# Supporting Information

## Rawat et al. 10.1073/pnas.1001878107

### SI Materials and Methods

**Phylogenetic Tree Construction.** The CLUSTAL algorithm in the ClustalX 2.0 program was used to align the protein sequences (see below) and to build phylogenetic trees using the neighbor-joining method and bootstrapping ($n$ = 1000) (1). *Drosophila* VENTX and Hydra Meis1 were used as outgroups. The phylogenetic trees were drawn with TreeViewPPC version 1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/rod.html). The graphs were annotated using a vector drawing program (Freehand MX; Macromedia). A protein BLAST search was performed to identify all *VENTX* homologous sequences in mammals. The following sequences were used for the *VENTX* tree: *Xenopus*—NP_001081607.1; chicken—XP_427647.2; zebrafish—NP_571775.1; opossum—XP_001380984.1; dog—XP_548821.2 (here the similarity region to the kinase noncatalytic C-lobe domain containing 1 was removed); rhesus monkey—XP_001092717.1; human—NP_055283.1; chimpanzee—XP_521666.1; lancelet—XP_002226868.1; and Drosophila—NP_523387.1. Not all available mammalian Meis1 protein sequences were used for the alignment. The following protein sequences were used for Meis1: human—NP_002389.1; mouse—Q60954.1; rat—NP_001128174.1; rhesus monkey—XP_001092069.1; chimpanzee—XP_001167056.1; opossum—XP_001374596.1; axolotl—AAZ95042.1; chicken—NP_001132949.1; *Xenopus*—AAI70273.1; zebrafish—NP_571968.1; lancelet—XP_002221242.1; sea urchin—XP_001197656.1; Drosophila—NP_476577.2; sea anemone—XP_001628502.1; and hydra—XP_002162967.1.

**Quantification of *VENTX* Expression.** Total RNA was isolated using an RNeasy micro or mini kit (Qiagen) to avoid genomic DNA contamination, and samples were treated with DNase enzyme according to the manufacturer's instructions. cDNA was prepared using random hexamers (Thermoscript RT–PCR system; Invitrogen) according to the manufacturer's instructions. For endogenous control, the human β-Actin gene (*ACTB*) was used. Reactions were run in triplicates with 1 μl of cDNA in a total reaction volume of 20 μl on an ABI PRISM 7900 Sequence Detection System (Applied Biosystems).

**Retroviral Vectors and Transduction.** The *VENTX* cDNA was provided by Paul Moretti (Institute of Medical and Veterinary Science, Adelaide, Australia). The *VENTX* cDNA was subcloned into the pMSCV-IRES-GFP vector. As a control, the MSCV vector carrying only the IRES–green fluorescent protein cassette (GFP virus) was used. High-titer, helper-free recombinant retrovirus was generated by transfecting the amphotropic Phoenix cell line and subsequently transducing the gibbon ape leukemia virus-pseudotyped PG13 packaging cell line as previously described (2). High-titer single-cell PG13 clones were isolated for each virus and used in subsequent experiments. Transduction efficiency was tested in K562 cells. The transduction of primary human cord blood cells was performed as previously described (2). For the in vitro experiments, successfully transduced CD34+ GFP+ cells were highly purified by FACS (FACS Vantage; Becton Dickinson).

**Preparation of Human CD34$^+$ Cells.** Cord blood MNCs were separated using density gradient centrifugation on Pancoll (PAN Biotech). CD34$^+$ cell enrichment was conducted using an immunoselection MACS CD34 Cell Isolation kit (Miltenyi Biotech). Enriched samples were cryopreserved in FBS and 10% dimethyl sulfoxide and stored in liquid nitrogen until use. CD34$^+$-enriched bone marrow cells were purchased from Cell Systems.

**In Vitro Progenitor Assays.** Hematopoietic colony-forming cells (CFC) were assayed using methylcellulose-based medium (MethoCult H4434; Stem Cell Technologies) as previously described (2). Briefly, GFP+ CD34+ cells were diluted with Iscove's Modified Dulbecco's Medium (IMDM) + 2% FBS to 10x the final concentration(s) required for plating. For a duplicate assay, 1.1 ml of cell-methylcellulose mixture were dispensed into 35 mm culture dishes. The 35 mm culture dishes were placed in a $CO_2$ incubator at 37 °C and >95% humidity. CFC numbers were evaluated after an incubation period of 12–14 days and distinguished into the different types of colonies as indicated. CFC assays were performed by replating aliquots of cells obtained by harvesting 14-day-old primary CFC cultures as previously described (2). Lymphoid cell development was assessed by plating 1x10$^5$ highly purified CD34$^+$ cells from both experimental arms on murine MS-5 cells in RPMI 1640 with 10% fetal calf serum (FCS) and 5% human antibody serum with 50 ng/mL steel factor, 10 ng/mL IL-2, and 10 ng/mL IL-15 as previously described (3). After 3 weeks, both adherent and nonadherent cells were collected and analyzed by FACS for the expression of CD19 and CD15. Six-week long-term culture-initiating cell (LTC-IC) limiting dilution assays were carried out using pre-established irradiated murine fibroblasts genetically engineered to produce human IL-3, G-CSF and SF as feeder layers. Cell numbers ranging from 1 to 12800 GFP$^+$ cells from both experimental arms were sorted into 96-well plates (Sarstedt, Nümbrecht, Germany). LTC-IC frequencies were calculated using Poisson statistics and the method of maximum likelihood with the assistance of the L-Calc software (Stem Cell Technologies).

**Flow Cytometric Analyses.** Cells isolated from colonies plated in methylcellulose or harvested nonobese diabetic/severe combined immunodeficient (NOD/SCID) bone marrow (BM) cells were stained according to standard protocols: cells were resuspended in cell-wash buffer (IMDM with 3% FCS, 0.2 mM EDTA) and treated with 1% ammonium chloride solution (20 min on ice) for selective removal of red cells. Approximately 10$^6$ cells were resuspended in PBS (with 3% FBS) and incubated on ice for 10 min with the anti-mouse IgG Fc-receptor antibody 2.4G2 (Miltenyi Biotech) to block nonspecific antibody binding. Separate aliquots were then incubated for 30 min on ice with a mouse isotype-matched control antibody or antibodies against the following human antigens: CD45-PE, CD34-PE, CD38-APC, CD33-PE, CD15-PE, and CD19-APC [Becton Dickinson (BD)-Pharmingen]. For quantification of *VENTX* expression by real-time PCR, human BM and cord blood (CB) samples were stained with the following antibodies: CD34-PE, CD38-APC, CD33-PE, CD15-PE, CD19-APC, CD3, and glycophorin A (Gly A). Cells were sorted using FACSvantage cell sorter (Becton, Dickinson).

**Western Blot Analysis of *VENTX* Expression Level.** Radioimmunoprecipitation assay (RIPA) lysates were prepared from the PG-13 VENTX and PG-13 GFP packaging cell lines. Extracts were separated on 12% SDS/PAGE, and Western blots were performed as described (3). The membrane was probed with anti-VENTX polyclonal antibody (provided by P. A. Moretti, Institute of Medical and Veterinary Science, Adelaide, Australia, or Abcam) and anti-β-actin antibodies (Santa Cruz Biotechnology).

**Gene Expression Profiling.** Gene expression profiling was performed using Affymetrix HG-U133plus2.0 oligonucleotide microarrays. Details regarding sample preparation, hybridization, and image acquisition have been described previously with some modifications (4). Briefly, RNA was isolated from human normal

$CD34^+$ cord blood cells successfully transduced with the MSCV-VENTX-IRES-GFP vector or the empty retroviral control 48 h after the end of the transduction protocol. Before RNA extraction, cells were sorted for GFP and $CD34^+$. RNA (4 μg) was used as starting material. First-strand cDNA synthesis and second-strand cDNA synthesis were carried out following the Invitrogen standard protocol (ThermoScript RT–PCR System; Invitrogen). Biotin-labeled cRNA synthesis was performed in a final reaction volume of 40 μl. Array hybridization was performed for 16 h. Array post-hybridization washing and scanning were performed following detailed protocols provided in the GeneChip Expression Analysis Technical Manual (Affymetrix).

**Microarry Data Analysis.** For the analysis of retrovirally transduced primary bone marrow cells, data processing was carried out using the R-package (CRAN) and Bioconductor (R version 2.8.1 and Bioconductor version 2.5). The raw expression data were normalized using the GeneChip RMA (GCRMA) method, generating log-transformed expression values. Evaluation of differentially expressed genes was carried out using Student's *t* test in the dChip software at a 95% confidence interval for a fold regulation of 1.2 with the false discovery rate permuation set to 1,000. The heat map of some selected genes was also generated in dChip. The list of differentially regulated genes was later used for Gene Ontology analysis, which was done using the KegArray (based on KEGG database) program (KegArray version 0.5.7. (Beta) build 20071210; Kanehisa Laboratories).

For the analysis of primary acute myeloid leukemia (AML) samples total mRNA was isolated from patient samples, processed, and analyzed on the Affymetrix HG-U133A and HG-U133B chips as described before (5). The CEL file data from the samples used in the comparison were normalized together according to the procedure described by Huber et al. (6). Normalized expression data were then analyzed with the R software package and the box-plot function (www.r-project.org). Expression signal intensities are given on a logarithmic scale.

**RNAi-Mediated Silencing of VENTX Expression.** The pLKO.1 construct contains a human U6 promoter that drives expression of the stemloop cassette and the puromycin resistance gene cloned 3′ of the human phosphoglycerate kinase promoter. VSV-G–pseudotyped lentiviral particles were produced by cotransfection of 293T cells (provided by Claudia Scholl, Department of Internal Medicine III, University Hospital of Ulm, Ulm, Germany) with pLKO.1 (Sigma) constructs and the compatible packaging plasmids pMD2G and psPAX2 (Addgene). Transfections were carried out using FuGENE 6 (Roche Diagnostics), and virus was harvested 48 and 72 h after transfection. Lentiviral supernatants were used to transduce cells in the presence of 8 μg/mL polybrene (Sigma-Aldrich), and infected cells were selected with 2 μg/mL puromycin (Sigma-Aldrich).


1. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
2. Ahmed F, et al. (2008) Constitutive expression of the ABC transporter ABCG2 enhances the growth potential of early human hematopoietic progenitors. *Stem Cells* 26: 810–818.
3. Buske C, et al. (2002) Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells. *Blood* 100:862–868.
4. Metzeler KH, et al.; Cancer and Leukemia Group B; German AML Cooperative Group (2008) An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood* 112:4193–4201.
5. Schoch C, et al. (2002) Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci USA* 99:10008–10013.
6. Huber W, von Heydebreck A, Sueltmann H, Poustka A, Vingron M (2003) Parameter estimation for the calibration and variance stabilization of microarray data. *Stat Appl Genet Mol Biol* 2:Article3.


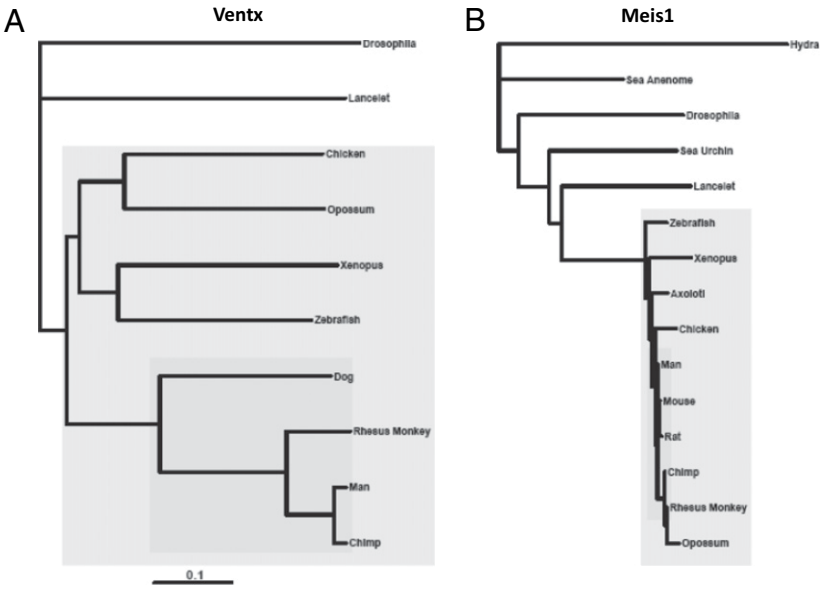


**Fig. S1.** Phylogenetic trees of VENTX and Meis1. (*A*) VENTX protein sequences of 10 animal species for which VENTX homologs were found after a BLAST search was aligned and a phylogenetic tree was constructed. (*B*) The Meis1 protein sequences of 15 species that were annotated as Meis1 or Meis1 homologs were used for alignment and phylogenetic tree construction. Vertebrate species are enclosed in the lightly shaded box, and mammals are enclosed in the slightly darker shaded box.

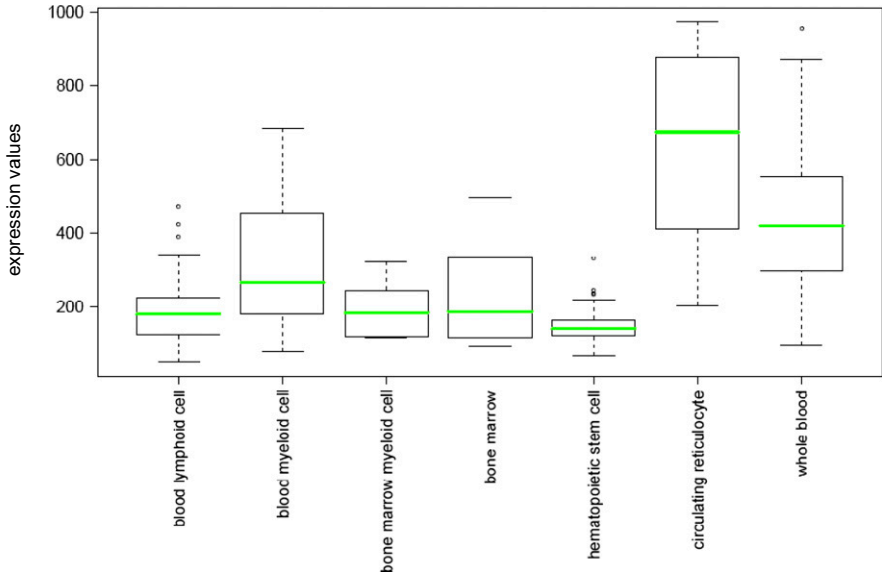


**Fig. S2.** Analysis of expression levels of *VENTX* in normal hematopoietic tissues by cDNA microarray (1). The normalized and variance stabilized expression values are shown on a linear scale (*y* axis).

1. Kilpinen S, et al. (2008) Systematic bioinformatic analysis of expression levels of 17,330 human genes across 9,783 samples from 175 types of healthy and pathological tissues. *Genome Biol* 9:R139.

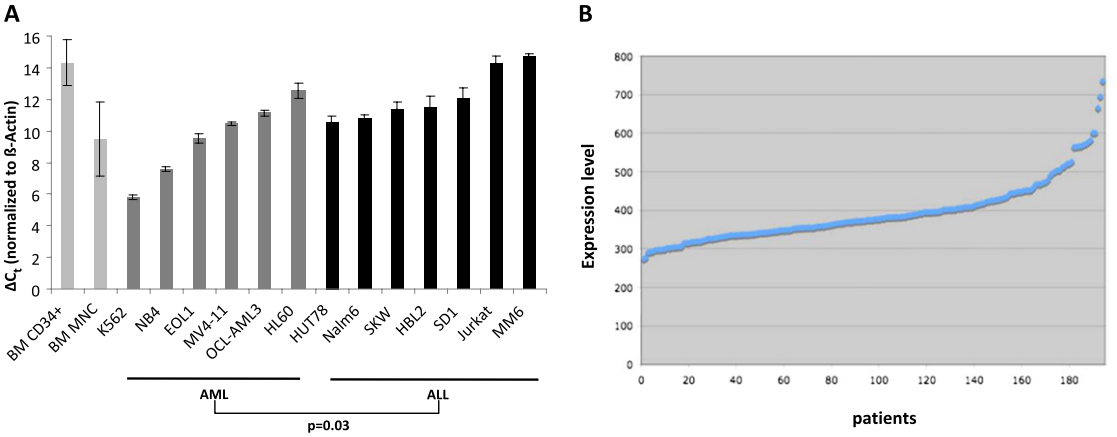


**Fig. S3.** VENTX expression in AML cell lines and primary AML samples. (*A*) VENTX expression of AML and ALL cell lines compared with BM MNCs and $CD34^+$ BM cells. Bars represent average expression ±SD. Note that ΔCT values are inversely correlated to expression levels. (*B*) VENTX expression levels in 194 patients with AML and normal karyotype. The normalized and variance stabilized expression values (microarray signal intensity values) are shown on a linear scale (y axis).

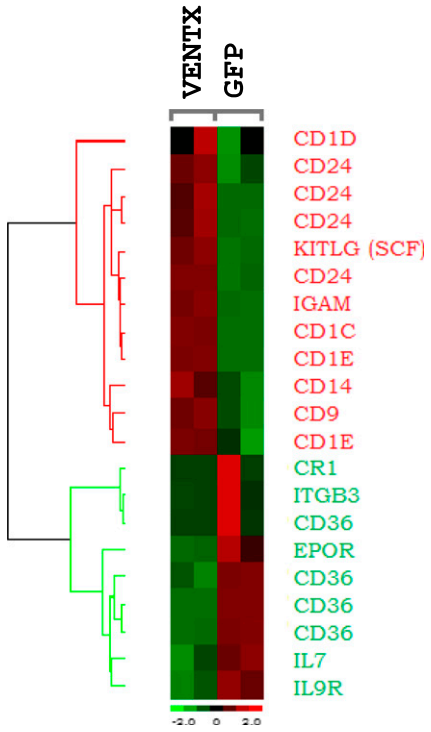


**Fig. S4.** Heat map showing differentially regulated hematopoietic genes after retrovirally induced expression of *VENTX* compared with the control. Red squares indicate up-regulation and green squares show down-regulation (color scale is given at the bottom of the figure). Red-shaded hierarchy (branch) shows the cluster of genes up-regulated in VENTX-GFP$^+$/CD34$^+$ cells, and the green branch indicates genes down-regulated in VENTX-GFP$^+$/CD34$^+$ cells compared with CD34$^+$ cells infected with the control vector. (Hierarchy was constructed on the basis of the Euclidean distance with average linkage.)

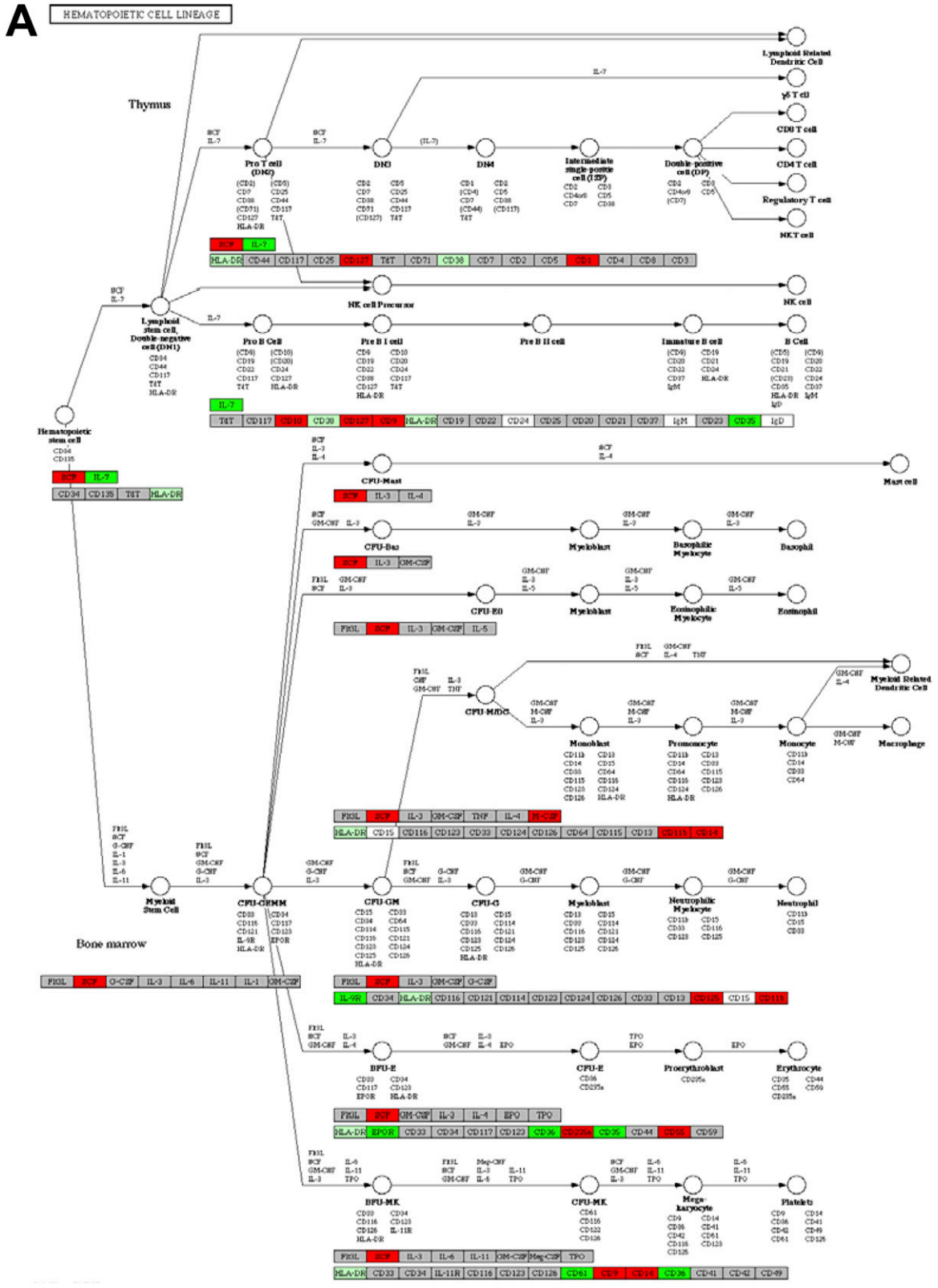


**Fig. S5.** (Continued)

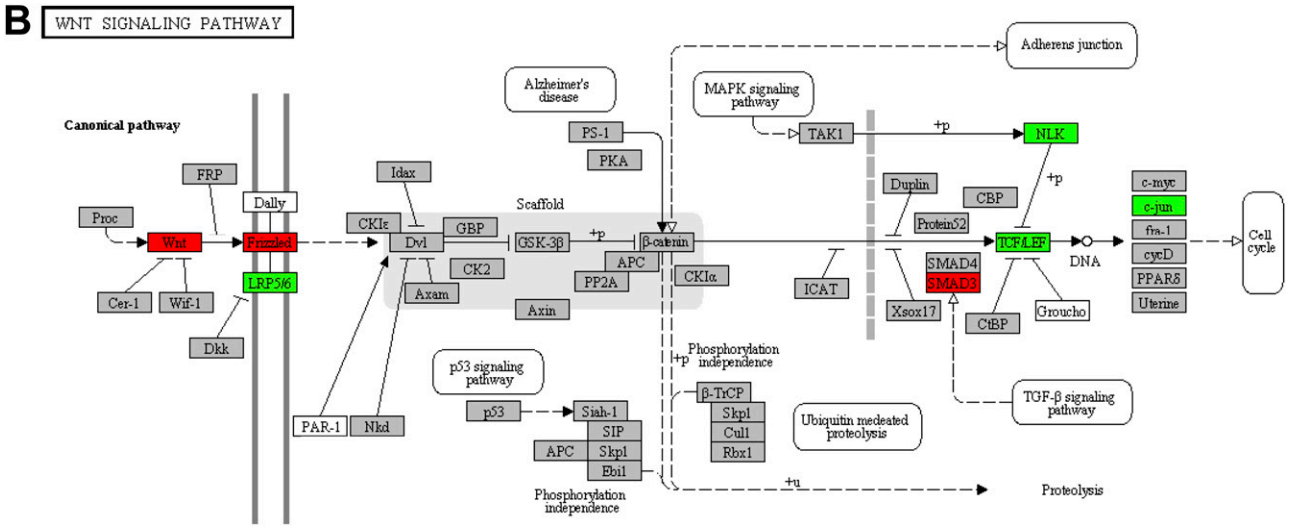


**Fig. S5.** (*A*) Pathway analysis by KegArray analysis originating from three independent experiments. The entire list of differentially regulated genes was used for KegArray analysis (KegArray v 0.5.7 Beta) to discover enrichment of functions and processes on the basis of the KEGG database. The "hematopoietic lineage" was one of the significantly enriched pathways/processes. Boxes beneath the different branches indicate genes involved in specific differentiation processes. Only the myeloid lineage is presented here for clarity. (*B*) Analysis of the WNT pathway by KegArray originating from three independent experiments. Genes labeled green or red are down- or up-regulated, respectively.

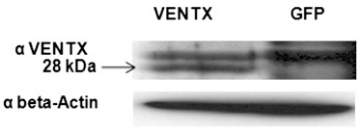


**Fig. S6.** VENTX protein expression as demonstrated by Western blotting in PG13 transduced with the MSCV-VENTX-IRES-GFP construct.

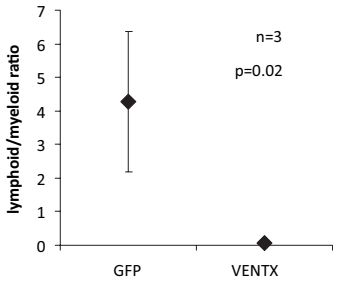


**Fig. S7.** VENTX generates an increased number of myeloid cells in a stroma-cell–dependent coculture system. After a 3-wk culture of cells on MS-5 feeder cells under lymphoid conditions, cells were harvested and immunophenotyped. The lymphoid/myeloid ratio was calculated by dividing the percentage of $CD19^+$ cells by the percentage of $CD15^+$ cells. Diamonds indicate the average ratio of three independent experiments ±SD.

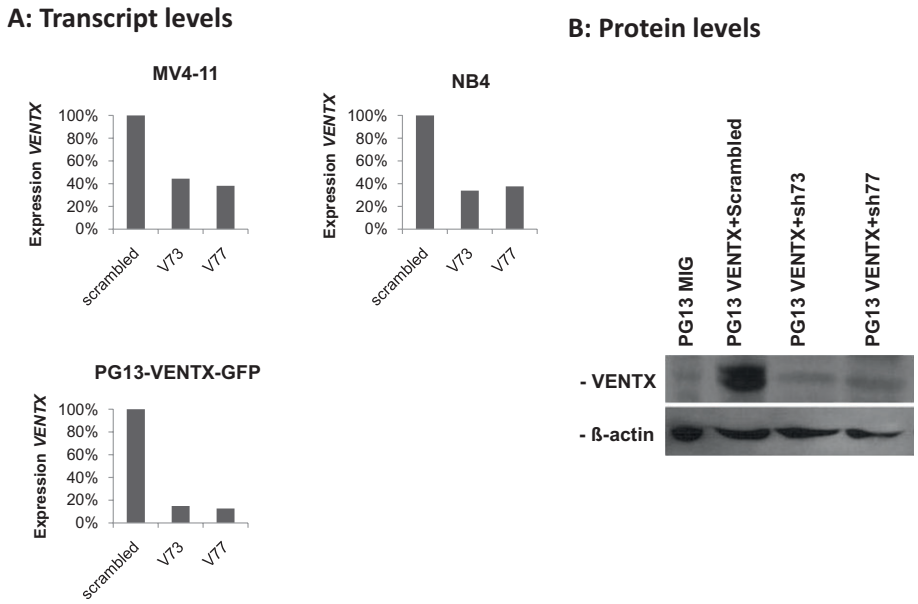


**Fig. S8.** Human AML cell lines and PG13 cells, stably expressing the MSCV-VENTX-IRES-GFP construct, were transduced with lentiviral vectors containing stem-loop cassettes encoding shRNAs (V77 and V73) targeted to the coding sequence of the human *VENTX* mRNA. Suppression of *VENTX* transcript (*A*) and protein levels (*B*) are shown compared with a control (scrambled) construct.

**Table S1. Characteristics of AML patients**

| Molecular subtype | Patient no. | Age (y) | Gender | Karyotype | Other relevant markers | *VENTX2* expression | CD34 (%) |
|---|---|---|---|---|---|---|---|
| AML1-ETO | 1 | 28 | M | 45,XY,der(4)t(4;16)(q12;q11),der(7)t(3;7)(?;q2?),t(8;21)(q22;q22),del(9)q11),-13 | | High | 43 |
| | 2 | 64 | F | 45,X,-X,t(8;21)(q22;q22),del(9)(q22) | Jak2-mutation+ | High | NA |
| | 3 | 31 | M | 47,XY,t(8;21)(q22;q22),+21c | NRAS-mutation (Codon12)+ | Low | NA |
| | 4 | 61 | F | 45,X,Y,t(2;6)(q37;q23),t(7;9;15)(q22;q22;q24),t(8;21)(q22;q22),t(3;11)(p24;q23),t(4;20)(q11;p13) | | Low | 75 |
| | 5 | 25 | M | 47,XY,der(7)t(7;11)(q31;?),t(8;21)(q22;q22),+15 | | High | 95 |
| | 6 | 44 | F | 47,XX,+8,t(8;21)(q22;q22),del(9)(q22q34) | tAML after breast cancer | High | NA |
| | 7 | 41 | M | 45,X,-Y,t(8;21)(q22;q22) | FLT3-ITD+ | High | 31 |
| | 8 | 75 | M | 47,XY,+4,t(8;21)(q22;q22) | cKIT D816+ | Low | NA |
| | 9 | 63 | M | 45,X,-Y,t(8;21)(q22;q22) | cKIT D816+ | Low | 25 |
| NPM1-/FLT3-ITD+ | 10 | 19 | M | 47,XY,+8 | | High | NA |
| | 11 | 57 | M | 45,XY,-21 | | Low | 93 |
| | 12 | 72 | F | 46,XX | | High | 31 |
| | 13 | 70 | M | 47,XYY | | High | 6 |
| | 14 | 60 | F | 46,XX | | High | 84 |
| | 15 | 44 | F | 47,XX,+8 | | Low | 67 |
| | 16 | 72 | M | 46,XY | | Low | 81 |
| | 17 | 55 | M | 46,XY | | Low | 49 |
| | 18 | 61 | F | 46,XX | | High | NA |
| NPM1+/FLT3-ITD- | 19 | 56 | M | 46,XY | | Low | NA |
| | 20 | 72 | F | n.a. | | Low | 1 |
| | 21 | 28 | M | n.a. | | Low | NA |
| | 22 | 53 | M | 46,XY | | High | NA |
| | 25 | 46 | F | 46,XX | | High | 3 |
| | 23 | 21 | F | 46,XX | | High | 1 |
| | 24 | 71 | F | n.a. | | High | 1 |
| | 25 | 64 | M | 46,XY | | High | 76 |
| | 26 | 56 | M | 46,XY | | Low | NA |

"Age" refers to age at the time of diagnosis in years. M: male; F: female. NA: not available. tAML: treatment-associated acute myeloid leukemia. *VENTX2* expression: high or low are defined as above or below the median.

**Table S2. Probesets that are at least twofold (log2 scale) differentially regulated**

| Probe set | Gene | GFP+_CD34+ | VENTX+_CD34+ | Fold change |
|---|---|---|---|---|
| 206647_at | HBZ: hemoglobin, ζ /// hemoglobin, ζ | 2.59 | 8.8 | 3.4 |
| 211959_at | IGFBP5: insulin-like growth factor binding protein 5 | 2.16 | 6.72 | 3.11 |
| 215784_at | CD1E: CD1e molecule | 2.29 | 6.86 | 3 |
| 203936_s_at | MMP9: matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92-kDa type IV collagenase) | 3.51 | 10.21 | 2.91 |
| 220811_at | PRG3: proteoglycan 3 | 2.87 | 7.79 | 2.71 |
| 1556314_a_at | (clone 1NIB-4) normalized cDNA library sequence | 2.34 | 5.85 | 2.5 |
| 239586_at | FAM83A: family with sequence similarity 83, member A | 3.15 | 7.74 | 2.46 |
| 205863_at | S100A12: S100 calcium-binding protein A12 (calgranulin C) /// S100 calcium binding protein A12 (calgranulin C) | 3.73 | 9.14 | 2.45 |
| 223828_s_at | LGALS12: lectin, galactoside-binding, soluble, 12 (galectin 12) | 2.6 | 6.22 | 2.39 |
| 202286_s_at | TACSTD2: tumor-associated calcium signal transducer 2 | 2.17 | 5.02 | 2.32 |
| 228170_at | OLIG1: oligodendrocyte transcription factor 1 | 3.19 | 7.05 | 2.21 |
| 205987_at | CD1C: CD1c molecule | 3.04 | 6.64 | 2.19 |
| 205249_at | EGR2: early growth response 2 (Krox-20 homolog, Drosophila) | 2.66 | 5.8 | 2.18 |
| 203903_s_at | HEPH: hephaestin | 2.78 | 5.98 | 2.15 |
| 209173_at | AGR2: anterior gradient 2 homolog (*Xenopus laevis*) | 3.36 | 7.17 | 2.13 |
| 205557_at | BPI: bactericidal/permeability-increasing protein | 3.4 | 7.02 | 2.07 |
| 203645_s_at | CD163: CD163 molecule | 2.82 | 5.8 | 2.06 |
| 202177_at | GAS6: growth arrest-specific 6 | 2.32 | 4.75 | 2.04 |
| 206134_at | ADAMDEC1: ADAM-like, decysin 1 | 3.52 | 7.13 | 2.03 |
| 201506_at | TGFBI: transforming growth factor, β-induced, 68kDa | 5.39 | 10.87 | 2.02 |
| 210262_at | CRISP2: cysteine-rich secretory protein 2 | 4.62 | 2.3 | −2.01 |
| 233425_at | ZCCHC2: zinc finger, CCHC domain containing 2 | 4.49 | 2.2 | −2.04 |
| 1553183_at | UMODL1: uromodulin-like 1 | 5.47 | 2.53 | −2.16 |
| 203854_at | CFI: complement factor I | 5.05 | 2.32 | −2.18 |
| 219255_x_at | IL17RB: interleukin 17 receptor B | 6.79 | 3.11 | −2.18 |
| 227812_at | TNFRSF19: tumor necrosis factor receptor superfamily, member 19 | 6.79 | 2.7 | −2.52 |
| 224009_x_at | DHRS9: dehydrogenase/reductase (SDR family) member 9 | 7 | 2.72 | −2.58 |
| 224156_x_at | IL17RB: interleukin 17 receptor B | 7.44 | 2.79 | −2.67 |

VENTX-GFP+/CD34+ were compared with the control (MIG-GFP+/CD34+).

**Table S3. MS-5 coculture assay of human progenitor cells transduced with VENTX-GFP or GFP**

| Vector* | $CD19^+$ cells per culture | $CD15^+$ cells per culture |
|---|---|---|
| GFP | 1530 ± 436 | 975 ± 713 |
| VENTX | 515 ± 173 | 10200 ± 3730 |
| *P* value | $P < 0.05$ | $P < 0.05$ |

Absolute numbers of $CD19^+$ and $CD15^+$ cells per culture are given.
*1 x $10^5$ $CD34^+$ $VENTX^+$ or $GFP^+$ cells plated initially ($n = 3$).

**Table S4. Limiting dilution competitive repopulating unit (CRU) assays of human progenitors transduced with VENTX-GFP or GFP vectors**

| Phenotype | No. of $CD34^+/GFP^+$ cells transplanted | No. of positively engrafted mice*/total no. | CRU frequency† (+/−1 SE) | 95% CI upper and lower frequencies |
|---|---|---|---|---|
| VENTX-GFP | 3,400 | 1/7 | 1 in 41,971 (1/30,630–1/57,512) | 1/22,636–1/77,821 |
| | 13,700 | 1/4 | | |
| | 37,500 | 0/2 | | |
| | 55,000 | 6/9 | | |
| | 75,000 | 0/2 | | |
| | 220,000 | 1/1 | | |
| | 600,000 | 2/2 | | |
| GFP | 3,400 | 0/4 | 1 in 74,042 (1/51,945–1/105,540) | 1/36,963–1/148,319 |
| | 13,700 | 0/3 | | |
| | 16,500 | 3/4 | | |
| | 37,000 | 2/2 | | |
| | 55,000 | 2/3 | | |
| | 67,000 | 2/2 | | |
| | 75,000 | 3/5 | | |
| | 220,000 | 1/1 | | |

*Mice with lymphoid/ myeloid engraftment 6 wk after transplantation.
†Differences were not significant.