Nucleic Acids Research, 2024, **52**, 12262–12280
https://doi.org/10.1093/nar/gkae864
Advance access publication date: 8 October 2024
Gene regulation, Chromatin and Epigenetics

Integrated multi-omics analysis of PBX1 in mouse adult neural stem- and progenitor cells identifies a transcriptional module that functionally links PBX1 to TCF3/4

Vera Laub[1], Elisabeth Nan[1], Lena Elias[1], Ian J. Donaldson[2], Mette Bentsen[3], Leona A. Rusling[4], Jonathan Schupp[1,5], Jennifer H. Lun[1,5], Karl H. Plate[1,5], Mario Looso[3], Julian D. Langer[4], Stefan Günther[6], Nicoletta Bobola[7] and Dorothea Schulte[1,*]

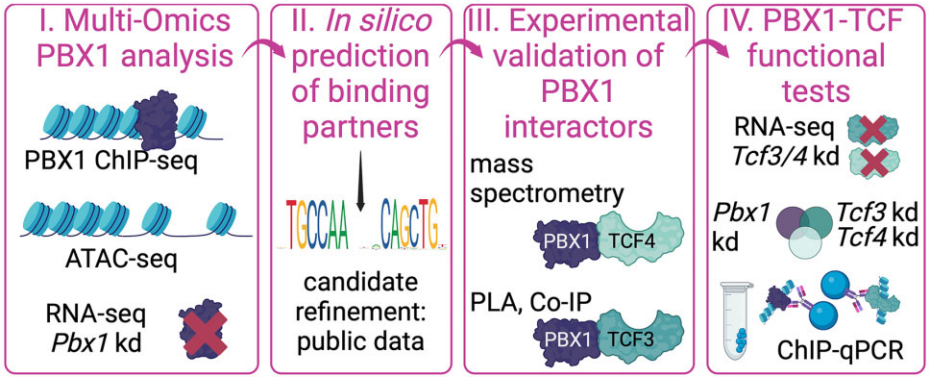
[1]Goethe University, University Hospital Frankfurt, Neurological Institute (Edinger Institute), 60528 Frankfurt am Main, Germany
[2]University of Manchester, Faculty of Biology, Medicine and Health, Bioinformatics Core Facility, Manchester, M13 9PT, UK
[3]Max Planck Institute for Heart and Lung Research, Bioinformatics Core Unit (BCU), 61231 Bad Nauheim, Germany
[4]Max Planck Institute for Biophysics, Proteomics, and Max Planck Institute for Brain Research, 60438 Frankfurt am Main, Germany
[5]Goethe University, Frankfurt Cancer Institute, 60528 Frankfurt am Main, Germany
[6]Max Planck Institute for Heart and Lung Research, Bioinformatics and Deep Sequencing Platform, 61231 Bad Nauheim, Germany
[7]University of Manchester, Faculty of Biology, Medicine and Health, Manchester, M13 9PT, UK

[*]To whom correspondence should be addressed. Tel: +49 6301 84159; Fax: +49 69 6301 84150; Email: d.schulte@em.uni-frankfurt.de
Present addresses:
Vera Laub, University of California, San Francisco, Program in Craniofacial Biology, Institute for Human Genetics, Eli and Edythe Broad Center of Regeneration Medicine & Stem Cell Research, San Francisco, CA 94143, USA.
Leona A. Rusling, Goethe University, Institute of Pharmaceutical Chemistry, 60438 Frankfurt, Germany.

Abstract

Developmental transcription factors act in networks, but how these networks achieve cell- and tissue specificity is still poorly understood. Here, we explored pre-B cell leukemia homeobox 1 (PBX1) in adult neurogenesis combining genomic, transcriptomic, and proteomic approaches. ChIP-seq analysis uncovered PBX1 binding to numerous genomic sites. Integration of PBX1 ChIP-seq with ATAC-seq data predicted interaction partners, which were subsequently validated by mass spectrometry. Whole transcriptome spatial RNA analysis revealed shared expression dynamics of *Pbx1* and interacting factors. Among these were class I bHLH proteins TCF3 and TCF4. RNA-seq following *Pbx1*, *Tcf3* or *Tcf4* knockdown identified proliferation- and differentiation associated genes as shared targets, while sphere formation assays following knockdown argued for functional cooperativity of PBX1 and TCF3 in progenitor cell proliferation. Notably, while physiological PBX1-TCF interaction has not yet been described, chromosomal translocation resulting in genomic *TCF3::PBX1* fusion characterizes a subtype of acute lymphoblastic leukemia. Introducing *Pbx1* into Nalm6 cells, a pre-B cell line expressing *TCF3* but lacking *PBX1*, upregulated the leukemogenic genes *BLK* and *NOTCH3*, arguing that functional PBX1-TCF cooperativity likely extends to hematopoiesis. Our study hence uncovers a transcriptional module orchestrating the balance between progenitor cell proliferation and differentiation in adult neurogenesis with potential implications for leukemia etiology.

Graphical abstract



Received: December 22, 2023. **Revised:** August 22, 2024. **Editorial Decision:** September 16, 2024. **Accepted:** September 23, 2024

Introduction

Metazoan development requires the orchestrated expression of thousands of genes. A key position in this process is occupied by transcription factors (TFs), DNA-binding proteins that decode regulatory sequences in the genome and translate this information into the temporally and spatially controlled production of transcripts. The human genome contains an estimated 1600 genes coding for TFs (1), a surprisingly low number, considering the diversity of biological processes that need to be started, executed, and completed in a coordinated manner during embryogenesis. Except for few specialized 'master TFs' that can activate cell identity pathways more or less independently, the majority of developmental TFs must act in concert, forming multimeric transcriptional complexes and interdependent gene regulatory networks (GRNs). Cooperativity among TFs together with the interconnectivity of GRNs ensures that gene expression is robust and cell type-specific but also adaptive to environmental signals, allowing existing transcriptional programs to be continuously adjusted to the changing cellular conditions that accompany developmental progression. How cell type-specific multi-TF complexes evolve, or how they and the GRNs they engage in are controlled in a spatially, temporally, and quantitatively dynamic manner are complex and still largely unanswered questions. Given that individual TFs must participate in more than one cell fate-specific transcriptional complex, a lasting question is how diverse such complexes must be for the same TF to contribute to the regulation of different developmental processes.

The three amino-acid loop extension (TALE) homeodomain (HD) transcription factor pre-B cell leukemia (PBX1) is a developmental TF involved in the control of pleiotropic biological processes (2). Initially discovered as part of the fusion protein that results from t(1;19) chromosomal translocation in human pre-B cell acute lymphoblastic leukemia (ALL) (3,4), *Pbx1* is expressed in a multitude of cell types and tissues (5). Loss-of-function models established *Pbx1* as a critical developmental regulator of many tissues and organs, including axial skeleton and bones of the limbs and girdle (6–8), muscles (9), heart (10), cranium and facial structures (6,11,12), hematopoiesis (13), as well as patterning of the cerebral cortex (14) and hindbrain (15). *Pbx1* is also critically involved in the generation of new neurons from resident stem cells in the adult mammalian ventricular- subventricular zone (V-SVZ), the largest neural stem cell niche in the adult rodent forebrain. In the V-SVZ, young neurons, termed neuroblasts, are generated from adult neural stem cells (adult NSC) via an intermediate population of transient amplifying progenitor cells (TAPs). Neuroblasts migrate en route the rostral migratory stream (RMS) towards the olfactory bulb (OB), where they terminally differentiate to GABAergic interneurons and integrate into existing circuitries (16). Consistent with the broad expression of *Pbx1* in TAPs, neuroblasts, and the majority of adult generated OB neurons, conditional deletion of *Pbx1* at different stages of adult V-SVZ neurogenesis established its involvement in multiple consecutive steps of neurogenesis (17,18).

Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) for PBX1 in different embryonic tissues mapped PBX1 binding to thousands of sites genome wide, indicating that it participates in the regulation of numerous gene expression programs (14,19,20). PBX1 achieves these tasks through interaction with other TFs, most notably HOX proteins and TALE-HD proteins of the related MEIS and PKNOX/PREP families (21–23). PAX6 is a presumed PBX1 interactor in adult V-SVZ neurogenesis (24), and PBX1 has been implicated in the interaction network of mental disorder proteins with OLIG2 (25). Still, PBX1 chromatin binding in cells of the V-SVZ adult neurogenic system has not yet been studied genome-wide and only few proteins that cooperate with PBX1 in this system are known.

With this study, we set out to decode PBX1-centered GRNs in adult neural progenitor cells. Combining ChIP-seq for PBX1 with ATAC-seq allowed us to predict candidate PBX1-interacting TFs, which were subsequently substantiated by data integration with a large number of ChIP-seq datasets taken from the public domain. Interactor candidates were validated by mass spectrometry and co-immunoprecipitation. This approach revealed a previously unknown physical interaction of PBX1 with the class I bHLH proteins TCF3 and TCF4 that provides a likely explanation for the strong leukemogenic activity of the t(1;19) translocation fusion protein in pre-B cell ALL.

Materials and methods

Preparation of murine primary V-SVZ cells and aNS culture

All procedures involving animals were approved by the local animal care committee and the government of Hesse and are in accordance with German and EU regulations. 6–30 weeks-old C57Bl/6J mice were provided by the MPI for Brain Research Frankfurt and the Central Animal Facility of the University Hospital Frankfurt. Adult neural stem-/progenitor cells were isolated from the lateral walls of the lateral ventricle and propagated at 37°C, 5% CO_2 in the presence of EGF and FGF2 to allow the formation of free-floating neurospheres (17). Adult neurospheres (aNS) were passaged every three to four days and passage 1 (p1) or passage 2 (p2) cells were used for all experiments.

siRNA mediated knockdown

For transient knockdown (kd), siRNAs targeting the coding sequences of *Pbx1* (Ambion, s71280), *Tcf3* (Ambion, s74858) or *Tcf4* (Ambion, s74830) were transfected into 1 Mio p1 aNS cells per treatment with Metafectene (Biontex) according to the manufacturer's instructions. Validated non-targeting siRNAs were used as controls in all experiments (Ambion, AM4611). After 48 h incubation at 37°C, cells were collected by centrifugation and used for qPCR, western blot or RNA-seq analysis. Kd efficiency of >70% was confirmed by qPCR.

Spheroid assay

To assess sphere growth under conditions of reduced *Pbx1 or Tcf3* expression, aNS cells were treated as described above with individual siRNAs alongside combination treatment. Per treatment, 25 000 cells were collected in the center of 96-well round bottom culture plates by brief centrifugation. Spheroid formation was observed by imaging for 72 h in a Brightfield CYTENA live-cell imaging system with CELLCYTE Studio Version 2.7.4 for analysis. Sphere size minimum was reached at 34 h post-treatment and was used for normalization.

Statistical analysis of control versus treatment was performed with unpaired, two-sided Student's t-test.

RT-qPCR

RNA was isolated with the RNeasy Mini Kit (Qiagen), cDNA was generated using the cDNA Synthesis Kit (Thermo Fisher Scientific). RT-qPCR of 1:100 diluted cDNA was performed using specific primers (Supplementary Table S1). Transcript quantity was assessed relative to *β-actin* or *GAPDH* for murine V-SVZ derived aNS and *RPLP0* for Nalm6 cells, as indicated. Statistical analysis of treatment vs. control was performed with paired student's t-test. Correlation analysis was performed with the Pearson correlation test in R Studio Version 1.3.1093.

Co-immunoprecipitation (co-IP)

To assess interaction between TCF3 and PBX1, *p3xFlag-Tcf3/E47* (murine *Tcf3* splice variant *E47*; Addgene #34585) or *p3xFlag-Tcf3/E47aa290-648* (coding for amino acids 290–648 of mouse E47) were introduced into HEK293T cells by calcium phosphate transfection. Cells were collected after incubation for 48 h at 37°C and lysed in cold JS+ buffer (50 mM HEPES at pH 7.5, 150 mM NaCl, 5 mM EGTA, 1.5 mM $MgCl_2$, 1% glycerol, 1% Triton X-100, in sterile H_2O; + indicates addition of cOmplete® protease inhibitor cocktail (PI), Roche) at 4°C for 30 min while rotating (26). The solution was sonicated for 5 cycles in a Bioruptor Plus (Diagenode) with high intensity, 30 s on, 30 s off, 4°C. 10% of the lysate were kept as input material. Immunoprecipitation was performed over night at 4°C, rotating with the remaining protein lysate and 2 µg of antibodies (PBX1 rb, Cell Signaling Technology, #4342 or GFP rb, abcam, ab290). Immunocomplexes were bound to Dynabeads protein A (30 µl/sample) by incubation for 3 h at 4°C, rotating. Beads were washed 3× with 1 ml NET + buffer (50 mM Tris–HCl at pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 1× PI in H_2O), centrifuged, the pellet was resuspended in LDS sample buffer (1×) along with 100 mM DTT (final conc.) and denatured at 70°C. The eluate was analyzed by western blotting using antibodies directed against the FLAG epitope (ms, 1:1000, Sigma, 1804).

Mass spectrometry (MS)

Proteomic assessment of the PBX1 interactome was performed in triplicates with antibodies specific for PBX1 or GFP together with aNS cells dissociated at day three of p1. Lysis, sonication and immunoprecipitation were performed as described above using 5 µg of either PBX1 rb (Cell Signaling Technology, #4342) or GFP rb (abcam, ab290) antibodies, but omitting the denaturation step in LDS sample buffer. The protein precipitates were instead processed using the S-Trap Micro High Recovery Protocol (ProtiFi, LLC) after the final washes. MS measurements and data analysis were performed at Max Planck-Institute for Biophysics, Frankfurt, Germany. Analytical separation was performed using a nanoElute nanoHPLC (Bruker Daltonics, Bremen, Germany) and liquid chromatography was coupled to a hybrid trapped ion mobility spectrometer (TIMS) – quadrupole time of flight (Q-ToF) mass spectrometer, timsTOF ProII (Bruker Daltonics, Bremen, Germany), with CaptiveSpray nano-electrospray ion source (nano-ESI). Data were acquired in positive, data dependent acquisition utilizing Parallel Accumulation Serial Fragmentation (dda-PASEF) and TIMS mode. For protein identifications and relative quantification of the MS/MS data acquired, MaxQuant version 2.2.0.0 was utilized (27,28). Details are described in Supplementary Material online. For visualization, mass spectrometry derived PBX1 interactors were filtered manually to exclude RNA-binding/ribosomal proteins and splicing factors, and subjected to protein-protein interaction network analysis with STRING 11.5 (29).

Proximity ligation assay (PLA)

For PLA, aNS cells in passage 1 were dissociated and transduced with *pWPI-Tcf3/E47-Flag* or *pWPI-empty vector* control virus (details described in Supplementary Material online) for 48 h. Per treatment condition, 150 000 singularized cells were plated on a PDL-coated 8-well Lab-Tek chamber slide. Attached cells were fixed, washed and permeabilized. After blocking, PLA was performed with Duolink in Situ Detection Reagents Red (Sigma-Aldrich) following the manufacturer's instructions. PBX1 rb (Cell Signaling Technology, #4342, 1:15 000, ChIP-grade high-concentrated formulation) or FLAG ms (Sigma-Aldrich, F3165, 1:1000) were used as primary, PLA-plus rb (Sigma-Aldrich, DUO82002-100RXN, 1:10) and PLA-minus ms (Sigma-Aldrich, DUO82004-100RXN, 1:10) antibodies as secondary antibodies. After ligation and amplification, cells were blocked and stained with GFP chk (abcam, ab13970, 1:1000), Alexa-488 α chk (Life Technologies, A11039, 1:1000) to visualize viral transduction and counterstained with DAPI. Specimen were mounted with cover slips in Aqua-Poly/Mount (Polysciences) and allowed to dry overnight. Images were captured using a Nikon Eclipse 80i microscope (60× water objective). Analysis was performed in ImageJ 1.53t with a customized analysis script (available in Supplementary Material online).

GST-pulldown

Pulldown experiments were performed with PBX1a or PBX1b and truncated forms thereof C-terminally fused to triple HA-tag, and full-length and truncated forms of mE47 N-terminally fused to GST-Flag, co-transfected into HEK293T. Transfection and extract preparation by sonification were performed as described above. Details of the pulldown are described in Supplementary Material online.

ChIP-qPCR

ChIP was performed on primary V-SVZ-derived aNS, cross-linked with 1% PFA, and fragmented by sonication to a size of 200–700 bp. Immunoprecipitation was performed with magnetic Dynabeads protein A/G and 2 µg of either α-Pbx1 rb (Cell Signaling Technology, #4342), α-FLAG ms (Sigma-Aldrich, F3165), or α-H3K27ac rb (Diagenode, C15410196) antibodies. Genomic primers for ChIP-qPCR can be found in Supplementary Table S2. Statistical analysis was done by two-way ANOVA with Bonferroni post-test. Details are described in Supplementary Material online.

ChIP-seq

PBX1 and H3K27ac ChIP-seq was performed in duplicates as described above, with 25 µg chromatin of fragment size 100–500 bp per IP. ChIP-enriched chromatin was purified on MicroChIP DiaPure columns (Diagenode) and successful ChIP was verified by qPCR. Library preparation was conducted with MicroPlex Library Preparation kit v2 (Diagen-

ode). Paired-end sequencing was performed at the Genomic Technologies Core Facility, University of Manchester, UK on an HiSeq 4000 or NovaSeq 6000 sequencer (both Illumina). Bioinformatic analysis of the raw data was carried out at the Bioinformatics Core Facility of University of Manchester, UK. Sequencing output in BCL format was converted to FASTQ using bcl2fastq v2.20.0.422, unmapped read pairs were tested by FastQC v11.3 and fastq-screen v0.9.2 or v0.14.0 to identify any contaminating DNA. Remaining sequence adapters were removed and reads were quality trimmed using Trimmomatic v0.36 (SLIDINGWINDOW:4:20 MINLEN:35) (30). Reads were mapped against the mouse reference genome, version mm10/GRCm38 using Bowtie2 v2.3.0 (31). Mapped reads were filtered using samtools v1.9 to retain only high confidence concordant pairs (-f 2 – q30) (32), followed by removal of reads mapping to unassembled contigs and mitochondrial genome. Peak calling was performed using MACS2 v2.1.2 using the default qvalue threshold of 0.05 (33); candidate regions were filtered by fold enrichment (FE) score. *De novo* PBX1 motif analysis was performed using MEME-ChIP v4.11.2 on the top 1000 ChIP-seq peaks of each replicate. The background sequences were obtained from randomly shuffled locations of PBX1 peaks.

ATAC-seq

Adult V-SVZ stem- and progenitor cells cultured for 48 h as free-floating aNS were purified by 3% BSA-gradient, and treated with accutase for singularization. For each of two replicates, cells (10 000/12 000, respectively) were subjected to cell lysis (Tris–HCl 10 mM, NaCl 2 mM, $MgCl_2$ 3 mM, Igepal 0.1%) and transposase reaction mix (Nextera) treatment with occasional snap mixing. DNA fragments were purified by MinElute PCR Purification Kit (Qiagen). Amplification of library together with indexing primers was performed as described elsewhere (34). Libraries were mixed in equimolar ratios and sequenced on the NextSeq500 platform using V2 chemistry and paired-end setup.

Raw reads were trimmed and aligned to the mouse genome version mm10 (GRCm38) using STAR 2.5.4b with the parameters '–outFilterMismatchNoverLmax 0.1 –outFilterMatchNmin 20 –alignIntronMax 1 –alignSJDBoverhangMin 999 –outFilterMultimapNmax 1 –alignEndsProtrude 10 ConcordantPair' (35) and retaining only unique alignments to exclude reads of uncertain origin. Reads were further deduplicated using Picard 2.21.1 to mitigate PCR artefacts leading to multiple copies of the same original fragment. Reads aligning to the mitochondrial chromosome were removed. The MACS peak caller version 3.0.0a6 was employed to accommodate for the range of peak widths typically expected for ATAC-seq (33). Minimum qvalue was set to −4 and FDR was changed to 0.0001. Peaks overlapping ENCODE blacklisted regions (known misassemblies, satellite repeats) were excluded. In order to compare peaks in different samples to assess reproducibility, the resulting lists of significant peaks were overlapped and unified to represent identical regions. Sample counts for union peaks were produced using bigWigAverageOverBed (UCSC Toolkit) and normalized with DESeq2 1.30.1 to compensate for differences in sequencing depth, library composition, and ATAC-seq efficiency (36). Peaks were annotated with the promotor of the nearest gene in range (TSS ± 5000 nt) based on reference data of GENCODE vM15.

Motif enrichment within ATAC-seq peaks was performed with AME (Analysis of Motif Enrichment) v5.0.2 using motifs from JASPAR and a background of mESC peaks (GEO source: GSM2156965) (37). ATAC-seq footprinting was performed within open chromatin regions using TOBIAS v0.16.0 with standard parameters (38). Footprints are visualized as the aggregate Tn5-corrected signal across TF binding sites predicted bound.

RNA-seq

RNA-seq of wild-type (WT) V-SVZ derived aNS, or of aNS 48 h post transfection with siRNAs directed against *Pbx1*, *Tcf3*, *Tcf4* or control siRNAs, was performed in triplicates. Total RNA was isolated by RNeasy Mini Kit (Qiagen). For exclusion of genomic DNA contamination, the samples were treated by on-column DNase digestion (RNase-free DNase Set, Qiagen). Total RNA and library integrity were verified with LabChip Gx Touch 24 (Perkin Elmer). 1 µg of total RNA was used as input for SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian (Clontech). Sequencing was performed on a NextSeq 500 instrument (Illumina) with 1 × 75 bp single end setup or NextSeq 2000 instrument with 1 × 72 bp single end setup. Details are described in Supplementary Material online.

Whole transcriptome spatial RNA analysis

The formalin-fixed paraffin embedded brain of C57Bl/6J ($n = 1$) was cut in coronal 5 µm sections and placed on Epredia™ SuperFrost Plus™ microscopy slides (Epredia). The GeoMx® Mouse Whole Transcriptome Atlas Assay (NanoString Technologies) was used as recommended by the manufacturer's protocol. Immunofluorescence staining was performed with α-PBX1 (rb, Cell Signaling Techology, #4342, 1:400) and α-DCX (gp, Merck, #AB2253, 1:4000) as primary, and Alexa488 α-rb (Thermo Fisher Scientific, A11008, 1:1000), Alexa647 α-gp (Thermo Fisher Scientific, AA21450, 1:1000) as secondary antibodies. A 2.5 nM solution of SYTO™ 82 Orange Fluorescent Nucleic Acid Stain (Thermo Fisher Scientific) was used to label the DNA. Data acquisition was performed on a GeoMx® Digital Spatial Profiler (NanoString). The barcodes of the assay were sequenced on a NextSeq 1000 (Illumina) and demultiplexed on the Illumina BaseSpace Sequence Hub by the NanoString GeoMx NGS Pipeline 2.0.21. Downstream data analysis was performed using the manufacturers software packages on the GeoMx® Digital Spatial Profiler.

Bioinformatic post-processing

Data retrieval from the public domain, bioinformatic post-processing and data integration of ChIP-/ ATAC- and RNA-seq was performed as summarized previously (39). Specifications of the applied processes and tools are described below. Graphs were produced with R Studio Version 1.3.1093 unless specified otherwise.

ChIP-seq analysis

GO term analysis for genomic loci occupied with PBX1 in .bed was performed using GREAT version 4.0.4 (40). Additional ChIP-seq data was retrieved from GEO (41), ChIP-Atlas (42), or the sources listed in Supplementary Table S3. For ChIP-seq visualization, mm10 .bigWig and .bed representations were loaded into IGV, version 2.11.9 (43). Multiple profiles were

co-visualized for inspection of local track overlaps and peak shape. In case .bed data was available in other formats than mm10 (also including human datasets), UCSC liftOver was used for coordinate conversion (44).

Query sequences in .bed format were converted to .fasta using bedtools getfasta (version 2.30.0) (45) on the Galaxy platform (46). Motif enrichment of PBX1 ChIP-seq peaks (.fasta) was performed with MEME-ChIP, using 'classic' motif discovery and enrichment mode for DNA, RNA or protein alongside Eukaryote DNA/vertebrates (*in vivo* and *in silico*) as motif enrichment paradigm. Significantly enriched selected motifs and motif profiles were retrieved in .eps format. Peak centering analysis was performed with CentriMo (version 5.5.1) using the 200 nucleotide (nt) summit of PBX1 ChIP-seq peaks (47). Secondary motif analysis was performed with SpaMo using entire overlapping region of PBX1 (aNS, this manuscript) and TCF3 (NPC) (48) ChIP-seq peaks, as a minimum peak length of 310 nt is required for this analysis (49). Peak annotation was performed with ChIPseeker (50), making use of mm10 comprehensive gene annotation file GRCm38.p6 (retrieved from GENCODE) as annotation source.

ChIP-seq datasets of potential PBX1 interactors (determined by aNS data integration of PBX1 ChIP-seq motifs, ATAC-seq motif enrichment and RNA-seq, for details see below) were subjected to *in silico* intersection to assess the context-specificity of PBX1-interactor overlaps. For each potential interactor, ChIP-seq data of developmental neurogenic (NSC or NPC) or adult neuronal contexts, as well as developmental and terminally differentiated control contexts (ESC, liver, lung etc., dependent upon availability) were intersected with full-length PBX1 ChIP-seq peaks from aNS cells (n1). Intersections were computed using bedtools intersect (option -wo to later remove duplicate intersections mapped to one reference peak) version v2.27.1 (45). Because coverage of the genome differed considerably between the datasets used for overlap computation, control intersections were conducted for each dataset. To this end, peaks were shifted in a random manner using a python script (see Supplementary Material online), while retaining their original extension. Background overlaps of 10 iterations of this script were computed, average overlap with PBX1 ChIP-seq was reported and visualized.

RNA-seq analysis

For visualization of RNA-seq data (*Pbx1*, *Tcf3* or *Tcf4* kd alongside respective controls), information was retrieved from the spreadsheet matrix and manually converted into CLARION format. Data was uploaded to the Webbased Interactive Omics visualizatioN tool (WIlsON), filtered as indicated in the respective figures, and visualized as heatmap with Euclidean clustering based on average (51).

GO term enrichment was performed with PANTHER Overrepresentation Test (version 20221013) with Fisher's Exact test, calculating false discovery rate (FDR) as correction (52). Redundancy of GO terms was removed with REVIGO version 1.8. (53). Gene regulatory network analysis of genes dysregulated upon *Pbx1* kd (ENCODE and ChEA Consensus TFs from ChIP-X) was performed with Enrichr (54,55).

Data integration of different sequencing paradigms

ChIP-seq and RNA-seq

For gene regulatory network prediction of PBX1 ChIP-seq, data were subjected to ISMARA analysis (56), and relevant predicted network interactors were filtered by integration with top predicted GRN candidates of dysregulated genes upon *Pbx1* kd in RNA-seq derived from Enrichr analysis (see above).

ChIP-/ATAC-/RNA-seq

Potential PBX1 interactors were identified by triangulation of PBX1 ChIP-seq motifs with TF motifs detected in open chromatin (ATAC-seq), and information about expression of the interactor candidate (RNA-seq). JASPAR 2022 CORE vertebrate non-redundant pfms motifs identified in both PBX1 ChIP-seq (MEME-ChIP) and ATAC-seq (AME), were integrated using a python script (see Supplementary Material online), and information of interactor candidate expression in RNA-seq (>50 reads in RNA-seq of WT aNS cells) was derived from corresponding spreadsheet matrix.

Identification of *PBX1* targets in leukemic context

Because PBX1 displays strong binding and regulation at promotors, E2A-PBX1 ChIP-seq peaks from leukemic 697 cells with t(1;19) fusion were annotated with ChIPseeker and filtered for promotor peaks. Information on dysregulation in shRNA-mediated *PBX1* kd in 697 cells was integrated (57). Promotor peaks were subjected to CiiiDER (58). Expression in different cell lines was assessed on DepMap and targets with increased expression in cell lines with t(1;19) translocation were selected. Filtered targets were screened for known possible significance in a leukemic context based on literature.

Results

Triangulation of PBX1 functionality with ChIP-seq, ATAC-seq and RNA-seq

Adult neural stem- and progenitor cells can be captured *in vitro* and cultivated in the presence of epidermal growth factor (EGF) and fibroblast growth factor 2 (FGF2) as adult neurospheres (aNS), free-floating cell aggregates that consist predominantly of TAPs and a few activated stem cells (59). Although aNS progressively change from neurogenic to gliogenic in the presence of growth factors, early passage aNS constitute an experimentally amenable population of neural stem- and progenitor cells (60,61). ChIP-seq analysis with PBX1-specific antibodies (Figure 1A, Supplementary Figure S1A) on chromatin of second passage adult murine aNS identified a large number of PBX1-bound sites, which could be assigned to a total of 16 044 genes (Supplementary Table S4). *De novo* motif analysis identified a short consensus sequence with maximum mapping to the center of PBX1-bound peaks, representing the PBX1 core binding motif in adult neural progenitor cells (Figure 1B). This motif shares a high degree of similarity with the short PBX1 consensus motif deduced from ChIP-seq data obtained for PBX1 in E12.5 and E15.5 embryonic mouse cortex (14) as well as the decameric motif bound by PBX-PREP in 3.5hpf zebrafish embryos (62) and E11.5 mouse embryonic trunk (19), but differs markedly from the motifs identified for PBX1-HOX targets (19). Peaks with fold enrichment (FE) of >50 ($n = 2453$) were primarily annotated to promotor-proximal regions (≤1 kb from putative transcriptional start site) (Figure 1C) and associated with terms such as 'protein folding', 'regulation of cell cycle G2/M phase transition,' and 'positive regulation of Notch receptor targets' (Figure 1E). In contrast, peaks with lower fold enrichment

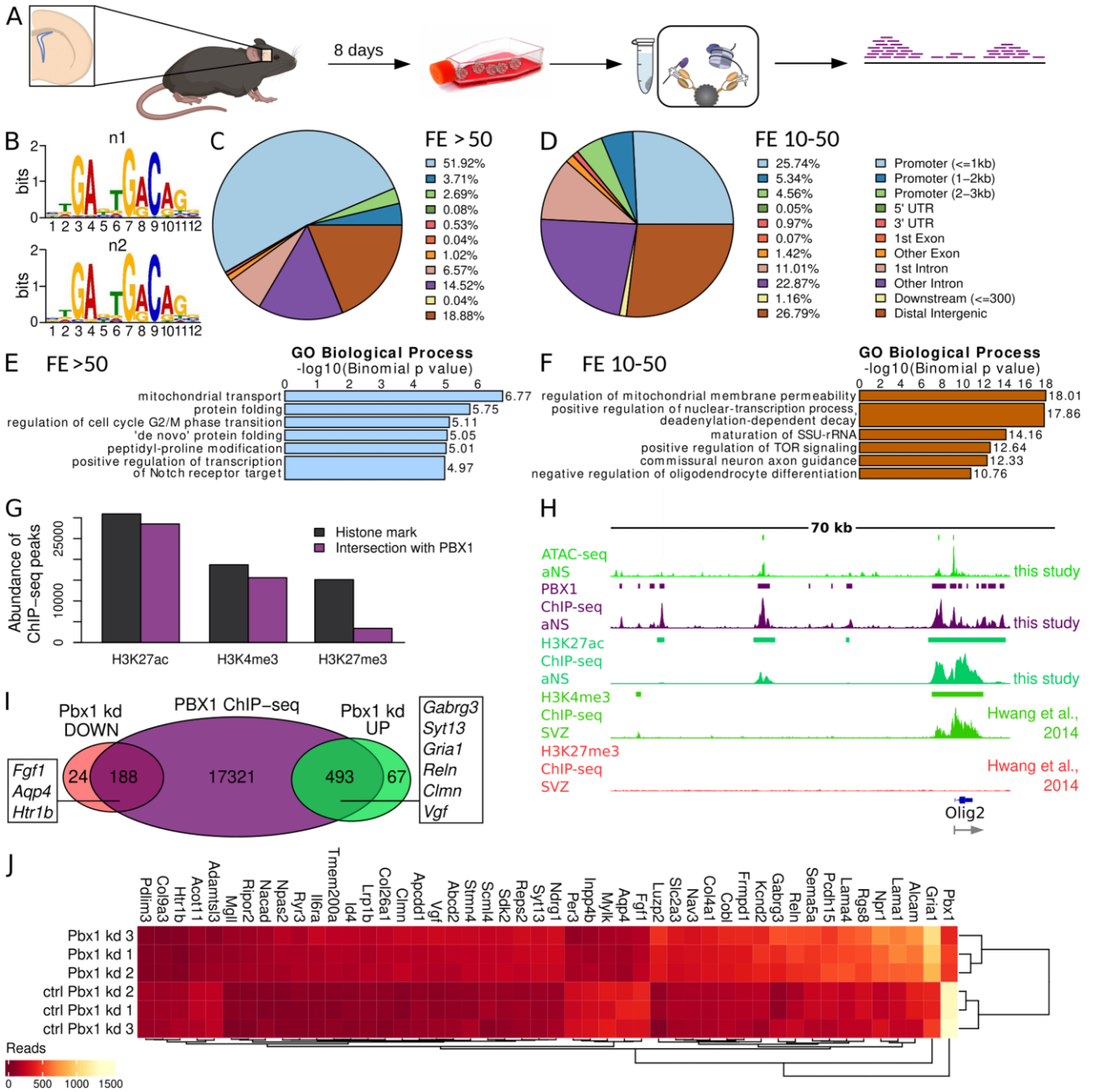


Figure 1. Genome-wide identification of PBX1 target genes in adult neural stem- /progenitor cells by ChIP-seq, RNA-seq and ATAC-seq. (**A**) Schematic depiction of PBX1 ChIP-seq experiment. (**B**) PBX1 binding motifs of both replicates (n1/n2) identified by *de-novo* motif analysis with MEME-ChIP. (**C, D**) ChIPseeker derived peak annotation of PBX1 ChIP-seq peaks with high fold enrichment (FE > 50, **C**) and low FE (10–50, **D**). (**E, F**) GO term analysis with GREAT of peaks with FE > 50 (**E**) and FE 10–50 (**F**). (**G**) Intersection of PBX1-bound genomic sites with regions marked by activating (H3K27ac, H3K4me3) or repressing (H3K27me3) histone modifications (63). (**H**) Exemplary co-localization of open chromatin (ATAC-seq) and PBX1 ChIP-seq occupancy, alongside histone modifications at the *Olig2* gene locus and upstream enhancer regions. Colored bars represent called peaks, grey arrow direction of transcription. (**I**) The majority of up- and downregulated genes upon *Pbx1* kd are associated with PBX1 ChIP-seq peaks. (**J**) Heatmap of top upregulated (FE > 0.9) and downregulated (FE < −0.9) genes following *Pbx1* kd.

(FE 10–50; n = 24 144) mapped more evenly to promotor proximal, intronic or distal intergenic regions (Figure 1D), and were linked to GO terms including 'commissural neuron axon guidance' and 'negative regulation of oligodendrocyte differentiation' (Figure 1F).

We next correlated the genomic sites of PBX1-binding to the transcriptional state of these loci by conducting ChIP-seq experiments for histone 3 lysine 27 acetylation (H3K27ac), an epigenetic modification associated with transcriptionally active promotors and enhancers (Supplementary Table S4). In the vast majority of cases, PBX1-bound genomic sites coincided with H3K27ac-modified chromatin (Figure 1G). Supporting the notion that PBX1 binds to transcriptionally active chromatin, intersecting PBX1-bound genomic sites with published ChIP-seq data from aNS uncovered prominent overlap of PBX1 binding with regions positive for histone 3 lysine 4 triple-methylation (H3K4me3), an epigenetic mark of active promotors, but minimal overlap with chromatin marked

by H3K27me3, a modification associated with transcriptional repression (Figure 1G, Supplementary Figure S1B, C) (63). To further narrow down the binding properties of PBX1 to chromatin, we applied ATAC-seq, a method grounded in the concept that protein binding shields DNA against cleavage. A total of 81 141 open chromatin regions were detected in early passage aNS of which 24 420 could be assigned to promotors (Supplementary Table S5). 18 466 (75.6%) of promoter-proximal open chromatin regions overlapped with PBX1 peaks, as exemplified for the gene locus of *Olig2* (Figure 1H). Visual inspection of PBX1 occupancy alongside histone marks and open chromatin at further example loci demonstrated binding to activated (*Cited2*), poised (*Prom1*) and repressed (*Dcx*) genes in aNS cells (Supplementary Figure S1D). In adult neural stem- and progenitor cells, PBX1 binding to the genome, hence, occurs at numerous positions, but is strongest at promotors. Further, PBX1 binding predominantly overlaps with regions of open chromatin and regulatory regions that carry histone modifications characteristic of actively transcribed chromatin.

To correlate PBX1 chromatin binding to a transcriptional outcome, we performed siRNA-mediated *Pbx1* knockdown in aNS cells (Supplementary Figure S2A-C; *Pbx1* kd: 84.31 ± 6.46% in RT-qPCR) followed by whole transcriptome RNA-seq. This approach revealed differential expression of 772 genes following PBX1-depletion (560 upregulated, 212 downregulated, Supplementary Figure S2D, Supplementary Table S6). 88% (493/560) of the upregulated and 88.6% (188/212) of the downregulated genes corresponded to PBX1-bound genes (Figure 1I). Notably, top upregulated genes encompassed neuronal marker genes, including the gamma 3 subunit of GABA-A receptor (*Gabrg3*), synaptotagmin XIII (*Syt13*) or glutamate receptor 1 (*Gria1*), as well as genes involved in nervous system development and neuronal maturation like reelin (*Reln*), calmin (*Clmn*) (64) or VGF nerve growth factor inducible (*Vgf*) (65) (Figure 1J). Corresponding GO terms were associated with neuronal development and differentiation including 'regulation of synaptic structure or activity', 'synapse organization' and 'regulation of cell differentiation' (Supplementary Figure S2E). Downregulated genes included genes linked to neural stem cell identity or maintenance of neurogenic proliferation, such as fibroblast growth factor 1 (*Fgf1*) (66), aquaporin 4 (*Aqp4*) (67), or serotonin (5-HT) receptor *Htr1b*, in line with the sensitivity of neural stem cell proliferation to 5-HT signaling (68) (Figure 1J). Accordingly, downregulated genes were enriched for GO terms related to DNA replication (Supplementary Figure S2F). These findings suggest that PBX1 may help to coordinate two competing gene expression programs in neural progenitor cells: maintenance of cellular proliferation and neurogenic differentiation, respectively.

In silico prediction of PBX1 binding partners

We subjected the PBX1 ChIP-seq dataset to MEME-ChIP motif analysis, reasoning that the TF consensus motifs near PBX1-occupied sites may provide clues to the identity of proteins that bind chromatin together with PBX1. Besides consensus binding motifs for PBX1, the most highly overrepresented motifs included MEIS2, OLIG2, NFIA and SOX2 (Figure 2A). To explore the presence of PBX1 binding motifs within accessible chromatin in aNS cells, we conducted an in-depth motif enrichment analysis of our ATAC-seq data with the analysis tool AME (Supplementary Table S5). This approach generated two separate lists of TF consensus binding motifs, those that were overrepresented near PBX1-bound sites according to our ChIP-seq data and those that were enriched in accessible chromatin as derived from analysis of our ATAC-seq results. Motifs included in both lists spanned diverse TF families, encompassing bHLH class I E-box binding TCF transcription factors, bHLH class II OLIG2, SMAD/NF-1 DNA-binding domain factors, and high-mobility group SOX2 (Figure 2B). MEME-ChIP enriched motifs were filtered for motifs that are listed in the JASPAR database and both lists were seamlessly integrated in subsequent computational analyses. This revealed that 187 motifs were shared between motifs associated with genomic regions characterized as 'open chromatin' in aNS (455 motifs in total) and motifs identified in PBX1-bound chromatin (354 motifs; Figure 2C).

To refine this list for further experimental analyses, we first filtered the 187 TFs for expression in aNS, only taking into account genes whose transcript level exceeded a minimum cutoff of 50 reads according to our WT RNA-seq results (Supplementary Figure S3A). Ninety-eight TF motifs fulfilled these criteria. To test the strength of our predictions, we first performed ATAC-seq footprinting analysis, a method which is based on the assumption that bound TFs hinder cleavage of DNA, thus enabling investigation of TF binding *in silico* (38). Indeed, footprints for many of these 98 factors were present at peak centers within accessible chromatin in aNS, including consensus motifs for OLIG2, NFIA/NFIB/NFIC/NFIX, TCF12, and TCF3-heterodimers (Supplementary Figure S4A, B; Supplementary Table S5). Next, we used the wealth of ChIP-seq data available in the public domain to examine genome-wide binding of TFs whose recognition motifs mapped near PBX1-occupied sites. We hypothesized that TFs that act in concert with PBX1 in aNS may also do so in other biological contexts, and that this may be more likely in settings related to neurogenesis than in others. We devised an intersection paradigm with bedtools and performed pairwise computational comparisons between the PBX1 ChIP-seq data from aNS and ChIP-seq datasets from various other cell sources. A total of 46 such comparisons were performed with datasets categorized into neural (early neural: 11 datasets; late neural: 7 datasets) or non-neural biological contexts (non-neural development: 9 datasets, hematopoietic lineage: 10 datasets, adult organ: 9 datasets; Figure 2D; Supplementary Table S3). These comparisons showed that the genomic sites bound by PBX1 in aNS were frequently also occupied by the same small group of TFs in other physiological settings. Prominent members of this group were TCF3, SREBF2, SMAD4, RFX1/2/3, PBX1/2/3, OLIG2 and NFI transcription factors (Figure 2D, Supplementary Table S3). This association was particularly high when ChIP-seq data from early neural cell populations were analyzed. In fact, over 80% of the genomic sites bound by RFX1/2/3, SMAD4, SREBF2 or TCF3 in ChIP-seq studies from early neural cellular contexts corresponded to positions bound by PBX1 in aNS, as did close to 70% of sites bound by NFI or OLIG2. As public domain-derived data differ in terms of width and number of annotated peaks, rigorous control analyses were conducted. When the peaks in the respective dataset were computationally shifted in a random manner and intersected with aNS PBX1 ChIP-seq data, the degree of overlap dropped to 3.48% on average, thus confirming the robustness of our intersections (Supplementary Figure S3A–C). Visual inspection of example locus *Olig2* furthermore demon-

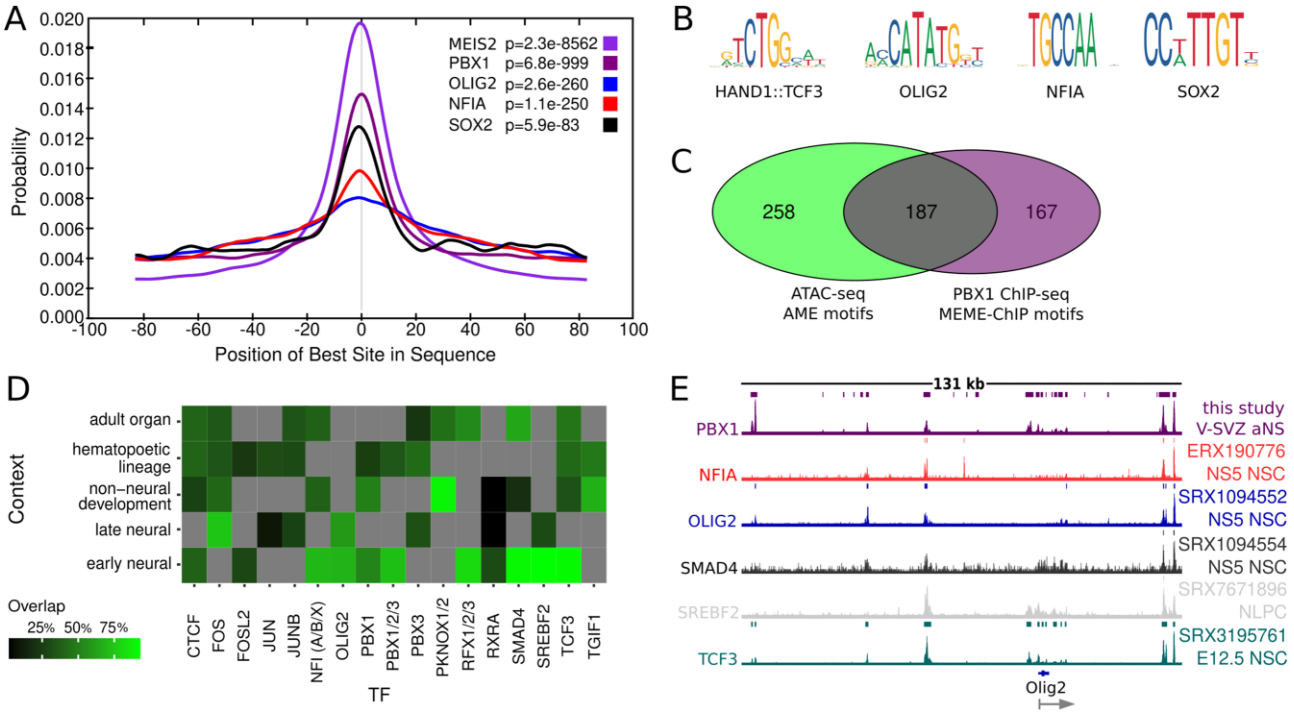


Figure 2. *In silico* prediction of PBX1 interaction partners. (**A**) Selection of overlapping enriched motifs found under PBX1 ChIP-seq peaks summit by MEME-ChIP analysis. (**B**) PBX1-associated motifs in aNS and motif enrichment of open chromatin regions (ATAC-seq) harbor E-box/bHLH class I and bHLH heterodimer, bHLH class II, SMAD/NF-1 DNA-binding domain factor and High-mobility group (HMG) domain factor motifs. (**C**) Intersection of JASPAR 2022 CORE vertebrate non-redundant pfms motifs derived from PBX1 ChIP-seq and open chromatin in aNS. (**D**) Genome-wide intersections of PBX1 ChIP-seq in aNS with ChIP-seq of the TFs indicated for potential interaction (references to datasets see Supplementary Table S3; overlaps between .bed files were identified pairwise with bedtools intersect). Percentage of overlapping intervals is color-coded (0% / black to 100% / bright green; grey: no data available). (**E**) Co-visualization of ChIP-seq profiles of *in silico* predicted interactors at example locus *Olig2* demonstrates shared binding pattern and distribution compared to PBX1 (ChIP-seq datasets can be identified by their ChIP-Atlas ID; NSC = neural stem cells with NS5 cells for NFIA, OLIG2 and SMAD4, E12.5 embryonic NSC for TCF3, NLPC = neural lens progenitor cells). Colored bars represent called peaks, grey arrow direction of transcription.

strated co-localization of PBX1 with similar peak shapes and patterns for several of these factors including OLIG2 itself, NFIA, SREBF2 and TCF3 (Figure 2E).

We also asked whether we can derive the underlying GRNs by performing bioinformatic network predictions on our RNA-seq and ChIP-seq data. We applied the collaborative gene list enrichment analysis tool Enrichr to the RNA-seq data obtained from *Pbx1* kd to predict possible TFs that contribute to the up- or downregulation of genes. In parallel, we subjected our PBX1 ChIP-seq data to reverse analysis by ISMARA to model GRNs associated with TFs bound near PBX1-occupied sites. Notably, computational modeling predicted both, chromatin binding together with PBX1 and participation in the regulation of PBX1-dependent genes, for the same small group of TFs, consisting of SMAD4, NFIA, SOX2, SREBF2 and TCF3 (Supplementary Figure S5). Several of these TFs, such as SMAD4, SOX2 and NFIA, are known regulators of adult V-SVZ neurogenesis or had been linked to PAX6-induced neurogenic differentiation in the V-SVZ (24,69,70). Hence, *in silico* predictions reconstructed a larger GRN in which PBX1 targets multiple other TFs, which are themselves connected by extensive cross-regulatory interactions (Supplementary Figure S5G).

In conclusion, by applying different bioinformatic approaches to our ChIP-seq, ATAC-seq and RNA-seq results and making extensive use of published data, we were able to derive a network of putative PBX1-interacting proteins in the adult V-SVZ neurogenic stem cell system, whereby interaction of PBX1 with its partners occurs in association with chromatin and is more prevalent in early neural tissues than others.

Novel PBX1 interactors and confirmation of *in silico* predicted partners by mass spectrometry

To experimentally test the predicted PBX1 interactions, we conducted immunoprecipitation experiments with PBX1 in aNS cells followed by LC–MS/MS. This analysis revealed a rich array of interacting proteins, including both known transcription factors and novel partners with diverse biological functions (Figure 3A, Supplementary Table S7). Known PBX1-interacting proteins MEIS1/2 and PKNOX1/2, as well as PBX3 were among the co-precipitated proteins (Figure 3A). Intriguingly, we detected associations between PBX1 and several TFs that we had predicted as potential PBX1-interactors, including OLIG2, SOX2, TCF12 and NFI factors. In addition, SOX2 and TCF4, which had not met our stringent inclusion criteria for the initial *in silico* analysis by genome-wide intersection, co-precipitated with PBX1 (Figure 3A). Drawing on our previous strategy, both were subjected to multiple pairwise comparisons of available public domain-derived ChIP-seq data (reference to respective datasets in Supplementary Table S3). This approach revealed a high degree of overlap between PBX1-bound genomic regions in aNS and sites bound by SOX2 (Figure 3B) or TCF4 (Figure 3C), with particularly high overlap in cells of neural or neuronal

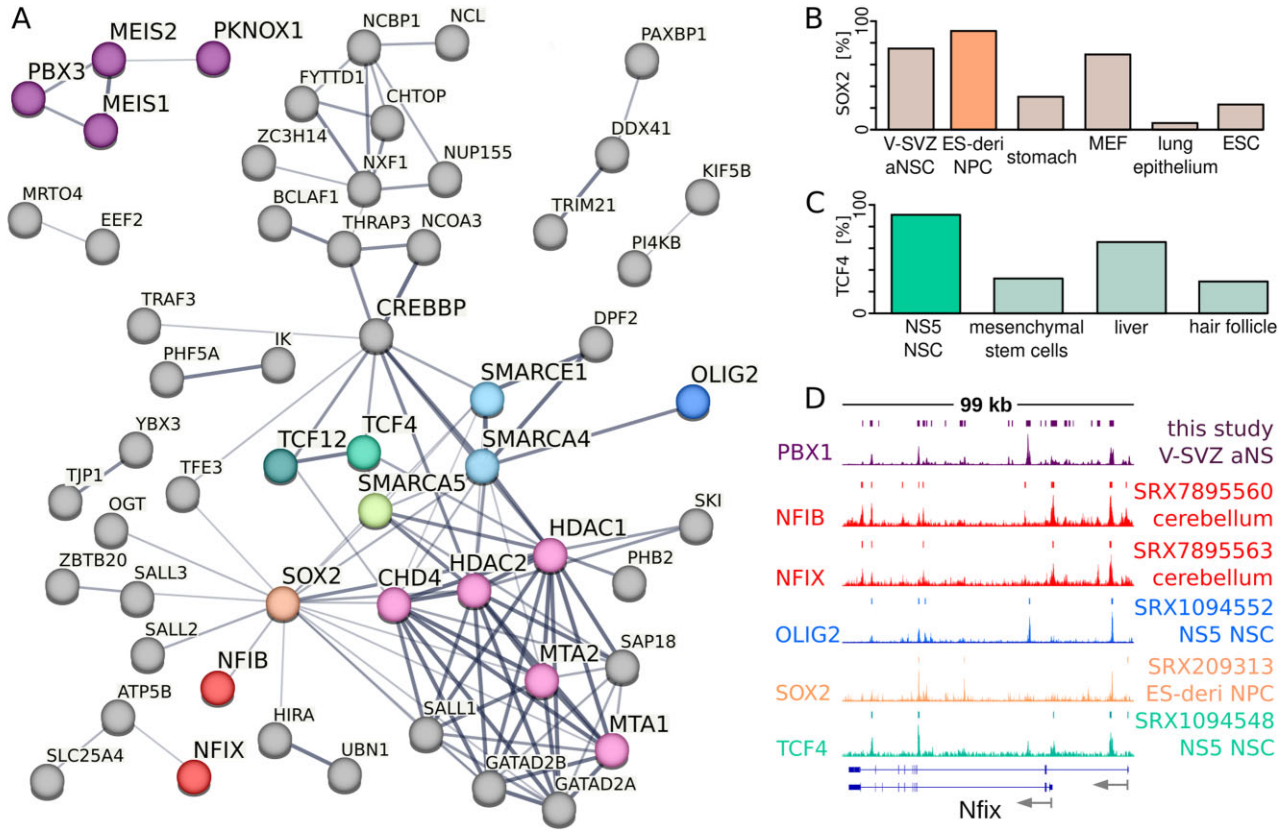


Figure 3. LC-MS/MS identifies novel PBX1 interactors and confirms *in silico* predicted interaction partners. (**A**) STRING Interaction network of PBX1 co-precipitators in mass spectrometry (physical interaction map; filter: no RNA-binding / ribosomal proteins, splicing factors and factors with no connection to network proteins). (**B, C**) Genome-wide intersections of V-SVZ derived aNS PBX1 ChIP-seq with ChIP-seq results obtained for SOX2 (**B**) or TCF4 (**C**). Overlaps between .bed files were identified pairwise with bedtools intersect; graphs show percentage of overlaps. ChIP-seq of various biological contexts shows pronounced co-localization in embryonic and adult neurodevelopmental contexts but low to minimal co-localization in non-neural cells and tissues (NPC: ESC-derived neural precursor cells, MEF: mouse embryonic fibroblast, ESC: embryonic stem cells, NS5 NSC: NS5 neural stem cell line; Supplementary Table S3). Tracks spanning the *Nfix* gene locus taken from ES-derived NPCs (SOX2, orange) and NS5 cells (TCF4, green) are shown in (D). (**D**) Example ChIP-seq profiles of PBX1 and mass spectrometry validated interaction partners NFIB, NFIX, OLIG2, SOX2 and TCF4 at the *Nfix* gene locus. Colored bars represent called peaks, grey arrows direction of transcription; ChIP-seq datasets can be identified by their ChIP-Atlas ID and are listed in Supplementary Table S3.

origin. Visualization of ChIP-seq data for PBX1-binding factors NFIB, NFIX, OLIG2, SOX2 and TCF4 obtained from early neural cell sources demonstrated striking similarities in position and shape of the ChIP-peaks, with the *Nfix* genomic locus given as example (Figure 3D; Supplementary Table S3). This is even more remarkable considering that the data originate from sources as diverse as cerebellar progenitors (NFIB, NFIX), NS5 NSC cells (OLIG2, TCF4) or ES-derived neural precursor cells (SOX2). Our mass spectrometry analysis also unveiled PBX1's association with key components of chromatin remodeling complexes active in neurodevelopmental contexts (71), including chromatin remodeling components of the CHD/NuRD (CHD4, MTA1/2 and HDAC1/2, RBP48) and SWI/SNF/ISWI (SMARCA4/BRG1, SMARCA5/SNF2H and SMARCE1/BAF57) complexes (Figure 3A). PBX1 also co-precipitated with various RNA-processing factors (Supplementary Table S7), indicative of potential involvement in the regulation of post-transcriptional RNA-related processes.

To gain insights into the biological implications, we subjected filtered interactors (excluding RNA-binding/ ribosomal proteins, and splicing factors) to GO term analysis with PANTHER. According to fold enrichment of terms, the top two terms were 'regulation of cell fate specification' (FE = 55.31, $P = 1.65$ E-06, statistical analysis with Fisher's Exact test using Bonferroni correction for multiple testing) and 'regulation of cell fate commitment' (FE = 37.30, $P = 1.88$ E-05). 'Gliogenesis' (FE = 7.53, $P = 3.75$ E-02), 'chromatin organization' (FE = 6.74, $P = 2.24$ E-05), 'negative and positive regulation of transcription by RNA polymerase II' (FE = 5.74, $P = 3.08$ E-08 and FE = 5.69, $P = 1.11$ E-11, respectively), and 'brain development' (FE = 5.11, $P = 2.55$ E-03) were among the top 25 terms. Mass spectrometry thus confirmed physical association of PBX1 with many of the proteins that we had predicted as potential interactors by computational modeling, including several TFs with established functions in neurodevelopment, such as SOX2, TCF4/TCF12, NFIB and NFIX, and OLIG2.

Whole transcriptome spatial RNA analysis reveals distinct distribution of PBX1-associated GRN components in the V-SVZ -OB neurogenic system

TFs, which participate in shared developmental processes, are expected to exhibit the same or at least overlapping spatial/temporal expression. Neurogenesis in the V-SVZ

follows a distinct differentiation trajectory in which progressive steps of cellular maturation take place at distinct positions along the migratory route from the V-SVZ to the OB, making it an excellent system to study gene expression dynamics in space and time (72). We therefore assessed the spatial distribution of *Pbx1* transcripts and its interactors along the neurogenic differentiation trajectory from the V-SVZ to the OB by whole transcriptome spatial RNA analysis at 11 different locations (Supplementary Table S8). Samples were taken from the left and right V-SVZ at ventral, medial and dorsal positions with the latter including the RMS, from caudal and rostral positions of the RMS, and from different areas in the OB, corresponding to GABAergic granule cells in the inner OB and periglomerular neurons and mitral cells in the outer OB. Hierarchical cluster analysis of spatial expression levels of all PBX1-interactors that had been identified by mass spectrometry demonstrated distinctly different expression among samples, with particularly pronounced differences between V-SVZ and OB (Figure 4A, B).

We followed the spatial expression of *Pbx1* and several of its interactors across the probe set (Figure 4C–G, Supplementary Figure S6A–D). *Pbx1*, its known heterodimerization partner *Meis2*, and the related *Pbx3* exhibited a biphasic expression dynamic whereby transcript levels were highest in the V-SVZ (samples 001–004), decreased in the RMS, before increasing again in the inner part of the OB, presumably in neuroblasts entering the OB and granule cells (samples 008 and 009; Figure 4C, Supplementary Figure S6E). Neurodevelopmental transcription factors of the *Nfi* family, *Tcf4* and *Sox2* showed strong expression in the V-SVZ but lacked the second wave of upregulation in the OB characteristic of *Pbx*/*Meis* (Figure 4D–F). These expression trajectories were further supported by the expression patterns of the respective genes reported in the Allen Mouse Brain Atlas (Supplementary Figure S7). Similar transcriptional dynamics were also observed for selected PBX1-co-precipitating epigenetic modifying enzymes and components of chromatin remodeling complexes (Figure 4G). Spatial whole transcriptome analysis therefore allowed us to define two phases of *Pbx1* expression: an early phase in the V-SVZ germinal niche during which *Pbx1* expression is matched by that of many of the interactors observed by mass spectrometry in aNS, and a second phase of upregulation that roughly coincides with neuroblast entry into the OB, during which no induction of expression of these interactors was observed. Therefore, the PBX1-centered GRN established here likely reflects the regulatory network in which PBX1 engages during early steps of V-SVZ neurogenesis.

To test this assumption in an independent dataset, we probed published single cell transcriptomics data obtained from the adult mouse V-SVZ for *Pbx1* and key PBX1-interacting TFs (73). This dataset only contains information on cells taken from the V-SVZ and therefore does not allow to draw conclusions about transcript expression in migrating neuroblasts or neurons in the OB. Still, we could confirm increased expression of *Pbx1* together with *Tcf12*, *Nfia* and *Sox2* in activated NSCs (aNSCs) and TAPs, and continuing strong expression of *Pbx1* but decreasing expression of interactors as neuroblasts mature (Figure 4H–K). Taken together, these results establish that *Pbx1* and many of its interactors share common transcriptional dynamics in cells associated with the V-SVZ germinal niche.

bHLH class I proteins are PBX1 interaction partners

To explore the functional relevance of these interactions, we focused on the class I bHLH proteins *Tcf3* and *Tcf4*. *Tcf4* was chosen for its prominent expression in the V-SVZ neurogenic system, while *Tcf3* was chosen because the oncogenic t(1;19) fusion event involving *PBX1* and *TCF3* (also known as *E2A*) is of considerable clinical relevance for ALL (3,4) and in light of the known functional compensation observed for this protein family (74). *Tcf12* had to be excluded due to the lack of suitable ChIP-seq data from neural progenitor cells or developing neurons. We first compared the overall peak distribution for PBX1 in aNS with ChIP-seq datasets for TCF3 in stem- and progenitor cells from E12.5 mouse cortices (48) and TCF4 in the NS5 NSC cell line (25). For completeness, all ChIP-seq peaks were included into the analysis. Peak annotation using ChIPseeker revealed parallel peak distribution for PBX1, TCF3 and TCF4 in close proximity to promotor regions (≤1 kb TSS), accounting for 22.63% (PBX1), 32% (TCF3) and 22.44% (TCF4) of peaks, respectively, demonstrating nuanced differences in the genomic binding profiles of both TCF proteins (Figure 5A-C). Comparison of PBX1, TCF3 and TCF4 ChIP-seq profiles demonstrated striking similarities in chromatin-binding distribution at example loci such as *Itgb5* (Figure 5D). Visualization of E-box motifs derived from MEME-ChIP motif analysis (Supplementary Table S5, Figure 2A) identified core motifs for TCF3, TCF4 and TCF12, co-localized with the PBX1 motif at the PBX1 ChIP-seq peak center (Figure 5E). Secondary motif analysis of PBX1- and TCF3-bound sites revealed that both motifs were significantly enriched with a distance of 18 nt (Figure 5F).

TCF3 was not among the PBX1-co-precipitating proteins detected by mass spectrometry, but this may be due to its relatively low expression in the V-SVZ niche compared to *Tcf4* (Figure 4D, Supplementary Figure S7). E47 is the prominent TCF3-isoform in the embryonic mouse forebrain and critically involved in the differentiation of cortical projection neurons (48), but commercial, high-affinity antibodies specific for the protein are missing. Co-immunoprecipitation experiments were therefore performed with nuclear protein extracts of HEK293T cells transfected with *Tcf3/E47* fused to a triple flag-tag (*Tcf3/E47-flag*). PBX1 precipitated together with TCF3/E47, consistent with direct, physical interaction between the two proteins (Figure 5G). Co-immunoprecipitation and GST-pull down experiments with truncated forms of PBX1 and E47 mapped association to the N-terminal PBC domains of PBX1 and the E47-C-terminus containing the bHLH domain (Supplementary Figure S8). PBX1–TCF3 interaction in the nuclear compartment was also confirmed *in cellulo* by proximity ligation assay (PLA) with FLAG- and PBX1-specific antibodies and adult V-SVZ neural stem- / progenitor cells that were transduced with pWPI-retroviruses carrying *Tcf3/E47-flag* together with IRES-GFP for visual detection (Figure 5H–H"'). Together, these results established that PBX1 and TCF proteins can engage in the formation of stable heteromeric complexes.

Shared gene expression changes following *Pbx1*, *Tcf3* and *Tcf4* depletion

To assess the biological function of PBX1–TCF interactions, we performed siRNA-mediated knockdown of *Pbx1*, *Tcf3* and *Tcf4* in aNS cells, followed by *in vitro* differentiation. Non-targeting siRNA validated for murine transcripts

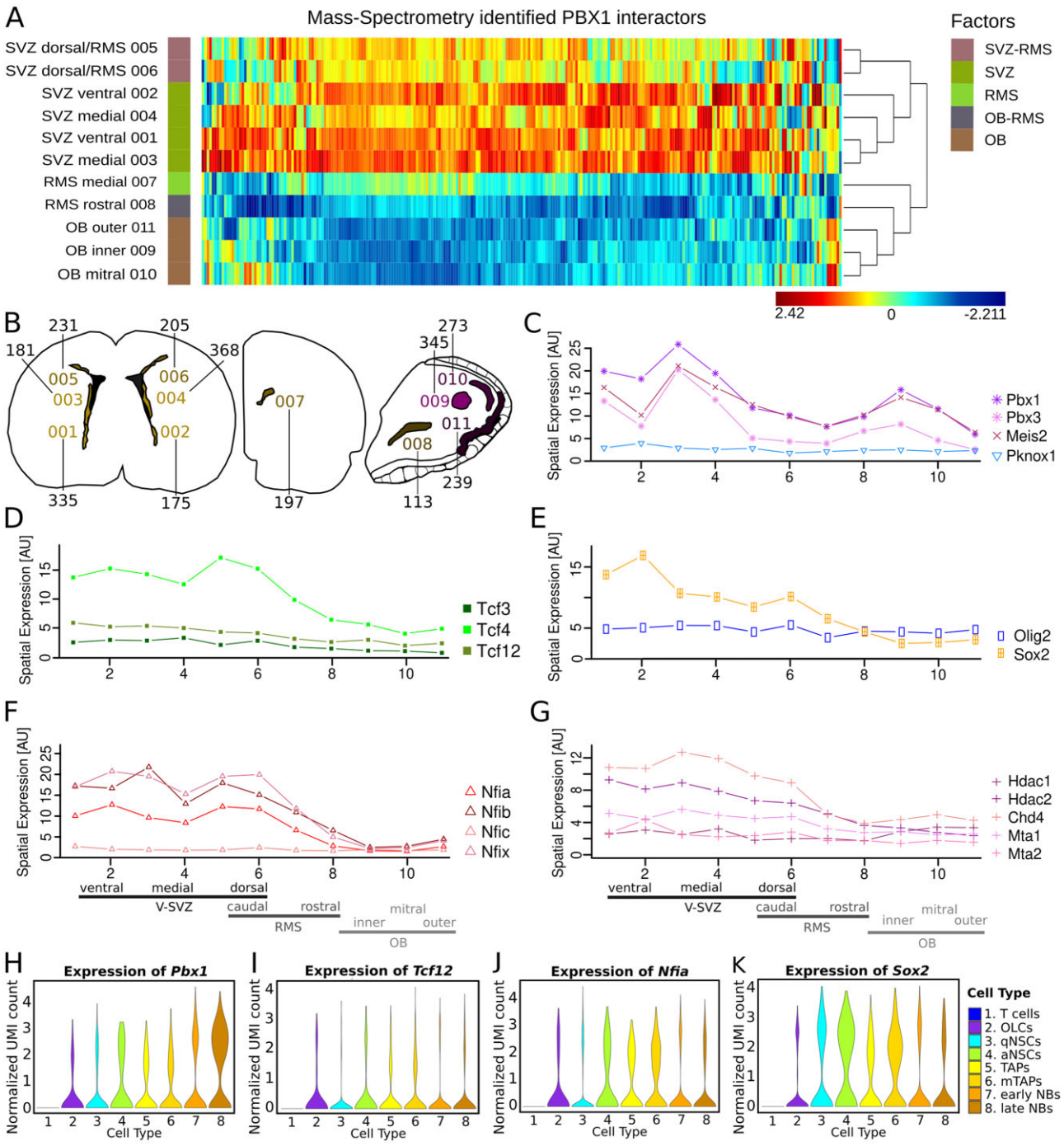


Figure 4. Whole transcriptome spatial RNA analysis and scRNA-seq support similar transcription dynamics of PBX1 interactors. (**A**) Cluster analysis of MS-identified PBX1 interactor transcription dynamics according to spatial RNA analysis demonstrates differential expression between RMS, OB and V-SVZ-associated regions. (**B**) Schematic representations of coronal sections through the forebrain at the level of the V-SVZ (left), RMS (middle) and OB (right); positions 001 to 011 from which cells were sampled for spatial transcriptomics are indicated in color; black numbers specify the number of cells sampled from each position; sections are not drawn to scale. (**C**) Expression profile of TALE-HD transcription factors in sample numbers 001 to 011. (**D-G**) Spatial expression along the V-SVZ differentiation trajectory of bHLH class I transcription factors (**D**), Sox2 and Olig2 (**E**), Nfi transcription factors (**F**), and PBX1-interacting chormatin remodeling complex components (**G**). (**H–K**) Selection of single-cell transcriptomics of the above factors from the V-SVZ neurogenic niche (73).

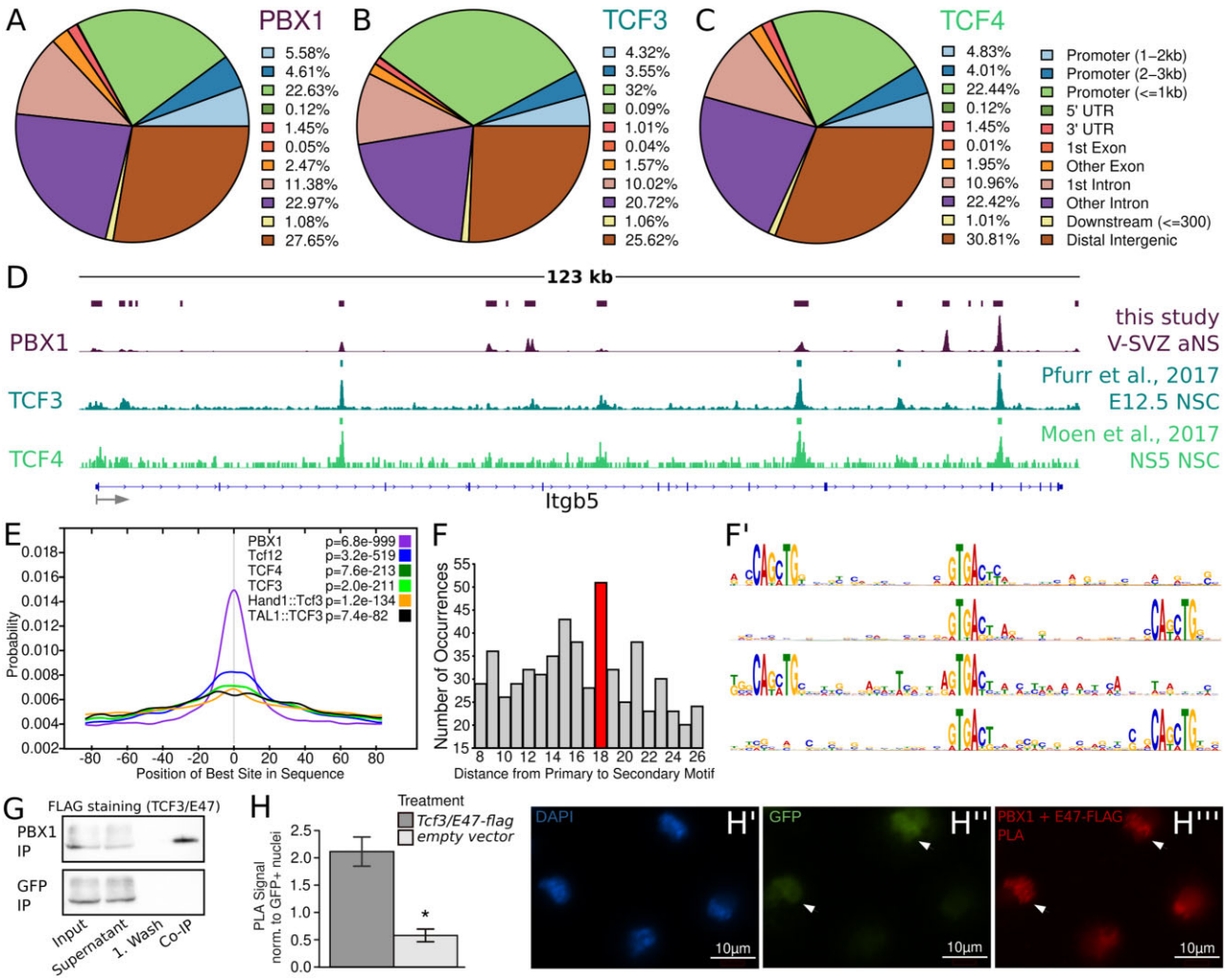


Figure 5. bHLH class I proteins are PBX1 interaction partners. (**A–C**) Similar peak distribution of neurodevelopmental ChIP-seq according to annotation with ChIPseeker: (**A**) PBX1 (V-SVZ aNS), (**B**) TCF3 (E12.5 NSC), (**C**) TCF4 (NS5 NSC). (**D**) ChIP-seq profiles of PBX1, TCF3 and TCF4 show similar pattern and distribution at example locus *Itgb5*. Colored bars represent called peaks, grey arrow direction of transcription. (**E**) CentriMo analysis of PBX1 ChIP-seq peaks demonstrates PBX1 and bHLH class I motif co-occurrence at peak center. (**F, F'**) Secondary binding analysis (SpaMo) of intersecting PBX1 (V-SVZ aNS) and TCF3 (E12.5 NSC) ChIP-seq peaks with MEME-ChIP; TALE-HD and bHLH class I motifs are spaced with a gap of 18 nt ($P = 8.23e\text{-}2$; F), on both strands and in all directions (**F'**). (**G**) PBX1 in aNS cells co-precipitates TCF3/E47-FLAG in Co-IP of HEK293T protein lysate. (**H-H'''**) Proximity ligation assay for PBX1 and TCF3/E47-FLAG; (**H**) Script-based automated quantification revealing significant enrichment of punctate PLA signal in *Tcf3/E47-flag* overexpressing cells compared to empty vector control treatment, paired Student's *t*-test ($P = 0.032$); (**H'–H'''**) Example staining showing speckled PLA signal in GFP + cells (arrow heads) as opposed to GFP low/- cells.

served as control. Depletion of each TF significantly reduced neuronal differentiation compared to the control, as evident from fluorescence activated cell sorting (FACS)-based quantification of PSA-NCAM positive neuroblasts that were generated under each condition (Figure 6A). To assess genome-wide gene expression changes, we subjected aNS cells following kd of *Pbx1*, *Tcf3* or *Tcf4* to RNA-seq (Supplementary Figures S2, S9, S10; $n = 3$ each; *Tcf3* kd: 69.44 ± 10.37%; *Tcf4* kd: 85.58 ± 8.26% in RT-qPCR). Each experimental sample was matched with a reference sample generated from aNS transfected with non-targeting siRNAs. 772 genes were significantly differentially regulated (up or down) following *Pbx1* kd, 651 genes after *Tcf3* kd and 1284 genes after *Tcf4* kd (Figure 6B, Supplementary Table S6). Assessing PBX1 occupancy at DEGs revealed that 88.2% (681) of *Pbx1* kd-dysregulated genes were bound by PBX1 in aNS ChIP-seq, while 87.9% (572) of *Tcf3* kd and 89.2% (1119) of *Tcf4* kd-dysregulated genes were occupied by PBX1. Intersection of these DEGs revealed 66 commonly dysregulated targets upon kd of either factor, while *Pbx1/Tcf3* kd affected additional 98 and *Pbx1/Tcf4* kd additional 160 genes (Figure 6B, Supplementary Figures S9, S10). GO biological process analysis of the 66 commonly dysregulated genes revealed 'axon guidance' (FE = 10.60, $P = 4.98$ E-06, FDR = 2.62E-02) and 'neuron projection guidance' (FE = 10.56, $P = 5.12$ E-06, FDR = 2.02 E-02) as the two top terms. Consistent with these GO-terms, GABAergic and glutamatergic receptors *Gria1* and *Grin2b*, and the extracellular matrix-associated *Reln* were prominent DEGs under all conditions (Figure 6C).

Intrigued by these similarities and motivated by the strong oncogenic activity of the TCF3-PBX1 fusion protein (also referred to as E2A-PBX1) that results from the t(1;19) fusion event in ALL, we examined the genes co-regulated by PBX1 and TCF3 in more detail. We subjected targets that

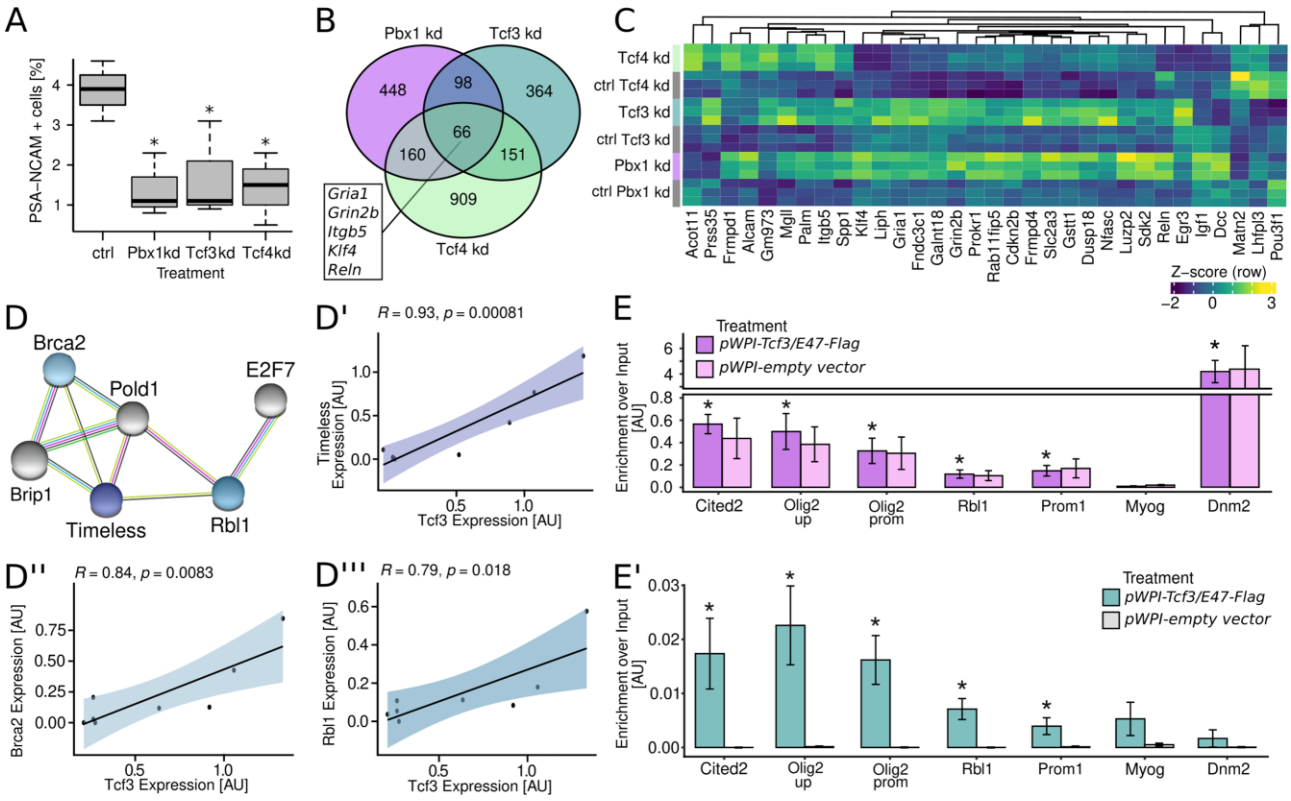


Figure 6. PBX1-TCF interaction in neural progenitor cells. (**A**) Proportion of PSA-NCAM positive-neurons generated from aNS upon *Pbx1*, *Tcf3* and *Tcf4* kd in aNS and quantified by FACS analysis ($n = 3$, $P < 0.05$, paired two-tailed Student's t-test). (**B**) Venn diagram of commonly dysregulated genes upon *Pbx1*, *Tcf3* and *Tcf4* kd identified by RNA-seq. (**C**) Heatmap and cluster analysis (euclidean clustering based on average) of commonly *Pbx1*, *Tcf3* and *Tcf4* kd dysregulated genes in RNA-seq with $\log_2$FoldChange >0.3. (**D**) Commonly downregulated genes upon *Pbx1* and *Tcf3* kd partake in replication-associated interaction network (STRING analysis). (**D'–D"'**) Expression of *Tcf3* correlated with that of *Brca2*, *Rbl1* and *Timeless* in scRNA-seq of V-SVZ (75); Pearson Correlation test. (**E, E'**) ChIP-qPCR of PBX1 (**E**) and TCF3 transcript variant E47-FLAG (**E'**) at *Olig2* upstream enhancer and promotor, *Cited2*, *Rbl1*, *Prom1*, *Myog* and *Dnm2* in *Tcf3/E47-Flag* transduced aNS cells. Statistic analysis with paired Student's *t*-test (FLAG IP: TCF3/E47-FLAG vs. empty vector control, PBX1 IP: respective gene against neg. ctrl. Myog, $n = 3$–4, $^*P < 0.05$).

were downregulated following *Pbx1*- and *Tcf3* kd to STRING analysis. This approach revealed a replication-associated target network including *Brca2*, *Rbl1* and *Timeless* (Figure 6D). Pearson correlation analysis of public scRNA-seq data from the V-SVZ (75) demonstrated correlation of *Tcf3* expression with all three targets (Figure 6D'-D"'). Commonly downregulated DEGs following depletion of *Pbx1* or *Tcf3* were primarily associated with GO-terms related to DNA replication, while commonly upregulated genes were linked to GO-terms related to neuronal differentiation and maturation (Supplementary Figure S9C, D). These findings raise the intriguing possibility that the PBX1–TCF3 module may have a role in balancing gene expression programs that act specifically in cell populations, which are in a state of transition from cellular proliferation to differentiation, such as adult neural progenitor cells. We also asked whether the transcriptional changes observed after *Pbx1* or *Tcf3* kd were associated with chromatin binding of PBX1 and TCF3 to the respective genomic loci. To this end, we assessed promotor occupancy of the commonly dysregulated targets *Cited2*, *Rbl1* and *Prom1* and two regulatory regions of *Olig2* (upstream enhancer and promotor) by ChIP-qPCR in *Tcf3/E47-Flag*-transduced aNS alongside two control loci, *Myog*, a known PBX1-target during skeletal muscle differentiation not expressed in the V-SVZ (9,17), and *Dnm2*, a gene whose promotor is occupied by PBX1 but not TCF3 (Supplementary Figure S11). PBX1 occupied all selected target sites, while FLAG-specific antibodies significantly enriched sites that were bound by PBX1 in common DEGs and *Olig2* regulatory regions, but neither *Dnm2* nor *Myog* (Figure 6E, E').

We next asked whether PBX1-TCF3 interaction may extend beyond neurogenesis. To this end, we compared PBX1 and TCF3 genome-wide binding in ChIP-seq datasets obtained from neural and hematopoietic cell populations. In the t(1;19) translocation-bearing cell line 697, 9098 genomic sites were bound by the TCF3-PBX1 fusion protein, 1816 of which overlapped with sites bound by PBX1 in aNS, and 274 positions additionally overlapped with sites bound by TCF3 in NSC from the mouse embryonic cortex (48,57). Genomic sites occupied by the TCF3-PBX1 fusion protein in 697 cells also exhibited similar genome-wide binding properties as PBX1 or TCF3 in neurodevelopmental cell populations, and genes associated with binding in all three settings were linked to GO-terms that suggest common involvement in cell proliferation and hematopoiesis (Supplementary Figure S12A–E). Moreover, comparing PBX1 and TCF3 binding profiles at example gene loci derived from our RNA-seq experiments, the retinoblastoma-like 1 (*Rbl1*) promotor and the promotor of minichromosome maintenance complex component 5 (*Mcm5*), a part of the DNA replication licensing complex, showed that both TFs occupied corresponding sites in neurodevelopmental cell populations, leukemic cell

lines with t(1;19) fusion (ICN12 and RCH-ACV), and B-lymphoblastoids (Supplementary Figure S12F-G'). Together, these observations suggest that PBX1 and TCF3 might interact to co-regulate proliferation-associated genes across biological contexts.

To obtain information on possible shared PBX1–TCF3 targets in leukemogenesis, we performed additional unbiased paradigms of data integration of ChIP-seq and RNA-seq data derived from the public domain (Figure 7A; (57)). This analysis identified 41 genes potentially regulated by both transcription factors, including B Lymphocyte Kinase (*BLK*) and *NOTCH3* (Supplementary Table S9). *Notch3* is a presumed PBX1-target gene as its expression was significantly downregulated upon *Pbx1* kd, whereas *Blk* is not expressed in aNS cells according to our RNA-seq results (Supplementary Table S6). Nevertheless, PBX1 and TCF3 ChIP-seq profiles at both *Blk* and *Notch3* were not only remarkably alike in cells of neural origin (Supplementary Figure S13A, B), they also shared many similarities across human hematopoietic and leukemogenic cell populations irrespective of their t(1;19) translocation status (Supplementary Figure S13A', B'). Of note, upon viral overexpression of *Pbx1a* in Nalm6 cells, a *PBX1*-negative but *TCF3*-positive human pre-B cell line (Supplementary Figure S13C, D), expression of *PBX1* was positively correlated with that of *BLK* and *NOTCH3* (Figure 7B, C). Hence, ectopic expression of PBX1 was sufficient to induce the expression of two *in silico* predicted, disease-relevant target genes. Finally, to investigate to which extent PBX1 and TCF3 cooperate in the control of cell proliferation, aNS cells were treated with siRNAs targeting *Pbx1*, *Tcf3* or with a combination of both, and subjected to spheroid assays (Figure 7D). Combination treatment with both siRNAs significantly augmented the growth-reducing effect of either siRNA alone. Together these findings suggest that PBX1 and TCF3 jointly control cell proliferation in aNS cells, a property which may be hijacked by the PBX1-TCF3 fusion protein in ALL.

Discussion

Tissue homeostasis of essentially all adult organs relies on the activation of resident stem cells, followed by precisely coordinated phases of cell proliferation, cell fate specification and cellular differentiation. Ordered transition through these phases requires the interaction and mutual cross-regulation of TFs. Focusing on the generation of neurons in the adult V-SVZ neurogenic niche and PBX1, a broadly expressed homeodomain TF in this germinal niche, we here provide an integrated multi-omics approach to decode a PBX1-centered GRN that underlies early steps of adult neurogenesis. On the one hand, our study exemplifies how the application of different bioinformatic analysis formats can be used to make predictions whose robustness can withstand subsequent experimental testing. On the other hand, our results shed light on how the successive maturation of adult neuronal progenitor cells can be controlled by an interacting network of TFs. At the core of this network is a PBX1–TCF3/TCF4 transcriptional module with likely relevance for leukemia etiology.

ChIP-seq analysis of PBX1 binding to chromatin demonstrated numerous occupied sites in V-SVZ derived adult neural stem- and progenitor cells, indicating a wide range of functions for PBX1 in this system. Integration of PBX1 ChIP-seq motif analysis with ATAC-seq motif enrichment analysis, followed by substantiation of the results with publicly available data allowed us to predict key neural TFs SOX2 and OLIG2, as well as members of the NFI- and TCF- families as novel PBX1-interactors in adult neural stem- and progenitor cells. These *in silico* predictions were not only confirmed by subsequent LC–MS/MS analysis, but association of individual components of this network had also been reported for other neurodevelopmental settings. For example, mass spectrometry experiments performed with the mouse embryonic NSC line NS5 had placed SOX2, TCF4 and OLIG2 into a common regulatory network related to mental disorders (25). In a comprehensive proteomic analysis conducted in the same cell line, TCF4, TCF12, NFIA, NFIB, OLIG2 and SOX2 copurified together with the Mediator complex (76). Computational analysis further identified SOX2 along with PBX1 and TCF4 as the most relevant factors in reprogramming epithelial-like cells to induced neural stem cells (77). Lastly, ISMARA-based gene regulatory network analysis had suggested NFIA as potential interactor of PBX1 and PAX6 in V-SVZ-derived aNS cells (24), and PBX1 and NFIA share common target genes, including *Aqp4* and *Nfasc*, during the generation of astrocytes from mouse embryonic stem cells (78). The results reported here bridge these diverse reports and place TALE-HD/SOX2/NFI/TCF-associated GRNs at the core of cell fate specification processes in diverse neurodevelopmental settings (Figure 7E).

Our findings also underscore the notion that reorganization of protein-protein interactions within GRNs drives developmental progression. Working with primary aNS cultures and, hence, with cells that are competent of differentiation towards all neural lineages, immunoprecipitation with antibodies specific for PBX1 captured multiprotein complexes containing diverse cell fate determinants. In separate experiments, spatial whole-transcriptome RNA analysis of the V-SVZ lineage revealed a biphasic expression dynamic for *Pbx1* with high expression in the V-SVZ germinal niche and a second wave of expression in the OB. These findings support the concept that PBX1 serves at least two roles during adult V-SVZ neurogenesis: as part of an organizing structure in highly proliferative, multipotent neural progenitor cells, as we report here, and as platform for neuronal cell fate specification at later stages of cellular maturation, consistent with published reports (17,18). In this regard, PBX1 resembles the bHLH class II factor OLIG2, which acts as regulator of self-renewal within a multifactorial GRN in the NS5 NSC line (79) and as critical determinant of oligodendroglial lineage acquisition (80). In addition, the fact that none of the proteins that PBX1 associates with in aNS exhibits a comparable transcriptional upregulation in the OB supports the notion that the composition of PBX1-containing TF complexes changes along the developmental trajectory of V-SVZ neurogenesis. In fact, many of the PBX1-interacting proteins identified here undergo dynamic complex formation to gain specificity. The TCF3 splicing isoform E47, for example, heterodimerizes with ASCL1 and NEUROG2 to activate neurogenic gene expression, but induces oligodendrocyte fate upon heterodimerization with OLIG2 (81). Similarly, NFI factors can promote proliferative quiescence, while also controlling cell specification of bipolar interneurons in late retinal progenitor cells (82) and NFIA has been described as gliogenic switch enabling astrocytic differentiation (78,83). Collectively these findings argue that the specificity of PBX1 and its binding partners depends on the syntax of the TF complexes they form in distinct cellular contexts.

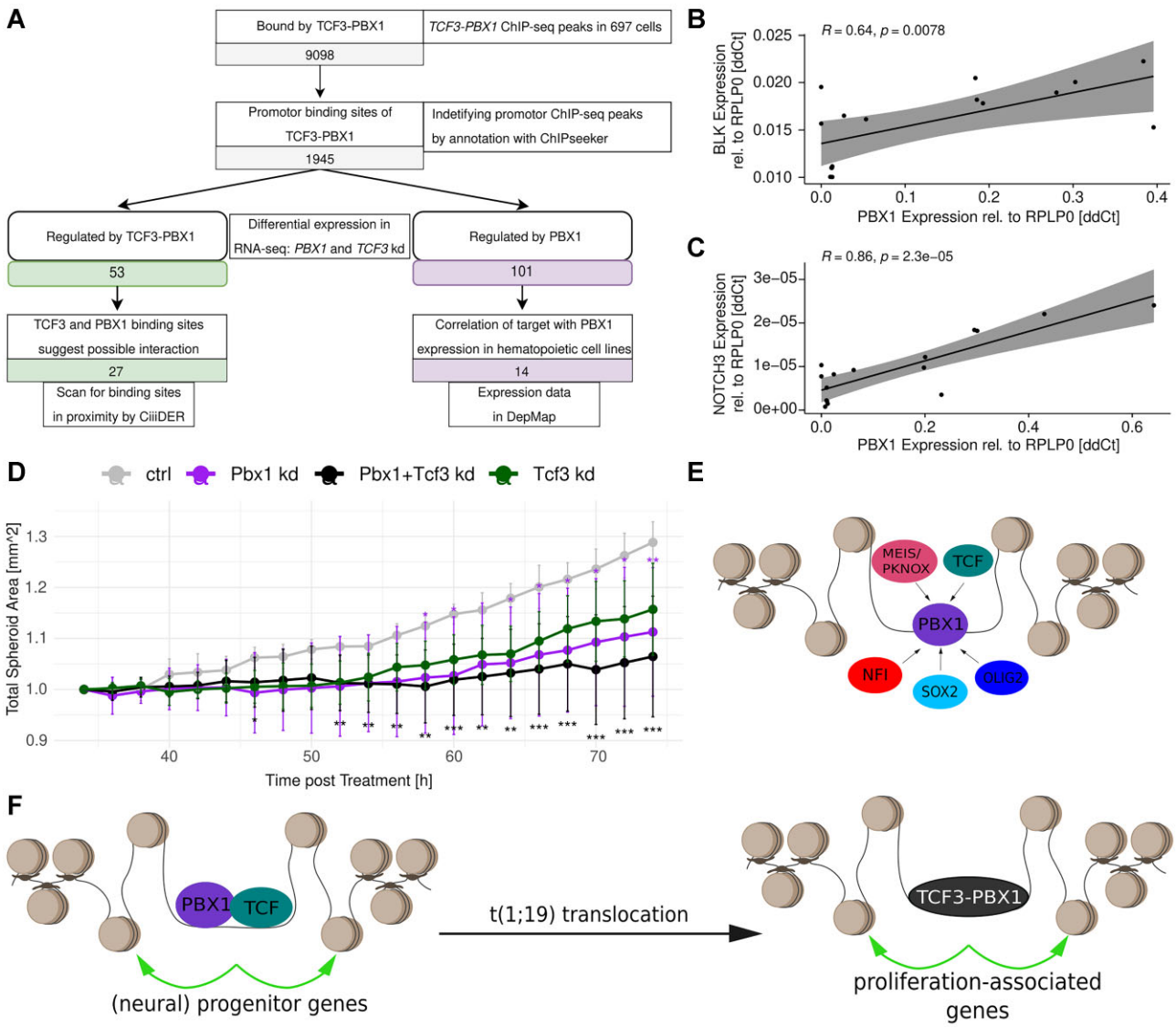


Figure 7. PBX1-TCF3 interaction in hematopoiesis. (**A**) *In silico* prediction of public domain ChIP-seq and RNA-seq data identifies *BLK* and *NOTCH3* as direct target of PBX1-TCF3 interaction in the hematopoietic context. (**B, C**) Transcript quantification of *BLK* ($P < 0.01$) (**B**) and *NOTCH3* ($P < 0.001$) (**C**) expression in Nalm6 cells following viral transduction of *Pbx1a* ($n = 16$, Pearson Correlation test). (**D**) Spheroid assay of aNS with *Pbx1*, *Tcf3* or combined knockdown ($n = 4$–7; statistical significance: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ with unpaired Student's *t*-test). (**E**) TALE-HD, TCF, NFI, SOX2 and OLIG2 are putative binding partners of PBX1 in adult neural progenitor cells. (**F**) The ability of PBX1 and TCF3 to regulate GRNs in (neural) progenitor cells might be hijacked by the leukemogenic TCF3-PBX1 fusion protein, leading to pathological upregulation of proliferation-associated gene expression programs.

In this study we further confirm the previous observation that *Pbx1* depletion compromises neuronal differentiation (17). Prima facie contradictory to this, our RNA-seq results indicate an upregulation of neuronal targets upon *Pbx1* depletion. However, upregulated neuronal targets, including *Gria1* and *Grin2b*, are primarily associated with glutamatergic rather than the *bona fide* V-SVZ GABAergic cell fate. This observation is consistent with a gatekeeping-like role for PBX1, preventing untimely differentiation to an adverse cell fate, disrupted by *Pbx1* kd. Interestingly, the effect of reduced neuronal differentiation was mimicked by knockdown of *Tcf3* or *Tcf4*, further strengthening a functional interaction of these three proteins. Cooperative DNA-binding of PBX1 and myogenic bHLH proteins had previously been investigated with electrophoretic mobility shift assays and *in vitro* generated proteins. This approach attributed complex formation to the PBX1-HD and a conserved tryptophan residue N-terminal to the bHLH domain (84). The association of PBX1 with bHLH class I factors that we report here differs because, first, the PBC domain not the HD of PBX1 is involved in binding and, second, the tryptophan that mediates binding of myogenic bHLH to PBX1 is not conserved in TCF3, TCF4 and TCF12. Indirect evidence for functional cooperativity of PBX1 and TCF3 in neurodevelopment exists. For instance, exploring the GRN of the TCF3 splice variant E47 in the mouse embryonic forebrain, Pfurr and colleagues detected no interaction with predicted binding partners of the NeuroD family and thus concluded that E47 might act in concert with previously unrecognized proteins (48). Based on the discoveries made here and in agreement with the results reported by Golonzhka *et al.* (14), PBX1 is a major candidate E47-interacting protein in forebrain neurogenesis.

While primarily focusing on proliferation and differentiation in the V-SVZ adult neural stem cell niche, our results may also have implications beyond neurogenesis. Translocation between chromosome 1q23 and chromosome 19p13 produces an oncogenic fusion protein consisting of the N-terminal transactivating domain of the human TCF3 homolog E2A and the C-terminal part of PBX1 (4). E2A is critically important for B-cell lineage commitment, evident from the fact that targeted deletion of the gene in mice blocks early stages of B-cell development (85) or from the observation that E2A overexpression can upregulate B-cell specific genes in non-lymphoid cells (86). The leukemogenic activity of the t(1;19) translocation event has been attributed to the strong transcriptional activity of the E2A-PBX1 fusion protein (85). In general, E2A-PBX1 is considered a neo-oncogene that brings together the transcriptional activation activity of E2A and the DNA-binding specificity of PBX1. Our observation that PBX1 and TCF3 physically interact and control the expression of shared target genes in adult neurogenic progenitor cells expands this view. In fact, we would argue that t(1;19) translocation does not simply join two independent TFs but rather stabilizes the association of two proteins, which are proven cooperators in an unrelated physiological context, into a fatal symbiosis. For transcriptionally active multi-TF complexes to form, the proteins involved must not only be able to assemble and interact in a sterically meaningful way, the regulatory elements of downstream genes must also possess the sequence motifs that can be bound and 'read' by these complexes. The evolution of physiologically active transcriptional complexes, hence, relies on the co-evolution of the protein-coding sequences of the TFs that constitute these complexes on one hand and of the regulatory sequences of their shared target genes on the other. Effective transcription is only possible when both elements come together: stoichiometrically correctly assembled TFs and matching sequence motifs in the promoters and enhancers of downstream genes. It therefore seems plausible to assume that a fusion protein consisting of two TFs that control progenitor cell proliferation as components of the same multiprotein complex will constitute a particularly aggressive oncogenic driver. Against this background, the forced heterodimer of TCF3 and PBX1 may freeze a proliferative cellular state and serve as turning point where malignant transformation of t(1;19) ALL cells occurs (Figure 7F).

In support of this hypothesis, we find that ectopic expression of *Pbx1* in Nalm6 cells, a B-cell precursor leukemia line that lacks t(1;19)-translocation and expresses *TCF3/E2A* but not *PBX1*, is sufficient to upregulate expression of *BLK* and *NOTCH3*. *BLK* has been associated with a multitude of hematopoietic cancers (87), and *NOTCH3* plays an oncogenic role in various neoplasms including leukemia (88). This observation provides additional support for the assumption that the ability of PBX1 and TCF3/E2A to function together during physiological conditions might be an important aspect of the leukemogenic effect of the E2A-PBX1 fusion protein. In fact, PBX1 in aNS and E2A-PBX1 in the human pre-B leukemia cell line 697 (57,89) share a number of similarities, including binding to the transcription machinery and preferential localization to chromatin marked by H3K27ac. In addition, we find that in adult neural stem- and progenitor cells PBX1 and TCF3 positively regulate the expression of genes involved in progenitor cell proliferation, while negatively regulating genes associated with cellular differentiation, a function that seems to be retained in the E2A-PBX1 fusion protein. However, clear differences with respect to the involvement of other TFs in the PBX1-containing complexes also exist in both settings. E2A-PBX1 in pre-B cells forms heteromeric complexes with RUNX1, a Runt-related TF linked to several types of leukemia (57) but not expressed in the V-SVZ neurogenic niche, while PBX1 and TCF3 associate with neural and neuronal cell fate determinants in adult neural stem- and progenitor cells.

In conclusion, using a broad spectrum of approaches our study establishes that the TALE-HD protein PBX1 and the class I bHLH TFs TCF3 and TCF4 function together under physiological conditions, an ability that appears to be conserved across physiological contexts and hijacked by the E2A-PBX1 oncogenic fusion protein in t(1;19) leukemia.

Data availability

- ChIP-seq: EBI ArrayExpress E-MTAB-13565
- RNA-seq: GSE248755
- ATAC-seq: GSE248756
- Mass spectrometry: PXD048222

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Sonja Thom for excellent technical assistance, Vijay Tiwari for sharing experimental protocols, Imke Wüllenweber for assistance with mass spectrometry sample workup, Ina Bartnik and Susanne tom Dieck for help with PLA, and Tanja Buhlmann and Ann-Christin Hau for help in setting up the spheroid assays. We are also grateful to Marta Losa, Katharina Imkeller and Marcel Schulz for bioinformatical insights and to Anjali Cremer for critically reading the manuscript. Finally, we thank all authors who have made their genome-wide datasets publicly available, serving as valuable resource in this project.

Author contributions: Vera Laub: Conceptualization, Formal analysis, Methodology, Visualization, Validation, Writing. Elisabeth Nan: Formal analysis, Methodology. Lena Elias: Methodology, Review and Editing. Jonathan Schupp, Jennifer Lun: Formal analysis of spatial transcriptomics. Karl H. Plate: spatial transcriptomics funding acquisition. Mette Bentsen, Mario Looso: Formal analysis of ATAC-seq and ChIP-seq. Stefan Günther: Methodology, Formal analysis of ATAC-seq and RNA-seq. Ian Donaldson: Formal analysis of ChIP-seq. Leona A Rusling, Julian D. Langer: LC/MS acquisition and analysis. Nicoletta Bobola: Formal analysis, Methodology. Dorothea Schulte: Conceptualization, Formal analysis, Methodology, Validation, Writing.

Funding

V.L. was supported by the Rosa Luxemburg Stiftung, Germany; E.N. was recipient of a FPF predoctoral scholarship from the Faculty of Medicine of Goethe University; research leading to the article was supported by the Deutsche Forschungsgemeinschaft, Germany [DFG SCHU 1218/4-1]; Stiftung Deutsche Krebsforschung, Germany [70114005] to D.S.; M.L. acknowledges support by DFG:EXC2026/1; J.H.L.

and J.S. were supported by the Uniscientia Foundation, Vaduz LI and LOEWE-Program FCI, respectively. Funding for open access charge: Open Access Publikationfonds of Goethe University DFG [SCHU 1218/4-1].

Conflict of interest statement

None declared.

References

1. Lambert,S.A., Jolma,A., Campitelli,L.F., Das,P.K., Yin,Y., Albu,M., Chen,X., Taipale,J., Hughes,T.R. and Weirauch,M.T. (2018) The Human transcription factors. *Cell*, **172**, 650–665.
2. Selleri,L., Zappavigna,V. and Ferretti,E. (2019) 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks. *Genes Dev.*, **33**, 258–275.
3. Kamps,M., Murm,C., Sun,X.-H. and Baitimore,D. (1990) A new homeobox.Gene contributes the DNA binding domain of the t(1;19) translocation protein in Pre-B ALL. *Cell*, **60**, 547–555.
4. Nourse,J., Mellentin,J.D., Galili,N., Wilkinson,J., Stanbridge,E., Smith,S.D. and Cleary,M.L. (1990) Chromosomal translocation t(1;19) results In synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, **60**, 535–545.
5. Roberts,V.J., Van Dijkb,M.A. and Murreb,C. (1995) Localization of pbxl transcripts in developing rat embryos. *Mech. Dev.*, **51**, 193–198.
6. Selleri,L., Depew,M.J., Jacobs,Y., Chanda,S.K., Tsang,K.Y., Cheah,K.S.E., Rubenstein,J.L.R., O'Gorman,S. and Cleary,M.L. (2001) Requirement for Pbx1 in skeletal patterning_Pbx1 KO. *Development*, **128**, 3543–3557.
7. Capellini,T.D., Zappavigna,V. and Selleri,L. (2011) Pbx homeodomain proteins: tALEnted regulators of limb patterning and outgrowth. *Dev. Dyn.*, **240**, 1063–1086.
8. Young,M., Selleri,L. and Capellini,T.D. (2019) Genetics of scapula and pelvis development: an evolutionary perspective. *Curr. Top. Dev. Biol.*, **132**, 311–349.
9. Berkes,C.A., Bergstrom,D.A., Penn,B.H., Seaver,K.J., Knoepfler,P.S. and Tapscott,S.J. (2004) Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell*, **14**, 465–477.
10. Stankunas,K., Shang,C., Twu,K.Y., Kao,S.C., Jenkins,N.A., Copeland,N.G., Sanyal,M., Selleri,L., Cleary,M.L. and Chang,C.P. (2008) Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ. Res.*, **103**, 702–709.
11. Amin,S., Donaldson,I.J., Zannino,D.A., Hensman,J., Rattray,M., Losa,M., Spitz,F., Ladam,F., Sagerström,C. and Bobola,N. (2015) Hoxa2 selectively enhances meis binding to change a branchial arch ground state. *Dev. Cell*, **32**, 265–277.
12. Ferretti,E., Li,B., Zewdu,R., Wells,V., Hebert,J.M., Karner,C., Anderson,M.J., Williams,T., Dixon,J., Dixon,M.J., *et al.* (2011) A conserved pbx-wnt-p63-Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. *Dev. Cell*, **21**, 627–641.
13. Ficara,F., Murphy,M.J., Lin,M. and Cleary,M.L. (2008) Pbx1 Regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell*, **2**, 484–496.
14. Golonzhka,O., Nord,A., Tang,P.L.F., Lindtner,S., Ypsilanti,A.R., Ferretti,E., Visel,A., Selleri,L. and Rubenstein,J.L.R. (2015) Pbx regulates patterning of the cerebral cortex in progenitors and postmitotic neurons. *Neuron*, **88**, 1192–1207.
15. Vitobello,A., Ferretti,E., Lampe,X., Vilain,N., Ducret,S., Ori,M., Spetz,J.F., Selleri,L. and Rijli,F.M. (2011) Hox and pbx factors control retinoic acid synthesis during hindbrain segmentation. *Dev. Cell*, **20**, 469–482.
16. Obernier,K. and Alvarez-Buylla,A. (2019) Neural stem cells: origin, heterogeneity and regulation in the adult mammalian brain. *Development (Cambridge)*, **146**, dev156059.
17. Grebbin,B.M., Hau,A.C., Groß,A., Anders-Maurer,M., Schramm,J., Koss,M., Wille,C., Mittelbronn,M., Selleri,L. and Schulte,D. (2016) PBX1 is required for adult subventricular zone neurogenesis. *Development (Cambridge)*, **143**, 2281–2291.
18. Remesal,L., Roger-Baynat,I., Chirivella,L., Maicas,M., Brocal-Ruiz,R., Pérez-Villalba,A., Cucarella,C., Casado,M. and Flames,N. (2020) PBX1 acts as terminal selector for olfactory bulb dopaminergic neurons. *Development (Cambridge)*, **147**, dev186841.
19. Penkov,D., SanMartín,D.M., Fernandez-Díaz,L.C., Rosselló,C.A., Torroja,C., Sánchez-Cabo,F., Warnatz,H.J., Sultan,M., Yaspo,M.L., Gabrieli,A., *et al.* (2013) Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with hox genes/proteins. *Cell Rep.*, **3**, 1321–1333.
20. Losa,M., Barozzi,I., Osterwalder,M., Hermosilla-Aguayo,V., Morabito,A., Chacón,B.H., Zarrineh,P., Girdziusaite,A., Benazet,J.D., Zhu,J., *et al.* (2023) A spatio-temporally constrained gene regulatory network directed by PBX1/2 acquires limb patterning specificity via HAND2. *Nat. Commun.*, **14**, 3993.
21. Chan,S.-K., Jaffe,L., Capovilla,M., Dotas,J. and Mann,R.S. (1994) The DNA binding specificity of ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, **76**, 603–615.
22. Chang,C.-P., Shen,W.-F., Rozenfeld,S., Lawrence,H.J., Largman,C. and Cleary,M.L. (1995) Pbx proteins display hexapeptode-dependent cooperative DNA binding with Hox. *Genes Dev.*, **9**, 663–674.
23. Bischof,L.J., Kagawa,N., Moskow,J.J., Takahashi,Y., Iwamatsu,A., Buchberg,A.M. and Waterman,M.R. (1998) Members of the Meis1 and Pbx homeodomain protein families cooperatively bind a cAMP-responsive sequence (CRS1) from bovine CYP17. *J. Biol. Chem.*, **273**, 7941–7948.
24. Hau,A.C., Mommaerts,E., Laub,V., Müller,T., Dittmar,G. and Schulte,D. (2021) Transcriptional cooperation of PBX1 and PAX6 in adult neural progenitor cells. *Sci. Rep.*, **11**, 21013.
25. Moen,M.J., Adams,H.H.H., Brandsma,J.H., Dekkers,D.H.W., Akinci,U., Karkampouna,S., Quevedo,M., Kockx,C.E.M., Ozgür,Z., Van Ijcken,W.F.J., *et al.* (2017) An interaction network of mental disorder proteins in neural stem cells. *Transl. Psychiatry*, **7**, e1082.
26. Basu,A., Mestres,I., Sahu,S.K., Tiwari,N., Khongwir,B., Baumgart,J., Singh,A., Calegari,F. and Tiwari,V.K. (2020) Phf21b imprints the spatiotemporal epigenetic switch essential for neural stem cell differentiation. *Genes Dev.*, **34**, 1190–1209.
27. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
28. Tyanova,S., Temu,T. and Cox,J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.*, **11**, 2301–2319.
29. Snel,B., Lehmann,G., Bork,P. and Huynen,M.A. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.*, **28**, 3442–3444.
30. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
31. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
32. 1000 Genome Project Data Processing Subgroup, Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

33. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biol.*, **9**, R137.
34. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
35. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
36. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
37. Castro-Mondragon,J.A., Riudavets-Puig,R., Rauluseviciute,I., Berhanu Lemma,R., Turchi,L., Blanc-Mathieu,R., Lucas,J., Boddie,P., Khan,A., Perez,N.M., *et al.* (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **50**, D165–D173.
38. Bentsen,M., Goymann,P., Schultheis,H., Klee,K., Petrova,A., Wiegandt,R., Fust,A., Preussner,J., Kuenne,C., Braun,T., *et al.* (2020) ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.*, **11**, 4267.
39. Laub,V., Devraj,K., Elias,L. and Schulte,D. (2023) Bioinformatics for wet-lab scientists: practical application in sequencing analysis. *BMC Genomics [Electronic Resource]*, **24**, 382.
40. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
41. Piwowar,H.A. (2010) A method to track dataset reuse in biomedicine: filtered GEO accession numbers in PubMed Central. *Proc. Natl. Acad. Sci. U.S.A.*, **47**, 1–2.
42. Zou,Z., Ohta,T., Miura,F. and Oki,S. (2022) ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and bisulfite-seq data. *Nucleic Acids Res.*, **50**, W175–W182.
43. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
44. Luu,P.L., Ong,P.T., Dinh,T.P. and Clark,S.J. (2020) Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. *NAR Genom Bioinform*, **2**, lqaa054.
45. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
46. Afgan,E., Nekrutenko,A., Grüning,B.A., Blankenberg,D., Goecks,J., Schatz,M.C., Ostrovsky,A.E., Mahmoud,A., Lonie,A.J., Syme,A., *et al.* (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.*, **50**, W345–W351.
47. Bailey,T.L. and MacHanick,P. (2012) Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.*, **40**, e128.
48. Pfurr,S., Chu,Y.H., Bohrer,C., Greulich,F., Beattie,R., Mammadzada,K., Hils,M., Arnold,S.J., Taylor,V., Schachtrup,K., *et al.* (2017) The E2A splice variant E47 regulates the differentiation of projection neurons via p57(KIP2) during cortical development. *Development (Cambridge)*, **144**, 3917–3931.
49. Whitington,T., Frith,M.C., Johnson,J. and Bailey,T.L. (2011) Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res.*, **39**, e98.
50. Yu,G., Wang,L.G. and He,Q.Y. (2015) ChIP seeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
51. Schultheis,H., Kuenne,C., Preussner,J., Wiegandt,R., Fust,A., Bentsen,M. and Looso,M. (2019) WIlsON: web-based Interactive Omics VisualizatioN. *Bioinformatics*, **35**, 1055–1057.
52. Mi,H., Ebert,D., Muruganujan,A., Mills,C., Albou,L.P., Mushayamaha,T. and Thomas,P.D. (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.*, **49**, D394–D403.
53. Supek,F., Bošnjak,M., Škunca,N. and Šmuc,T. (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**, e21800.
54. Chen,E.Y., Tan,C.M., Kou,Y., Duan,Q., Wang,Z., Meirelles,G.V., Clark,N.R. and Ma'ayan,A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf.*, **14**, 128.
55. Kuleshov,M.V., Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A., *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
56. Balwierz,P.J., Pachkov,M., Arnold,P., Gruber,A.J., Zavolan,M. and Van Nimwegen,E. (2014) ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.*, **24**, 869–884.
57. Pi,W.-C., Wang,J., Shimada,M., Lin,J.-W., Geng,H., Lee,Y.-L., Lu,R., Li,D., Wang,G.G., Roeder,R.G., *et al.* (2020) E2A-PBX1 functions as a coactivator for RUNX1 in acute lymphoblastic leukemia. *Blood*, **136**, 11–23.
58. Gearing,L.J., Cumming,H.E., Chapman,R., Finkel,A.M., Woodhouse,I.B., Luu,K., Gould,J.A., Forster,S.C. and Hertzog,P.J. (2019) CiIIder: a tool for predicting and analysing transcription factor binding sites. *PLoS One*, **14**, e0215495.
59. Reynolds,B.A. and Weiss,S. (1992) Generation of neurons and astrocytes from isolated cells of the adult mammalian Central nervous system. *Science*, **255**, 1707–1710.
60. Costa,M.R., Ortega,F., Brill,M.S., Beckervordersandforth,R., Petrone,C., Schroeder,T., Götz,M. and Berninger,B. (2011) Continuous live imaging of adult neural stem cell division and lineage progression in vitro. *Development*, **138**, 1057–1068.
61. Schramm,J. and Schulte,D. (2014) A fast and simple differentiation protocol to study the pro-neurogenic activity of soluble factors in neurospheres. *Neurosci. Lett.*, **562**, 69–74.
62. Ladam,F., Stanney,W., Donaldson,I.J., Yildiz,O., Bobola,N. and Sagerström,C.G. (2018) TALE factors use two distinct functional modes to control an essential zebrafish gene expression program. *eLife*, **7**, e36144.
63. Hwang,W.W., Salinas,R.D., Siu,J.J., Kelley,K.W., Delgado,R.N., Paredes,M.F., Alvarez-Buylla,A., Oldham,M.C. and Lim,D.A. (2014) Distinct and separable roles for EZH2 in neurogenic astroglia. *eLife*, **3**, e02439.
64. Marzinke,M.A. and Clagett-Dame,M. (2012) The all-trans retinoic acid (atRA)-regulated gene Calmin (Clmn) regulates cell cycle exit and neurite outgrowth in murine neuroblastoma (Neuro2a) cells. *Exp. Cell. Res.*, **318**, 85–93.
65. Sato,H., Hatakeyama,J., Iwasato,T., Araki,K., Yamamoto,N. and Shimamura,K. (2022) Thalamocortical axons control the cytoarchitecture of neocortical layers by area-specific supply of VGF. *eLife*, **11**, e67549.
66. Lee,D.-C., Hsu,Y.-C., Chung,Y.-F., Hsiao,C.-Y., Chen,S.-L., Chen,M.-S., Lin,H.-K. and Chiu,I.-M. (2009) Isolation of neural stem/progenitor cells by using EGF/FGF1 and FGF1B promoter-driven green fluorescence from embryonic and adult mouse brains. *Mol. Cell. Neurosci.*, **41**, 348–363.
67. Kong,H., Fan,Y., Xie,J., Ding,J., Sha,L., Shi,X., Sun,X. and Hu,G. (2008) AQP4 knockout impairs proliferation, migration and neuronal differentiation of adult neural stem cells. *J. Cell Sci.*, **121**, 4029–4036.
68. Tong,C.K., Chen,J., Cebrián-Silla,A., Mirzadeh,Z., Obernier,K., Guinto,C.D., Tecott,L.H., García-Verdugo,J.M., Kriegstein,A. and Alvarez-Buylla,A. (2014) Axonal control of the adult neural stem cell niche. *Cell Stem Cell*, **14**, 500–511.

69. Colak,D., Mori,T., Brill,M.S., Pfeifer,A., Falk,S., Deng,C., Monteiro,R., Mummery,C., Sommer,L. and Götz,M. (2008) Adult neurogenesis requires Smad4-mediated bone morphogenic protein signaling in stem cells. *J. Neurosci.*, **28**, 434–446.
70. Marqués-Torrejón,M.Á., Porlan,E., Banito,A., Gómez-Ibarlucea,E., Lopez-Contreras,A.J., Fernández-Capetillo,Ó., Vidal,A., Gil,J., Torres,J. and Fariñas,I. (2013) Cyclin-dependent kinase inhibitor p21 controls adult neural stem cell expansion by regulating Sox2 gene expression. *Cell Stem Cell*, **12**, 88–100.
71. Sokpor,G., Castro-Hernandez,R., Rosenbusch,J., Staiger,J.F. and Tuoc,T. (2018) ATP-dependent chromatin remodeling during cortical neurogenesis. *Front. Neurosci.*, **12**, 226.
72. Imayoshi,I., Sakamoto,M., Ohtsuka,T., Takao,K., Miyakawa,T., Yamaguchi,M., Mori,K., Ikeda,T., Itohara,S. and Kageyama,R. (2008) Roles of continuous neurogenesis in the structural and functional integrity of the adult forebrain. *Nat. Neurosci.*, **11**, 1153–1161.
73. Zywitza,V., Misios,A., Bunatyan,L., Willnow,T.E. and Rajewsky,N. (2018) Single-cell transcriptomics characterizes cell types in the subventricular zone and uncovers molecular defects impairing adult neurogenesis. *Cell Rep.*, **25**, 2457–2469.
74. Ravanpay,A.C. and Olson,J.M. (2008) E protein dosage influences brain development more than family member identity. *J. Neurosci. Res.*, **86**, 1472–1481.
75. Xie,X.P., Laks,D.R., Sun,D., Poran,A., Laughney,A.M., Wang,Z., Sam,J., Belenguer,G., Fariñas,I., Elemento,O., *et al.* (2020) High-resolution mouse subventricular zone stem-cell niche transcriptome reveals features of lineage, anatomy, and aging. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 31448–31458.
76. Quevedo,M., Meert,L., Dekker,M.R., Dekkers,D.H.W., Brandsma,J.H., van den Berg,D.L.C., Ozgür,Z., van IJcken,W.F.J., Demmers,J., Fornerod,M., *et al.* (2019) Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nat. Commun.*, **10**, 2669.
77. Omrani,M.R., Yaqubi,M. and Mohammadnia,A. (2018) Transcription factors in regulatory and protein subnetworks during generation of neural stem cells and neurons from direct reprogramming of non-fibroblastic cell sources. *Neuroscience*, **380**, 63–77.
78. Tiwari,N., Pataskar,A., Péron,S., Thakurela,S., Sahu,S.K., Figueres-Oñate,M., Marichal,N., López-Mascaraque,L., Tiwari,V.K. and Berninger,B. (2018) Stage-specific transcription factors drive astrogliogenesis by remodeling gene regulatory landscapes. *Cell Stem Cell*, **23**, 557–571.
79. Mateo,J.L., Van Den Berg,D.L.C., Haeussler,M., Drechsel,D., Gaber,Z.B., Castro,D.S., Robson,P., Crawford,G.E., Flicek,P., Ettwiller,L., *et al.* (2015) Characterization of the neural stem cell gene regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal. *Genome Res.*, **25**, 41–56.
80. Sock,E. and Wegner,M. (2021) Using the lineage determinants Olig2 and Sox10 to explore transcriptional regulation of oligodendrocyte development. *Dev. Neurobiol.*, **81**, 892–901.
81. Kageyama,R., Shimojo,H. and Ohtsuka,T. (2019) Dynamic control of neural stem cells by bHLH factors. *Neurosci. Res.*, **138**, 12–18.
82. Clark,B.S., Stein-O'Brien,G.L., Shiau,F., Cannon,G.H., Davis-Marcisak,E., Sherman,T., Santiago,C.P., Hoang,T.V., Rajaii,F., James-Esposito,R.E., *et al.* (2019) Single-cell RNA-seq analysis of retinal development identifies NFI factors as regulating mitotic exit and late-born cell specification. *Neuron*, **102**, 1111–1126.
83. Tchieu,J., Calder,E.L., Guttikonda,S.R., Gutzwiller,E.M., Aromolaran,K.A., Steinbeck,J.A., Goldstein,P.A. and Studer,L. (2019) NFIA is a gliogenic switch enabling rapid derivation of functional human astrocytes from pluripotent stem cells. *Nat. Biotechnol.*, **37**, 267–275.
84. Knoepfler,P.S., Bergstrom,D.A., Uetsuki,T., Dac-Korytko,I., Sun,Y.H., Wright,W.E., Tapscott,S.J. and Kamps,M.P. (1999) A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with Pbx-Meis1/Prep1. *Nucleic Acids Res.*, **27**, 3752–3761.
85. Bain,G., Robanus Maandag,E.C., Izon,D.J., Amsen,D., Kruisbeek,A., Weintraub,B.C., Krop,I., Schlissel,M.S., Feeney,A.J., Van Roon,M., *et al.* (1994) E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell, **79**, 885–892.
86. Choi,J.K., Shen1,C.-P., Radomska,H.S., Eckhardt,L.A. and Kadesch,T. (1996) E47 activates the ig-heavy chain and TdT loci in non-B cells. *EMBO J.*, **15**, 5014–5021.
87. Petersen,D.L., Berthelsen,J., Willerslev-Olsen,A., Fredholm,S., Dabelsteen,S., Bonefeld,C.M., Geisler,C. and Woetmann,A. (2017) A novel BLK-induced tumor model. *Tumor Biol.*, **39**, 101042831771419.
88. Aburjania,Z., Jang,S., Whitt,J., Jaskula-Stzul,R., Chen,H. and Rose,J.B. (2018) The role of Notch3 in cancer. *Oncologist*, **23**, 900–911.
89. Lee,Y.-L., Ito,K., Pi,W.-C., Lin,I.-H., Ch,C.-S., Malik,S., Cheng,I.-H., Chen,W.-Y. and Roeder,R.G. (2021) Mediator subunit MED1 is required for E2A-PBX1-mediated oncogenic transcription and leukemic cell growth. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e1922864118.

Received: December 22, 2023. **Revised:** August 22, 2024. **Editorial Decision:** September 16, 2024. **Accepted:** September 23, 2024