REGULAR ARTICLE

blood advances

The ParaHox gene *Cdx4* induces acute erythroid leukemia in mice

Silvia Thoene,[1,*] Tamoghna Mandal,[2,*] Naidu M. Vegi,[2] Leticia Quintanilla-Martinez,[3] Reinhild Rösler,[4] Sebastian Wiese,[4] Klaus H. Metzeler,[5] Tobias Herold,[5] Torsten Haferlach,[6] Konstanze Döhner,[7] Hartmut Döhner,[7] Luisa Schwarzmüller,[8] Ursula Klingmüller,[8] Christian Buske,[2] Vijay P. S. Rawat,[2,†] and Michaela Feuring-Buske[2,7,†]

[1]Institute of Clinical Chemistry and Pathobiochemistry, Technical University of Munich, Munich, Germany; [2]Institute of Experimental Cancer Research, Comprehensive Cancer Center and University Hospital of Ulm, Ulm, Germany; [3]Institute of Pathology, University of Tübingen, Tübingen, Germany; [4]Core Unit Mass Spectrometry and Proteomics, University Hospital Ulm, Ulm, Germany; [5]Laboratory for Leukemia Diagnostics, Department of Medicine III, University Hospital, Ludwig-Maximilians University, Munich, Munich, Germany; [6]Munich Leukemia Laboratory, Munich, Germany; [7]Department of Internal Medicine III, University Hospital Ulm, Ulm, Germany; and [8]Division of Systems Biology of Signal Transduction, German Cancer Research Center, Heidelberg, Germany

Key Points

- CDX4 induces AEL in mice and suppresses expression of genes associated with erythroid differentiation.
- Mutations detectable in the Cdx4-induced AEL model occur in genes reported mutated in patients with AEL.

Acute erythroid leukemia (AEL) is a rare and aggressive form of acute leukemia, the biology of which remains poorly understood. Here we demonstrate that the ParaHox gene *CDX4* is expressed in patients with acute erythroid leukemia, and that aberrant expression of Cdx4 induced homogenously a transplantable acute erythroid leukemia in mice. Gene expression analyses demonstrated upregulation of genes involved in stemness and leukemogenesis, with parallel downregulation of target genes of Gata1 and Gata2 responsible for erythroid differentiation. Cdx4 induced a proteomic profile that overlapped with a cluster of proteins previously defined to represent the most primitive human erythroid progenitors. Whole-exome sequencing of diseased mice identified recurrent mutations significantly enriched for transcription factors involved in erythroid lineage specification, as well as TP53 target genes partly identical to the ones reported in patients with AEL. In summary, our data indicate that Cdx4 is able to induce stemness and inhibit terminal erythroid differentiation, leading to the development of AEL in association with co-occurring mutations.

Introduction

Acute erythroleukemia (AEL) is a rare subtype of acute myeloid leukemia (AML) that accounts for less than 5% of all de novo AML cases. Previously, this subtype was characterized by the presence of a predominant erythroid population, which, in the case of AML M6a, was mixed with myeloid blasts. In contrast, in pure erythroid leukemia (AML M6b), the leukemic clone exclusively consisted of erythroblasts. The 2016 revision of the World Health Organization classification merged the M6a into a hybrid subtype of myelodysplasia and AML (MDS or AML not otherwise specified [NOS], nonerythroid subtype), based on the number of blasts present in the bone marrow. Only M6b remained as a subtype of AML NOS, acute erythroid leukemia, pure erythroid type if more than 30% proerythroblasts are present.[1,2] There have been several efforts to characterize AEL at a molecular level[3,4]: Bacher et al[4] described 77 AEL and 7 pure erythroid leukemia cases and described an association with aberrant and unfavorable karyotypes including *TP53* alterations, as well as recurrent mutations in the *NPM1* and *FLT3* gene, although at lower frequency compared with the overall AML cohort. Just recently, a large comprehensive genomic analysis of 159 childhood and adult AEL cases confirmed genomic complexity of this AML subtype, but succeeded into grouping AEL into 5 age-related subgroups characterized by distinct expression profiles. Furthermore, this report demonstrated druggable mutations in signaling pathways in nearly every second patient with AEL, opening an avenue for developing novel targeted approaches in this disease.[5] Despite these advances and the identification of driver mutations in AEL,

Submitted 29 July 2019; accepted 4 October 2019. DOI 10.1182/bloodadvances.2019000761.

*S.T. and T.M. contributed equally to the manuscript.

†V.P.S.R. and M.F.-B. contributed equally to the manuscript.

Send data sharing requests to Michaela Feuring-Buske (michaela.feuring-buske@uni-ulm.de).

The full-text version of this article contains a data supplement.

the underlying biology of AEL is still not precisely defined. This is also because there are only few models recapitulating human AEL. One of the models of murine erythroleukemia, the Friend-virus-induced erythroleukemia described nearly 30 years ago, is based on 2 retroviruses, the replication-defective spleen focus-forming virus and the replication-competent Friend murine leukemia virus. Friend virus induces an acute erythroleukemia that proceeds through a characteristic 2-stage progression, triggered by spleen focus-forming virus proviral insertional activation of the *Spi1/Pu.1* gene and Hedgehog-dependent signaling in a self-renewing population of stress erythroid progenitors in the spleen .[6,7] On the basis of the observation that the *spi-1* gene was a target for insertional mutagenesis with subsequent overexpression of Pu.1 in the latter model, Pu.1 transgenic mice were generated that also are developing erythroleukemia, mainly by blocking differentiation at the level of proerythroblasts.[8]

Here, we report that constitutive expression of the caudal-related homeobox gene *Cdx4* robustly induces AEL in mice, shedding light on the role of homeobox genes in the pathobiology of erythroid leukemia.

Materials and methods

Patient samples, cell lines, and mouse experiments

Mononuclear cells were isolated from diagnostic bone marrow of 8 patients with AEL. As a control, sorted subpopulations of 6 cord blood (CB) samples were analyzed. Cytomorphology, cytochemistry, cytogenetics, and molecular genetics were applied in all cases, as described. Cases were classified according to the French-American-British criteria and World Health Organization classification.[1,2] The study was approved by the ethics committees of all participating institutions, and informed consent was obtained from all patients before they entered the study in accordance with the Declaration of Helsinki (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/). Mice experiments were performed in compliance with the German Law for Welfare of Laboratory Animals and were approved by the Regierungspräsidium Oberbayern (AZ 55.2-1-54-2531-129-06) and the Regierungspräsidium Tübingen, Germany (No. 997).

Microarray analyses

Affymetrix gene expression microarray data from 548 newly diagnosed patients with AML were analyzed as reported previously.[9] CDX4 expression levels (probe set GC0XP072583_at) were compared between the AML M6 subset (n = 22) and all remaining patients with known FAB subtype (n = 538), using the Wilcoxon rank sum test.

qRT-PCR and linker-mediated PCR

Expression of *CDX4* was assayed by TaqMan real-time quantitative polymerase chain reaction (qRT-PCR) in sorted subfractions of human CB and unfractioned primary AEL patient samples. Expression analyses were performed by predesigned gene expression assays purchased from Applied Biosystems (Foster City, CA; assay ID CDX4 Hs00193194_m1). For normalization, the TATA binding protein gene was used (Hs4333769F). A minimum of 20 ng input of cDNA was used per 20 μL PCR reaction volume. In murine tissues, expression of Hox genes was quantified as described earlier, using *Gapdh* as housekeeping gene (4352932E). Fold expression was determined by the $\Delta\Delta C_t$ method: *Cdx2*, Mm01212280_m1; *Cdx4*, Mm00432452_m1; *Hoxa5*, Mm00439362_m1; *Hoxa7*, Mm00657963_m1; *Hoxa9*, Mm00439364_m1; *Hoxa10*, Mm00433966_m1; *Hoxb3*, Mm00650701_m1; *Hoxb6*, Mm00433970_m1; *Hoxb8*, Mm00439368_m1; and Hoxb9, Mm01700220_m1. For linker-mediated PCR, integrated long-terminal repeats and flanking genomic sequences were amplified and then isolated, using a modification of the bubble linker-mediated polymerase chain reaction strategy as previously described.[10,11]

Retroviruses and plasmids

Murine stem cell virus–based retroviral vectors were used for overexpression of Cdx4, Cdx2, and MIY.[12,13] The Cdx4 cDNA was provided by R. K. Humphries (Terry Fox Laboratory, Vancouver, BC, Canada) and subcloned into the retroviral pMSCV-IRES-YFP vector, as previously described.[10] As control, an empty vector (MSCV vector carrying only the IRES and YFP-cassette) was used (vector control).

Transduction and BM transplantation

Stable packaging cell lines were generated for the different constructs and used for BM experiments, as reported previously.[12,14] Five-fluorouracil BM from (C57Bl/6Ly-Pep3b × C3H/HeJ) F1 (PepC3) mice was transduced with empty vector (control) or Cdx4 in 5 independent experiments. Directly after the end of transduction, 1.4×10^4 to 1.1×10^6 total BM cells (average, 16.25% ± 5.9% standard error of the mean enhanced green fluorescent protein positivity) were transplanted into primary mice lethally irradiated with 800 to 850 cGy together with 5×10^5 to 1×10^6 untransduced BM helper cells of a normal syngenic animal. Lethally irradiated secondary recipients were injected with 10^6 BM cells from a diseased primary mouse together with 5×10^5 to 1×10^6 untransduced BM helper cells of a normal syngenic animal.

CFC assay and cytospin

Hematopoietic colony-forming cell (CFC) and replating assays were performed using methylcellulose supplemented with murine cytokines (MethoCult GF M3434, Stem Cell Technologies, Cologne, Germany), as previously described.[14] Seven days after setting up the CFC, number and morphology of the colonies were assessed.[14]

Proliferation assay

For determination of the proliferative potential of BM cells in vitro, 2.5×10^4 transduced BM cells were cultured in the DMEM medium containing 15% fetal bovine serum and 6 ng/mL murine interleukin–3, 10 ng/mL mIL-6, 100 ng/mL murine SCF. Cells were counted every week; expanded and immunophenotypic analyses were performed as indicated.

CFU-S assay

Colony-forming unit assay (CFU) spleen assay was performed by BM transplantation as previously published.[11]

Flow cytometric analysis

Cells were stained according to standard protocols.[10] Antibodies used for flow cytometric staining: Ter119APC (BioLegend), B220 and Ter119 labeled with eFluor450 (both eBioscience, Frankfurt, Germany), Mac-1 labeled with qDot605 (BioLegend). Antibodies labeled with Alexa Fluor 700 were used against the epitope of c-Kit (eBioscience), Gr-1 was labeled with APC-Cy7 (both BioLegend), Sca-1 with PE-Cy5.5 (Caltag), and CD71 with PE-Cy7 (Becton Dickinson and BioLegend, respectively). Cells were analyzed using a FACS Fortressa LSR II (Becton Dickinson, Heidelberg, Germany).

Sequencing

PCR products and plasmids were commercially sequenced by Sequiserve (Vaterstetten, Germany).

RNA-Seq analysis

RNA-Seq analyses were performed on 5-fluorouracil–stimulated mouse BM transduced with either the empty vector control or the Cdx4 vector. RNA was extracted after sorting of GFP or YFP transduced cells 48 hours after the end of transduction. The analysis was performed in duplicates from libraries prepared by using TruSeq RNA, sample preparation kit version 2. The samples were loaded on the flow cell, using the cBot from Illumina and run on a HiSeq2000 as a Paired End 100 base pair read. The raw double-end reads were adapter trimmed and quality filtered (phred score of > 20), using the cut adapt wrapper trim galore. RNA-Seq data were analyzed using Basepair software (https://www.basepairtech.com/), with a pipeline that included the following steps: reads were aligned to the transcriptome derived from University of California–Santa Cruz genome assembly mm10, using STAR[15] with default parameters. Read counts for each transcript were measured using feature counts.[16] Differentially expressed genes were determined using DESeq2[17] and cutoff parameters of read count higher than 10, $P < .05$ and P adjusted (false discovery rate, corrected for multiple hypotheses testing) $<.1$, were used for pathway and target gene analysis.

Liquid chromatography–MS-based proteomics analysis

Sample preparation and mass spectroscopy (MS) analysis: Culture samples were dissolved in lysis buffer (7 M Urea, 2 M Thiourea, 30 mM Tris at pH 8.5). A total of 50 μg protein per sample was loaded on sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and after staining, each lane was cut into 10 slices and further processed and analyzed via liquid chromatography-MS/MS, as described earlier.[18]

A database search was performed using MaxQuant version 1.6.3.4 (https://www.maxquant.org/).[19] Employing the build-in Andromeda search engine,[20,21] MS/MS spectra were correlated with the UniProt mouse reference proteome set (https://www.uniprot.org/) for peptide identification. Carbamidomethylated cysteine was considered as a fixed modification along with oxidation (M), and acetylated protein N termini as variable modifications. False discovery rates were set on both peptide and protein level to 0.01. Mouse Genome Informatics homology database[21,22] was used to assign corresponding human homologs to the obtained mouse data. Copy numbers of proteins per cell were estimated by proteomic ruler plugin for Perseus Software 1.6.2.3 (https://www.maxquant.org/).[19,23]

Whole-exome sequencing

Whole-exome sequencing and statistical analyses were performed on BM of leukemic mice by StarSeq (Mainz, Germany). The sequencing reads were analyzed using TREVA pipeline, using murine genome alignment, mm9, and genome analysis toolkit, gatk3. The annotations for the variants were annotated according to SNP databases, dbSNP129, and μ tect.

Flow cytometry-activated cell sorting of human bone marrow populations

Hematopoietic stem cells, common lymphoid progenitors, common myeloid progenitors, granulocyte-macrophage progenitors, and myeloid-erythroid progenitors from human CB (Lonza, Cologne, Germany) were isolated by density gradient centrifugation. Cells were preincubated with human Fc receptor binding inhibitor (14-9161, eBioscience) and stained with propidium iodide and antibodies against CD235a, CD41a, and CD71, as described earlier. Cell sorting was conducted on a BD Bioscience Aria.

Statistics

All pairwise statistical tests were performed as nonparametric tests, using Wilcoxon and Mann-Whitney U tests, except for gene expression TaqMan assays, for which column-wise multiple comparisons were performed with the Holm-Šídák method 1-way analysis of variance. Figure legends specify the test and P value used. Analyses were performed with the GraphPad Prism v.6 software (La Jolla, CA). FlowJo (Tree Star Inc, Ashland, OR) was applied for analysis of fluorescence-activated cell sorting plots.

Results

Cdx4 is expressed in patients with AEL

Using publicly available cDNA microarray analyses, CDX4 was expressed in patients with AEL with significantly higher expression levels compared with other AML subtypes ($P = .002$)[9] (Figure 1A). CDX4 expression was confirmed by qRT-PCR analysis in 3 of 8 patients with AEL, whereas $CD34^+$ BM cells as well as CB subpopulations were negative for *CDX4* expression, with the exception of 1 myeloid-erythroid progenitor sample from CB (Table 1; Figure 1B). Similarly, published microarray data showed CDX4 expression in AEL cell lines, such as K562, LAMA84, and OCIAML2 (https://portals.broadinstitute.org/ccle).

Constitutive expression of Cdx4 increases erythroid cell formation in vitro

To test the functional relevance of CDX4 expression in AEL development, we constitutively expressed Cdx4 in murine stem and progenitors (HSPCs) and analyzed the effect on erythroid development. Overexpression of Cdx4 in HSPCs induced a significant 643-fold increase in cell formation in liquid expansion cultures compared with the control at day 14 (Figure 1C). This was accompanied by a doubling of the yield of erythroid $Ter119^+$ cells at day 7, with further expansion to day 21 compared with the control (Figure 1D). Flow cytometry data of primary CFC demonstrated a significant increase of the percentage of $Ter119^+$ cells (6.2-fold increase; $P = .020$), c-Kit^-/Sca-1^+ cells ($P = .023$), and $Gr1^-Mac1^+$ mature macrophages compared with the vector control ($P = .015$). In contrast, overexpression of CDX4 led to a decrease in the number of $Gr1^+Mac1^+$ myeloid cells (data not shown). Furthermore, Cdx4 conferred serial replating in the CFC assay with a 2.7-fold higher yield of $Ter119^+$ erythroid cells and persisting generation of erythroid cells at replating (Figure 1E-F). Furthermore, Cdx4 increased the frequency of spleen colonies (CFU-S) 86-fold and tripled the percentage of immature $GFP^+CD71^+Ter119^-$ erythroid cells ($P < .02$) with a parallel decrease of more mature $GFP^+CD71^+Ter119^+$ in the spleen colonies compared with the control, indicating a partial block in erythroid differentiation (Figure 2A-B).

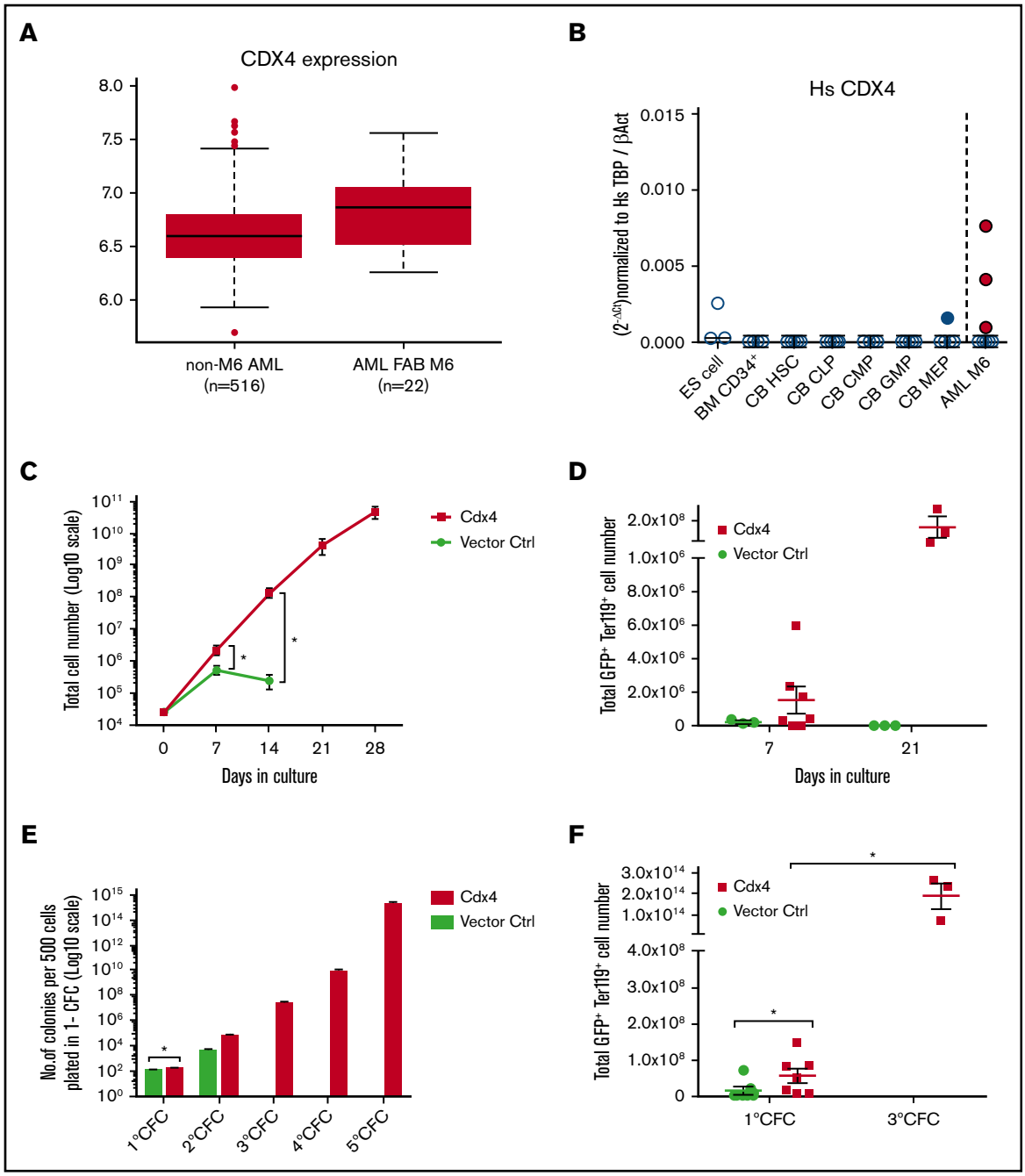

Figure 1. CDX4 is expressed in patients with AML M6 and induces an expansion of erythroid cells in vitro. (A) Microarray analyses showing expression of *CDX4* in samples from patients with FAB M6 compared with non-M6 AML.[9] (B) Fold expression measured by qRT-PCR of *CDX4* in embryonic stem cells as a positive control, $CD34^+$ human bone marrow cells, CB-derived hematopoietic stem cells (HSC), common lymphoid progenitors (CLP), common myeloid progenitor (CMP), granulocyte-macrophage progenitors (GMP), myeloid-erythroid progenitors compared with and human AEL patient samples (n = 8). Fold values were obtained through normalization to expression of TATA binding protein/β-actin. (C) Total cell number in liquid culture expansion assay of HSPCs transduced with Cdx4 (n = 6) or vector control (n = 7). Fluorescence-activated cell sorter-purified cells were plated 72 hours after retroviral infection, and viable cells were counted every 7 days. After day 14, all cells in the control group showed mast cell characteristics and were counted as 0. Total cell number difference was significant on day 7 and day 14; $*P < .05$ (Wilcoxon test, $P = .0313$). (D) Total number of $Ter119^+$ cells at day 7 and day 21 of proliferation assay (n = 3-7). (E) CFC assay for HSPCs transduced with vector control and Cdx4 plated 72 hours after retroviral infection. The primary CFCs were euthanized and replated 4 times every 7 to 9 days. Total number of colonies generated from 500 cells plated in primary CFCs (n = 7) derived from BM cells transduced with the vector control (ctrl) or Cdx4 ($*P < .05$; Wilcoxon test, $P = .027$). (F) Total number of GFP^+ $Ter119^+$ cells generated in primary and tertiary CFCs (n = 3). Total GFP^+ $Ter119^+$ cell number in Cdx4 primary CFC compared with vector ctrl, $*P < .05$ (Wilcoxon test, $P = .0313$). Total increase in GFP^+ $Ter119^+$ cell number in Cdx4 primary CFC compared with tertiary CFC is indicated. $*P < .05$ (Mann-Whitney *U* test, $P = .017$).

Table 1. Patient characteristics

Diagnosis	Sex	Age, y	Karyotype	Other relevant mutations
AML M6a	M	77	46,XY,del(5)(q22q34),+8,dic(15;17)(p11;p11),der(20;21)(p10;q10),+der(20;21)(p10;q10) [16]	DNMT3A, TP53
AML M6a	M	48	46,XY,del(11)(p11p14) [11] 46,XY [9]	
AML M6	M	74	48,XY,+8,+8 [4] 48,XY,der(6)t(6;6)(p25;q12),+8,+8 [15] 46,XY [1]	DNMT3A, RUNX1, TP53
AML M6	F	81	46,XX,t(4;10)(q13;p12) [3] 45,XX,t(4;10)(q13;p12),del(5)(q31q35),-7 [3] 46,XX [4]	TET2, TP53
AML M6	M	71	47,XY,+8 [10] 46,XY [10]	ASXL1
AML M6	F	42	46,XX[20]	FLT3 TKD, NPM1
AML M6	F	49	46,XX[20]	NPM1
AML M6	F	56	46,XX[20]	NPM1

F, female; M, male.

CDX4 induces AEL in mice

All mice transplanted with *Cdx4*-transduced BM HSPCs succumbed to transplantable AEL with a median latency of 309 days posttransplantation (range, 154-504 days; Figure 2C). AEL was diagnosed according to the Bethesda criteria for nonlymphoid neoplasms by histopathology.[24] The spleens of all mice showed diffuse infiltration by erythroid precursors, which were strongly positive for CD71 and partly for Ter119. All secondary recipients showed multiorgan infiltration into liver, lung, and kidney (Figure 3A-F; supplemental Figure 1A-I; Table 2). Sequencing of retroviral integration sites in the Cdx4 leukemic mice did not identify recurrent integrations into genes listed in the Retrovirus Integration Database (https://rid.ncifcrf.gov), indicating no major role of retroviral insertional mutagenesis in the development of the erythroid leukemias (supplemental Table 1).

The erythroid phenotype is shaped by differential *Hox* gene expression

Cdx2 is highly related to Cdx4 and is known to induce AML but not AEL, in the same experimental mouse model.[12,13] We analyzed expression of *Hox* genes known to be involved in normal erythropoiesis and known to be downstream targets of *Cdx2* and *Cdx4*.[25] Both ParaHox genes upregulated *Hox* gene expression compared with the control in primary CFC assays (Figure 4A). These findings were supported by the *Hox* expression profile in leukemic mice suffering from CDX2-induced AML vs CDX4-induced AEL. Of note, *Hoxb3* and *Hoxb4*, with the latter previously shown to induce erythroid colony formation in humans,[26] were significantly higher expressed in Cdx4-induced AEL compared with Cdx2-induced AML, pointing to the possibility that the erythroid phenotype is shaped by differential *Hox* gene expression (Figure 4B).

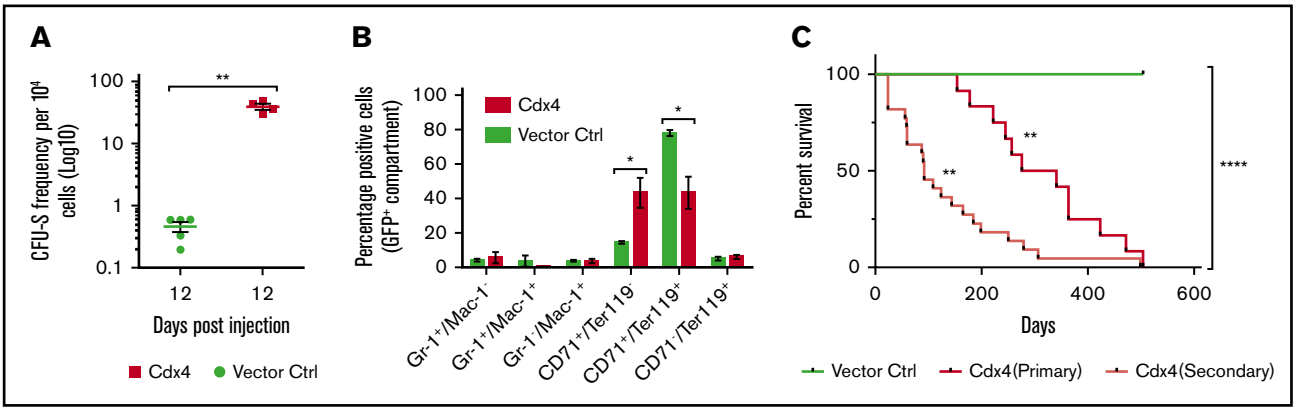

Figure 2. Cdx4 induces an expansion of erythroid cells in vivo and an AEL in transplanted mice. (A) Spleen colony formation (CFU-S) was compared between Cdx4 (n = 4) and control vector (n = 5) transduced HSPCs. Transduced cells were sorted 72 hours after retroviral infection and injected in the tail vein of lethally irradiated mice 12 days before euthanasia. CFU-S frequency was calculated per 10^4 cells. Data are represented in a Log10 scale. Values are shown as mean ± standard error of the mean. Significance was calculated by the Mann-Whitney *U* test (**$P < .01$). (B) Spleen cells from panel A were stained and analyzed by fluorescence-activated cell sorting for the expression of myeloid and erythroid markers. The bar graph shows the mean percentage (± standard error of the mean) myeloid marker (Gr-1$^+$/Mac-1$^+$, Gr-1$^+$/Mac-1$^-$, Gr-1$^-$/Mac-1$^+$) and erythroid marker (CD71$^+$/Ter119$^-$, CD71$^+$/Ter119$^+$, and CD71$^-$/Ter119$^+$)-expressing cells and refers to the GFP$^+$ compartment. Significance was calculated by the Mann-Whitney *U* test (*$P = .0159$). (C) Kaplan-Meier survival curves of primary (n = 12) and secondary (n = 22) mice transplanted with 5FU-stimulated BM cells expressing Cdx4 or the vector control. Mantel-Cox log-rank test was performed on mice injected with control vector vs *Cdx4* (primary recipient mice; **$P = .0045$) and *Cdx4* (secondary recipient mice; **$P = .0019$), respectively. Mantel-Cox log-rank test on all 3 survival curves was also found significant (****$P < .0001$).

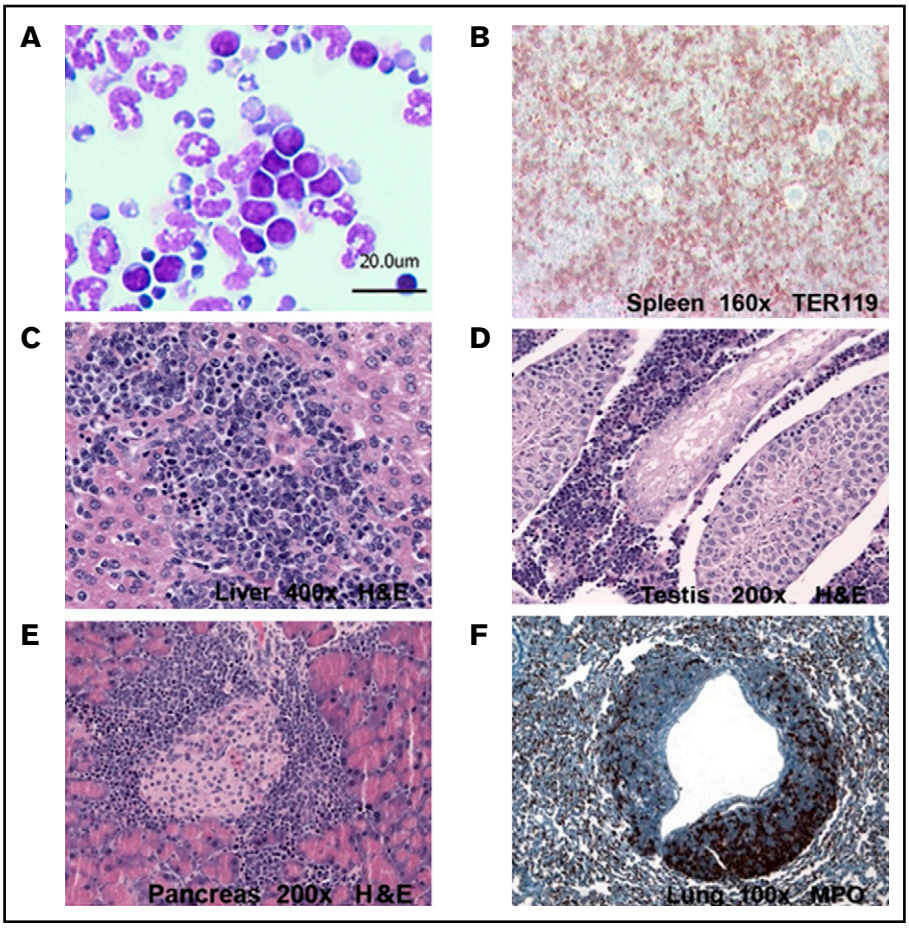

Figure 3. The histopathological analysis of a representative *Cdx4* mouse showing the presence of erythroblasts in different mouse organs. (A) Wright-Giemsa staining of peripheral blood. (B) Ter119 staining of the spleen. Hematoxylin and eosin (H&E) staining of the liver (C), testis (D), and pancreas (E). (F) Myeloperoxidase (MPO) staining of the lung.

Cdx4 promotes expression of genes involved in stemness and oncogenic pathways and suppresses genes associated with erythroid differentiation

When RNA-Seq was performed in HSPCs 48 hours after transduction, 651 genes were differentially expressed (DEGs) between *Cdx4* and the control, with most of the genes being upregulated (Figure 5A-B; supplemental Table 2, sheet 1). Among the upregulated genes, we detected several genes involved in AML such as *Meg3, Dlk1*, and *Hox* genes, whereas genes associated with erythroid development such as *Smad3*, *Gata1*, and its targets such as *Ets1* were downregulated. Of note, DEGs downregulated by *Cdx4* were also enriched for target genes of *Gata1*, *Gata2*, and *Spi1*, which are key players in the process of late erythroid differentiation and were shown to be downregulated in AELs[5] (Figure 5C-D; supplemental Table 2, sheets 2-3). In contrast, *Cdx4*-upregulated DEGs were enriched for oncogenic signatures and pathways in cancer, as well as in embryonic stem cell signature, using the Enrichr platform, the MSigDB Oncogenic Signatures, and the murine KEGG 2019 pathway analysis (Figure 5E-F; supplemental Table 2, sheets 4-7). We analyzed the BM from diseased mice for the expression of *GATA1* and *DLK1*, the latter among the top 3 DEGs after 48 hours of gene transduction, and saw the same changes (up for *DLK1* and down for *GATA1*; data not shown). Interestingly, human AML LSC signature genes were enriched in our upregulated genes, using gene set enrichment analysis (Figure 6A).[27] Of note, more than 70% of DEGs in our model overlapped with protein coding genes identified in human AELs (Figure 6B; supplemental Table 2, sheet 8).[5]

The protein profile in Cdx4 overexpressing HSPCs mostly resembles the protein signature of immature erythroid stages

Furthermore, proteins expressed after 48 hours of *Cdx4* gene exposure in murine HSPCs were compared with protein clusters of human erythroid progenitors and mature erythroid subpopulations: 75.6% of the cluster 1 proteins and 72.8% of the cluster 2 proteins representing the most primitive erythroid progenitors overlapped with Cdx4-induced proteins, indicating that the protein profile in *Cdx4* overexpressing HSPCs resembled the protein signature of immature erythroid stages (Figure 7A; supplemental Table 2, sheets 9-10) [28]. Most differentially expressed proteins in clusters 1 and 2, defined by Gautier et al,[28] showed a higher expression in erythroid progenitor cells (BFU-E and CFU-E as well as proerythroblast stages) and decreased during differentiation toward polychromatophilic and orthochromatic erythroblasts. Functional annotation analysis of these 2 clusters showed a significant enrichment of proteins related to general cellular processes, such as RNA processing, translation, metabolism, intracellular transport, proteasome degradation, and DNA replication and repair, and lacked

Table 2. Hematological parameters of experimental mice

Construct	Mouse	Survival, d	RBC, $\times 10^{12}$/L	WBC, $\times 10^{9}$/L	Spleen weight, mg	Blasts BM, %	Blasts spleen, %	Blasts PB, %	Lymphoid/ myeloid ratio in PB
GFP	1	90*	6.0	4.5	150	0	0	0	5
GFP	2	90*	4.8	3.2	200	0	0	0	2
GFP	3	90*	5.0	3.6	200	0	0	0	2
Cdx4	4	423	1.5	2.3	164	57	>95	56	<0.01
Cdx4	5	178	0.7	2.6	449	>95	>95	61	0.02
Cdx4	6	222	1.3†	2.5†	n.d.	n.d.	n.d.	52†	<0.01†
Cdx4	7	504	0.6	1.2	401	35	62	52	<0.01
Cdx4	8	364	12	9.1	129	20	35	52	0.8
Cdx4	9	364	12	8.9	134	11	39	29	0.9
Cdx4	10	341	0.9†	85†	n.d.	n.d.	n.d.	63†	0.02†
Cdx4	11	276	2.7†	3.4†	n.d.	n.d.	n.d.	50†	0.1†
Cdx4	12	257	3.4	2.9	505	27	39	2	<0.01
Cdx4	13	472	6.1	5.5	345	65	75	25	<0.01
Cdx4	14	245	10†	4.0†	398	n.d.	n.d.	66†	0.1†
Cdx4	15	154	8.5	3.3	296	43	55	24	<0.01

Healthy control mice were euthanized at day 90 post transplantation.
L/M, lymphoid/myeloid ratio in PB; n.d., not determined; RBC, red blood cell; WBC, white blood cell.
*Control mice were euthanized for analysis.
†As spleen, BM, and PB of some mice could not be analyzed because of the sudden death of the animals; the results of the respective last PB analysis are indicated.

genes associated with erythroid differentiation such as Hb and band 3[28].

Genes mutated in Cdx4-induced murine AEL are also reported in patients with AEL

As AELs developed after long latency in this model, we performed whole-exome sequencing in 3 diseased vs nontreated control mice, generated in independent gene transfer experiments, to identify potential cooperating mutations: there were 34 mutated genes detected in the leukemic Cdx4 mice, which were also reported in patients with AEL.[5] These commonly mutated genes were significantly enriched for transcription factors involved in erythroid lineage specification such as GATA 1, GATA 2, and GATA 4, as well as SMAD3 and SMAD4. Furthermore, mutated genes comprised TP53 target genes, in line with observations in patients with AEL[5] (supplemental Table 2, sheet 11). Three of 34 mutated genes were recurrently found in the AEL data set. From these 3 genes, 2 genes were mutated in all 3 leukemic mice. Both genes (Tns3 and Ralgapa2) carried mutations in the same domain as reported in patients with AEL (Figure 7B; supplemental Table 2, sheet 12).

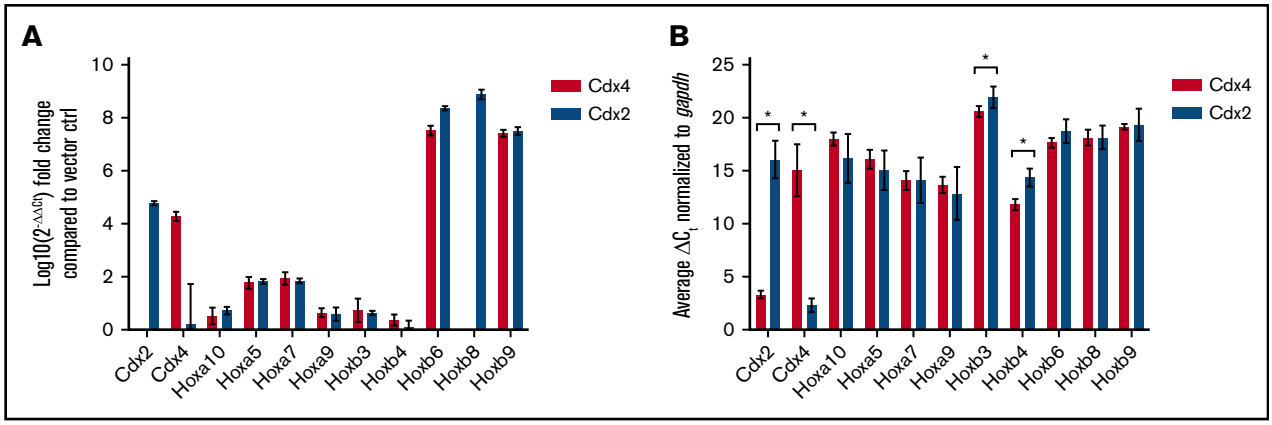

Figure 4. The erythroid phenotype is shaped by differential Hox gene expression. (A) qRT-PCR analysis of members of *HoxA* and *B* cluster genes along with *Cdx* genes in *Cdx2* and *Cdx4* transduced cells of primary CFC. ΔC_t values were obtained by normalizing to *Gapdh*, and fold expression compared with empty vector was calculated. The diagram shows average expression levels of 3 independent experiments ± SD. Vector control transduced cells negative for *Cdx2* expression; *significant difference between vector control, Cdx2 and Cdx4, respectively. (B) Expression of *Hox* cluster genes in diseased *Cdx4* transplanted mice. Mice that died of an acute leukemia after transplantation with Cdx2 or Cdx4 transduced BM were analyzed for their expression of *Hox* genes by TaqMan qRT-PCR. Multiple tests using Holm-Sidak method were performed on each gene. *Cdx2, Cdx4, Hoxb3* and *Hoxb4* were significantly differentially expressed. *Significant difference between *Cdx2* and *Cdx4*, respectively.

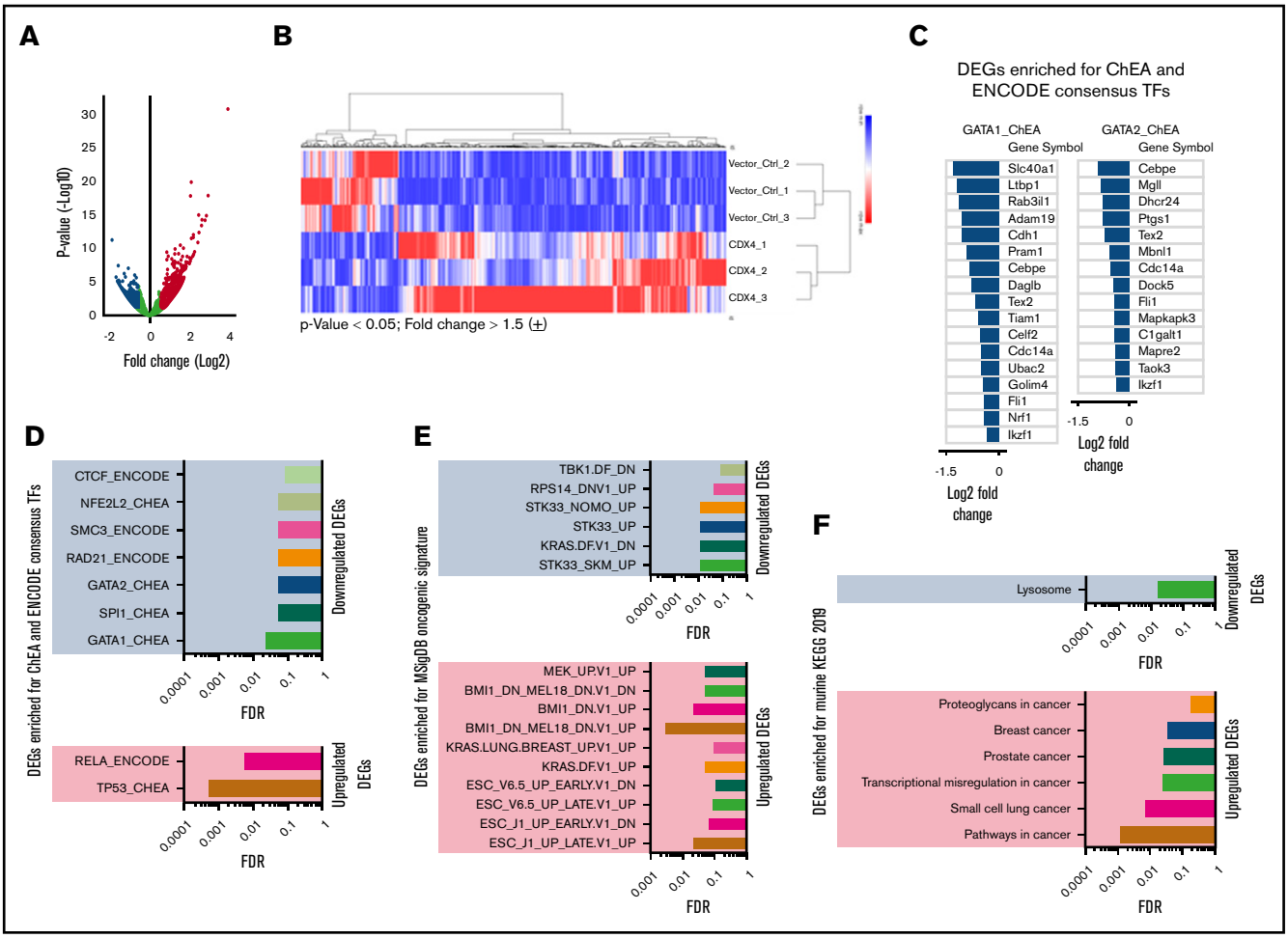

Figure 5. Cdx4 promotes expression of genes involved in oncogenic pathways and suppresses genes associated with erythroid differentiation. (A) Volcano plot showing all genes in the RNA sequencing analysis $>P < .05$. Each green dot represents a gene that is not differentially regulated. Upregulated DEGs in HSPCs, fold change >1.5 are indicated in red, downregulated DEGs in blue. Cumulative data of 3 biological independent experiments are shown for the vector control and Cdx4. (B) Heat map showing selected DEGs determined by RNA-Seq of Cdx4 and vector control-transduced HSPCs. Heat map representing unsupervised hierarchical clustering of all 6 samples for DEGs (fold change >1.5; $P < .05$). (C) Enrichr, ChEA, and ENCODE databases consensus transcription factor (TFs) enrichment analysis of down-regulated (DN) DEGs, demonstrating enrichment for known targets of *Gata1* and *Gata2*. (D) ENCODE and ChEA consensus TFs from ChIP-X analysis of downregulated DEGs (upper) and upregulated DEGs (lower). (E) The Molecular Signatures Database (MSigDB) oncogenic signature analysis of *Cdx4* induced DEGs. (F) Kyoto Encyclopedia of Genes and Genomes (KEGG) 2019 pathway analysis of *Cdx4* induced DEGs. The analysis in panels D-F was performed on Enrichr, web-based analysis platform.[40]

Discussion

Despite major recent progress in molecularly characterizing AEL, as reflected in the work by Iacobucci et al, it is still an open question how the erythroid phenotype is shaped in this AML subtype.[5] It could be successfully demonstrated that there are distinct molecular differences between AEL and non-AEL AML, but these differences were mostly in number, placing the mutational spectrum of adult AEL between those of MDS and AML, not explaining the fundamental differences in phenotypes.[5] One explanation would be that the cell of origin of AEL lies in the primitive erythroid compartment, transferring the erythroid phenotype to the full-blown leukemia. However, it is very unlikely that the complex molecular evolvement only occurs in a distinct erythroid precursor or that 1 key transforming factor within the erythroid compartment triggers the generation of the highly complex molecular landscape of AEL. Much more likely is a scenario in which the co-occurrence of mutated and/or aberrantly expressed factors leads to the emergence of erythroblasts beside myeloblasts, the hallmark of AEL. Interestingly, when genomic subtype-defining lesions were grouped by gene expression analysis in 4 distinct AEL subgroups, groups 2 and 3 were characterized by overexpression of several nonmutated HOX genes such as *HOXB5*, *HOXB6*, *HOXB8*, and *HOXB9*. There are several reports emphasizing the relevance of *HOX* gene expression for erythropoiesis; early reports have described that expression of *HOX* genes of the B cluster are predominantly expressed in erythroid cells, and that in particular, expression of HOXB6 is intimately associated with erythroid development and the TER119-positive phenotype in adult erythropoiesis.[29] Interestingly, we could demonstrate in our own previous experiments that constitutive expression of HOXB4 increases human erythroid colony formation, clearly distinct from HOXA10 in the same experimental system.[26,30]

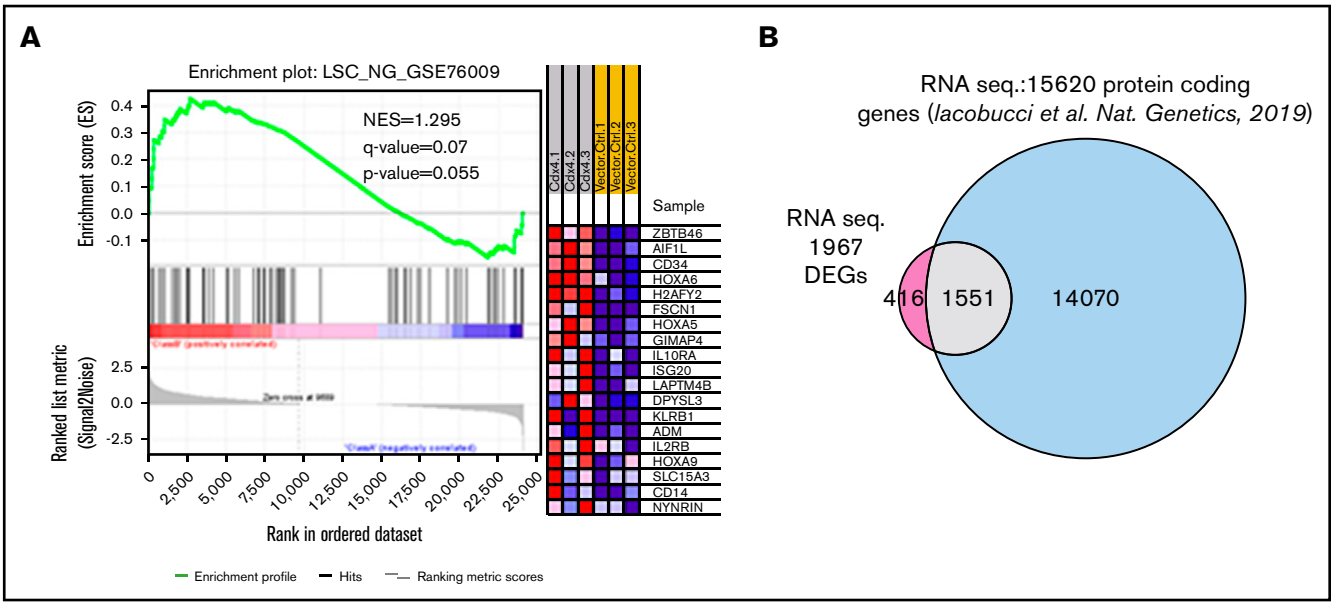

Figure 6. Genes upregulated in the RNA-Seq analysis of Cdx4 transduced murine HSPCs show an enrichment of AML LSC signature genes and overlap with protein coding genes in human AEL. (A) GSEA analysis based on RNA-Seq of Cdx4 transduced HSPCs, showing enrichment of human LSC signature.[27] Right, upregulated genes significantly enriched for the human LSC signature are shown as heat map. (B) Venn diagram representing the overlap of the DEGs on Cdx4 overexpression ($P < .05$) with published RNA sequencing data (protein coding genes) in human patients with AEL.[5] NES, normalized enrichment score.

CDX4 is a known upstream regulator of multiple HOX genes and, importantly, functions upstream of Gata1 in the zebrafish model.[31] This is in line with observations in the *kugelig* (*kgg*) zebrafish mutant, characterized by inactivated mutant Cdx4, which showed loss of Gata-1-positive hematopoietic cells in zebrafish embryos. This defect in Gata-1–positive cells could be nearly completely rescued by injecting hoxb7a and hoxa9 AMRNA into mutants.[32,33] Here we now demonstrate that constitutive expression of Cdx4 in adult hematopoietic stem and progenitor cells is able to induce AEL in the murine BM transplantation model. This effect of Cdx4 is remarkable for several reasons: first, there are very few models of mimicking human AEL in mice. Second, Cdx4 expression was sufficient to induce robustly AEL and was not depending on simultaneous coexpression of a second oncogene. Third, we did not see a mixture of different AML phenotypes as, for instance, observed in other murine AML models described by us and others.[11,34] An important feature of the presented Cdx4 AEL model is its long latency, going along with recurrent mutations also observed in patients with AEL. This distinguishes this model from the recently reported murine model, which observed rapid AEL development after coexpression of mutant NRK1 and Tp53, but no occurrence of mutations described in patients with AEL, probably because of the rapid kinetics of leukemia development in this model.[5] This underlines the value of the Cdx4 model for identifying putative co-occurring driver mutations in human AEL. Indeed, we detected 3 gene mutations recurrently found in the AEL data set. From these 3 genes, 2 genes were mutated in all 3 leukemic mice, and both genes carried mutations in the same domain as reported in patients with AEL. In addition to the emergence of co-occurring mutations, Cdx4 itself shaped the gene expression profile by downregulating target genes of *Gata1*, *Gata2*, and *Spi1.* Simultaneously, it upregulated stem cell and oncogenic signatures, in line with observations in an AEL mouse model driven by mutated NTRK1 in a TP53 background.[5] These changes were observed already 48 hours after successful gene transduction, thereby excluding confounding effects by co-occurring alterations in the in vivo settings. Strikingly, also on the protein level, expression resembled protein clusters of normal human erythroid progenitors already after 48 hours of gene exposure, underlining that Cdx4 as single factor is able to shape the expression profile of murine HSPCs. We could demonstrate that Cdx4 was able to increase expression of Hox genes such as Hoxb6 and Hoxb4 in CFCs formerly associated with erythropoiesis compared with the vector control.[26,29] However, our own published data showed that Hoxb4 cannot phenocopy the effect of Cdx4, not inducing AML as full-length protein.[35,36] Furthermore, we did not see any differences of HOXB3 and HOXB4 expression between human AEL vs other AML subtypes (data not shown). Despite all this, yet-unknown factors will contribute to the erythroid phenotype in our model. This is underlined by the fact that constitutive expression of Cdx4 caused AML, but not AEL, in a BM transplantation model, using the BALB/c genetic background.[37] This is in line with other murine AML models, showing divergent effects of bona fide AML-specific oncogenes in different mouse strains. This was shown for FMS-like tyrosine kinase 3 internal tandem duplication, which caused myeloproliferation in BALB/c mice but did not induce any hematological abnormalities in C57BL/6 mice,[11,38] possibly because of multiple intermingling factors such as environment, allelic heterogeneity, and stochastic effects, as well as the presence of modifier genes in different mouse strains.[39] In addition, CDX4 is expressed in human erythroid cell lines and patients with AEL, but at lower levels and only in a fraction of patients when measured by quantitative RT-PCR. This indicates that not 1 single gene will

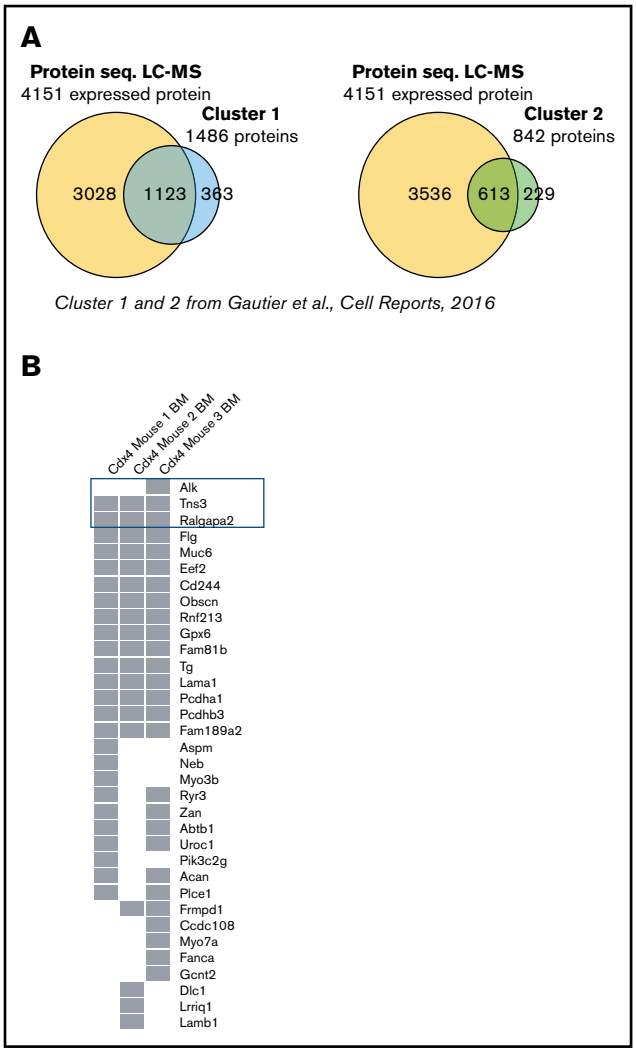

Figure 7. The protein profile of Cdx4 overexpressing HSPCs resembles the protein signature of immature erythroid stages. (A) Venn diagram representing the overlap of liquid chromatography-MS–based proteomics data (expressed proteins) to the published protein clusters, cluster 1 and cluster 2, derived from primitive stages of erythroid development.[28] (B) Gene variants represented in a matrix plot found in 3 mice indicating 34 genes that were also found in patients with AEL.[5] The 3 genes marked on top of the list were recurrently mutated (>2 patients) in patients with AEL.[5]

be important for developing the AEL phenotype, but that other factors or combination of genes will play an additional role in this AML subtype. Importantly, it has to be considered that the murine model presented here tests the effect of Cdx4 at the earliest stages of AEL development, overexpressing the gene in normal murine HSPCs and giving Cdx4-positive cells time to drive AEL development in vivo, most likely by acquiring additional mutations. In contrast, expression analyses of CDX4 in human patients with AEL mirrors the situation of full-blown leukemia, potentially underestimating the role of CDX4 at initiation and earlier stages of AEL development. To the end, the data emphasize a yet-unknown role of CDX4 in AEL development and demonstrate that aberrant expression of homeobox genes can play a pivotal role in the development of this AML subtype. We furthermore provide a novel murine AEL model, allowing us to test the functional relevance of co-occurring mutations by gain-of-function or loss-of-function experiments.

Acknowledgments

The authors thank all members of the Core Facility FACS, Core Facility Genomics of the University Ulm, and the animal facility of the Helmholtz Centre Munich and the University of Ulm for breeding and maintenance of the animals.

This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 1074 Project A6 [M.F.-B], Project Z1 and Project A4 [C.B.], and Project Z3 [S.W.]).

Authorship

Contribution: S.T., T.M., L.Q.-M., R.R., S.W., L.S., K.H.M., T. Herold, and U.K. performed research and analyzed the data; N.M.V. transplanted mice; K.D., T. Haferlach, and H.D. contributed research material; and C.B., V.P.S.R., and M.F.-B. contributed to data interpretation, wrote the manuscript, and designed the project.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: S.T., 0000-0003-3155-7293; L.Q.-M., 0000-0001-7156-5365; S.W., 0000-0001-5697-2608; K.H.M., 0000-0003-3920-7490; T. Herold, 0000-0002-9615-9432; U.K., 0000-0001-9845-3099.

Correspondence: Michaela Feuring-Buske, Department of Internal Medicine III, University Hospital of Ulm, Albert-Einstein Allee 23, 89081 Ulm, Germany; e-mail: michaela.feuring-buske@uni-ulm.de; and Vijay P. S. Rawat, Institute of Experimental Cancer Research, University Hospital of Ulm, Albert-Einstein Allee 23, 89081 Ulm, Germany; e-mail: vijay.rawat@uni-ulm.de.

References

1. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114(5):937-951.
2. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
3. Cervera N, Carbuccia N, Garnier S, et al. Molecular characterization of acute erythroid leukemia (M6-AML) using targeted next-generation sequencing. *Leukemia*. 2016;30(4):966-970.
4. Bacher U, Haferlach C, Alpermann T, Kern W, Schnittger S, Haferlach T. Comparison of genetic and clinical aspects in patients with acute myeloid leukemia and myelodysplastic syndromes all with more than 50% of bone marrow erythropoietic cells. *Haematologica*. 2011;96(9):1284-1292.

5. Iacobucci I, Wen J, Meggendorfer M, et al. Genomic subtyping and therapeutic targeting of acute erythroleukemia. *Nat Genet*. 2019;51(4):694-704.
6. Hegde S, Hankey P, Paulson RF. Self-renewal of leukemia stem cells in Friend virus-induced erythroleukemia requires proviral insertional activation of Spi1 and hedgehog signaling but not mutation of p53. *Stem Cells*. 2012;30(2):121-130.
7. Mager D, MacDonald ME, Robson IB, Mak TW, Bernstein A. Clonal analysis of the late stages of erythroleukemia induced by two distinct strains of Friend leukemia virus. *Mol Cell Biol*. 1981;1(8):721-730.
8. Moreau-Gachelin F, Wendling F, Molina T, et al. Spi-1/PU.1 transgenic mice develop multistep erythroleukemias. *Mol Cell Biol*. 1996;16(5):2453-2463.
9. Li Z, Herold T, He C, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. *J Clin Oncol*. 2013;31(9):1172-1181.
10. Deshpande AJ, Cusan M, Rawat VP, et al. Acute myeloid leukemia is propagated by a leukemic stem cell with lymphoid characteristics in a mouse model of CALM/AF10-positive leukemia. *Cancer Cell*. 2006;10(5):363-374.
11. Schessl C, Rawat VP, Cusan M, et al. The AML1-ETO fusion gene and the FLT3 length mutation collaborate in inducing acute leukemia in mice. *J Clin Invest*. 2005;115(8):2159-2168.
12. Rawat VP, Cusan M, Deshpande A, et al. Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2004;101(3):817-822.
13. Rawat VP, Thoene S, Naidu VM, et al. Overexpression of CDX2 perturbs HOX gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated HOX gene expression in human acute myeloid leukemia. *Blood*. 2008;111(1):309-319.
14. Rawat VP, Arseni N, Ahmed F, et al. The vent-like homeobox gene VENTX promotes human myeloid differentiation and is highly expressed in acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2010;107(39):16946-16951.
15. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
16. Liao R, Zhang R, Guan J, Zhou S. A new unsupervised binning approach for metagenomic sequences based on N-grams and automatic feature weighting. *IEEE/ACM Trans Comput Biol Bioinformatics*. 2014;11(1):42-54.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
18. Groß A, Kracher B, Kraus JM, et al. Representing dynamic biological networks with multi-scale probabilistic models. *Commun Biol*. 2019;2(1):21.
19. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372.
20. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-1805.
21. Smith CM, Hayamizu TF, Finger JH, et al. The mouse Gene Expression Database (GXD): 2019 update. *Nucleic Acids Res*. 2019;47(D1):D774-D779.
22. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE; Mouse Genome Database Group. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2019;47(D1):D801-D806.
23. Wiśniewski JR, Hein MY, Cox J, Mann M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics*. 2014;13(12):3497-3506.
24. Kogan SC, Ward JM, Anver MR, et al; Hematopathology subcommittee of the Mouse Models of Human Cancers Consortium. Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood*. 2002;100(1):238-245.
25. Rawat VP, Humphries RK, Buske C. Beyond Hox: the role of ParaHox genes in normal and malignant hematopoiesis. *Blood*. 2012;120(3):519-527.
26. Buske C, Feuring-Buske M, Abramovich C, et al. Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells. *Blood*. 2002;100(3):862-868.
27. Ng SW, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433-437.
28. Gautier EF, Ducamp S, Leduc M, et al. Comprehensive proteomic analysis of human erythropoiesis. *Cell Reports*. 2016;16(5):1470-1484.
29. Zimmermann F, Rich IN. Mammalian homeobox B6 expression can be correlated with erythropoietin production sites and erythropoiesis during development, but not with hematopoietic or nonhematopoietic stem cell populations. *Blood*. 1997;89(8):2723-2735.
30. Buske C, Feuring-Buske M, Antonchuk J, et al. Overexpression of HOXA10 perturbs human lymphomyelopoiesis in vitro and in vivo. *Blood*. 2001;97(8):2286-2292.
31. Pillay LM, Forrester AM, Erickson T, Berman JN, Waskiewicz AJ. The Hox cofactors Meis1 and Pbx act upstream of gata1 to regulate primitive hematopoiesis. *Dev Biol*. 2010;340(2):306-317.
32. Davidson AJ, Ernst P, Wang Y, et al. cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes. *Nature*. 2003;425(6955):300-306.
33. Lengerke C, Daley GQ. Caudal genes in blood development and leukemia. *Ann N Y Acad Sci*. 2012;1266(1):47-54.
34. McKenzie MD, Ghisi M, Oxley EP, et al. Interconversion between tumorigenic and differentiated states in acute myeloid leukemia. *Cell Stem Cell*. 2019;25(2):258-272.
35. Cusan M, Vegi NM, Mulaw MA, et al. Controlled stem cell amplification by HOXB4 depends on its unique proline-rich region near the N terminus. *Blood*. 2017;129(3):319-323.
36. Antonchuk J, Sauvageau G, Humphries RK. HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. *Cell*. 2002;109(1):39-45.
37. Bansal D, Scholl C, Fröhling S, et al. Cdx4 dysregulates Hox gene expression and generates acute myeloid leukemia alone and in cooperation with Meis1a in a murine model. *Proc Natl Acad Sci USA*. 2006;103(45):16924-16929.
38. Kelly LM, Liu Q, Kutok JL, Williams IR, Boulton CL, Gilliland DG. FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. *Blood*. 2002;99(1):310-318.
39. Nadeau JH. Listening to genetic background noise. *N Engl J Med*. 2005;352(15):1598-1599.
40. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90-7.