 (2014).
12. Ottema, S. *et al.* Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia. *Blood* **136**, 224–234 (2020).
13. Ottema, S. *et al.* The leukemic oncogene EVI1 hijacks a MYC super-enhancer by CTCF-facilitated loops. *Nat. Commun.* **12**, 5679 (2021).
14. Montefiori, L. E. *et al.* Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell Leukemia. *Cancer Discov.* **11**, 2846–2867 (2021).
15. Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **24**, 103–112 (2018).
16. Romana, S. P. *et al.* The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* **85**, 3662–3670 (1995).
17. Schwaller, J. Novel insights into the role of aberrantly expressed MNX1 (HLXB9) in infant acute myeloid leukemia. *Haematologica* **104**, 1–3 (2019).
18. Balgobind, B. V. *et al.* Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. *Haematologica* **96**, 221–230 (2011).
19. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
20. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
21. Cruz-Molina, S. *et al.* PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell* **20**, 689–705.e9 (2017).
22. Ragusa, D., Dijkhuis, L., Pina, C. & Tosi, S. Mechanisms associated with t(7;12) acute myeloid leukaemia: from genetics to potential treatment targets. *Biosci. Rep.* **43**, (2023).
23. Dalla-Favera, R. *et al.* Mechanism of activation and biological role of the c-myc oncogene in B-cell lymphomagenesis. *Ann. N. Y. Acad. Sci.* **511**, 207–218 (1987).
24. Duan, H., Xiang, H., Ma, L. & Boxer, L. M. Functional long-range interactions of the IgH 3' enhancers with the bcl-2 promoter region in t(14;18) lymphoma cells. *Oncogene* **27**,

6720–6728 (2008).
25. Ryan, R. J. H. *et al.* Detection of Enhancer-Associated Rearrangements Reveals Mechanisms of Oncogene Dysregulation in B-cell Lymphoma. *Cancer Discov.* **5**, 1058–1071 (2015).
26. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
27. Zhang, Y. *et al.* High-coverage whole-genome analysis of 1220 cancers reveals hundreds of genes deregulated by rearrangement-mediated cis-regulatory alterations. *Nat. Commun.* **11**, 736 (2020).
28. Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* **167**, 1188–1200 (2016).
29. Qiu, Y. & Huang, S. CTCF-mediated genome organization and leukemogenesis. *Leukemia* **34**, 2295–2304 (2020).
30. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature* **611**, 387–398 (2022).
31. Cai, Y. *et al.* H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat. Commun.* **12**, 719 (2021).
32. Nagel, S., Kaufmann, M., Scherr, M., Drexler, H. G. & MacLeod, R. A. F. Activation of HLXB9 by juxtaposition with MYB via formation of t(6;7)(q23;q36) in an AML-M4 cell line (GDM-1). *Genes Chromosomes Cancer* **42**, 170–178 (2005).
33. Ingenhag, D. *et al.* The homeobox transcription factor HB9 induces senescence and blocks differentiation in hematopoietic stem and progenitor cells. *Haematologica* **104**, 35–46 (2019).
34. Wildenhain, S. *et al.* Homeobox protein HB9 binds to the prostaglandin E receptor 2 promoter and inhibits intracellular cAMP mobilization in leukemic cells. *J. Biol. Chem.* **287**, 40703–40712 (2012).
35. Waraky, A. *et al.* Aberrant MNX1 expression associated with t(7;12)(q36;p13) pediatric acute myeloid leukemia induces the disease through altering histone methylation. *Haematologica* (2023) doi:10.3324/haematol.2022.282255.
36. Thompson, N., Gésina, E., Scheinert, P., Bucher, P. & Grapin-Botton, A. RNA profiling and chromatin immunoprecipitation-sequencing reveal that PTF1a stabilizes pancreas progenitor identity via the control of MNX1/HLXB9 and a network of other transcription factors. *Mol. Cell. Biol.* **32**, 1189–1199 (2012).
37. Kodgule, R. *et al.* ETV6 Deficiency Unlocks ERG-Dependent Microsatellite Enhancers to Drive Aberrant Gene Activation in B-Lymphoblastic Leukemia. *Blood Cancer Discov* **4**, 34–53 (2023).
38. Cameron, D. L. *et al.* GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol.* **22**, 202 (2021).
39. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
40. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
41. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
42. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–

140 (2010).
43. van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
44. Weichenhan, D. *et al.* Tagmentation-Based Library Preparation for Low DNA Input Whole Genome Bisulfite Sequencing. *Methods Mol. Biol.* **1708**, 105–122 (2018).
45. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
46. Carter, B. *et al.* Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat. Commun.* **10**, 3747 (2019).
47. Orsi, G. A., Kasinathan, S., Zentner, G. E., Henikoff, S. & Ahmad, K. Mapping regulatory factors by immunoprecipitation from native chromatin. *Curr. Protoc. Mol. Biol.* **110**, 21.31.1–21.31.25 (2015).
48. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
49. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17–32 (2020).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
52. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).
53. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
54. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
55. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
56. Brandstetter, K. *et al.* Differences in nanoscale organization of regulatory active and inactive human chromatin. *Biophys. J.* **121**, 977–990 (2022).
57. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).
58. Wurm, C. A., Neumann, D., Schmidt, R., Egner, A. & Jakobs, S. Sample preparation for STED microscopy. *Methods Mol. Biol.* **591**, 185–199 (2010).
59. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
60. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).