

Supplemental Figure Legends:

Fig. S1. (A) Alignment of HoxB cluster proteins (ref. sequences: HOXB1: gi|64654265; HOXB2: gi|4504465; HOXB3: gi|4204923; HOXB5: gi|4504469; HOXB6: gi|23503237; HOXB7: gi|85068580; HOXB8: gi|11993911; HOXB9: gi|11993915; HOXB13: gi|47125370), and (B) selected oncogenic proteins: HOXA7 (gi|3097078), HOXA9 (gi|14603109), HOXA10 (gi|182765442). Bolded red sequence indicates the proline region of HOXB4-wt and the underlined sequence is absent in the HOXB4- Δ PRD mutant used in the present work. * indicates conserved residues; : and . indicate partially conserved residues. (C) The track shows multiple alignments of 100 vertebrate species and measurements of evolutionary conservation of HOXB4 using two methods (*phastCons* and *phyloP*) from the PHAST package for all species. Multiple alignments were generated using multiz and other tools in the UCSC/Penn State Bioinformatics comparative genomics alignment pipeline. Conservation scores are displayed as a *wiggle track* (histogram) in which the height reflects the size of the score. Horizontal black bar indicates the proline region deleted in this work.

Fig. S2. (A) Schematic representation of the MSVC-based retroviral constructs and structure of the HOXB4 proteins used in this study (LTR: long term repeat, IRES: internal ribosomal entry site, GFP: green fluorescent protein, PID: PBX interaction domain; HD: homeodomain). (B) Left: Western blot analysis from 293T and 32D cell lines overexpressing the different constructs. Equal amounts of whole cell protein extracts were loaded. Right: FACS analysis showing the proportion and intensity of GFP⁺ BM cells 48h after transduction. (C) Intracellular staining for HOXB4 in 293T cells (anti-HOXB4 antibody, N-18 Santa Cruz).

Fig.S3 (A) Primary CFC (n=5) and (B) secondary CFC (n=3) assays. (C) Overtime chimerism for myeloid markers within the GFP⁺WBCs in the peripheral blood of transplanted mice (HOXB4-wt n=14; HOXB4- Δ PRD₂₂₉₋₃₆₀ n=10; GFP n=3). The statistically significant differences between HOXB4- Δ PRD₂₂₉₋₃₆₀ mice and GFP mice at 36-86 weeks post

transplantation ($p < 0.0003$) are indicated. * indicate statistically significant differences between the immunophenotype at 36th-48th weeks in comparison to the corresponding immunophenotype at 13th-16th weeks post transplantation (HOXB4-wt Mac-1⁺ $p < 0.0475$, HOXB4- Δ PRD₂₂₉₋₃₆₀ Mac-1⁺ $p < 0.000004$, and HOXB4- Δ PRD₂₂₉₋₃₆₀ Gr-1⁺ $p < 0.0013$, respectively). (D) Blood smears and cytopsin preparations from different tissues of representative mice are shown (bars: 8 μ m, X630). (E) Flow cytometry analysis of peripheral blood, bone marrow, and spleen cells isolated from representative primary GFP, HOXB4-wt, and HOXB4- Δ PRD₂₂₉₋₃₆₀ mice.

Fig.S4 (A) Schematic representation of the different constructs used: HOXB4- Δ PRD₂₂₉₋₃₆₀ and constructs with re-insertion of parts of the proline stretch as indicated. (B) Secondary CFC assay from BM progenitor cells freshly transduced with the constructs indicated. Results of three independent experiments and significances are given.

Fig. S5. (A) Schematic representation of the Hoxa9 constructs used in the present work. The proline region was inserted by synthesis after the C-terminal end (bottom). (B) Structure prediction analysis using the IUPred web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content¹. Blue bars indicate the homeodomain, red bars indicate the proline region. Dotted red bars indicate scramble sequences. (C) Cytopsin obtained from liquid culture from BM progenitor cells after overexpression of Hoxa9 constructs alone or in combination with Meis1. (Bar: 8 μ m). (D) Expression analysis of Hoxa9 3 weeks after co-transduction with Meis1 and Hoxa9 constructs as determined by qRT-PCR ($n=3$). (E) FACS analysis showing the proportion and intensity of the GFP signal from transduced cells *in vitro*. (F) Primary CFC assay performed with freshly co-transduced and highly purified BM progenitors overexpressing Meis1 and the Hoxa9 constructs as indicated ($n=3$).

Fig. S6. (A) Gene ontology analysis of genes differentially expressed in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt in BM cells (FDR \geq 0.25; GSEA). Vertical bars indicate % of genes differentially expressed, which contribute to the enrichment score. (B) Volcano plot representing differentially expressed genes in 32D cells overexpressing HOXB4- Δ PRD₂₂₉₋₃₆₀ in comparison to 32D cells overexpressing HOXB4-wt. (C) Venn diagrams showing the overlap between the genes differentially expressed in 32D cells upon overexpression of HOXB4- Δ PRD₂₂₉₋₃₆₀ and HOXB4-wt in comparison to the YFP control. (D) GSEA analysis: comparison of expression signatures (HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt) with published HSC specific signatures. (a) upregulation of early hematopoietic progenitors signature in 32D cells²; (b) downregulation of late hematopoietic signature² in BM progenitors; (c) upregulation of mouse and human embryonic stem cells signature³ and (d) upregulation of the ME14 primitive signature⁴ in 32D cells. (e) Enrichment of genes also down-regulated in leukemic stem cells (LSC), defined as CD34⁺CD38⁻ cells, from AML patients compared to the CD34⁺CD38⁺ cells in BM cells⁵. (f) Negative correlation with HSC-related signature in 16 primary AML patient samples and BM cells⁶; (g) Positive enrichment of LSC maintenance associated genes⁷; (h) GSEA-Leading edge analysis identifying core-enriched genes in normal and leukemic expression signatures, as shown in previous panels, in HOXB4- Δ PRD₂₂₉₋₃₆₀ in comparison to HOXB4-wt in BM cells. (F) GSEA analysis for oncogenic signatures in HOXB4- Δ PRD₂₂₉₋₃₆₀ in comparison to HOXB4-wt. Enrichment plots for *kras/myc*-upregulation and *Sfn5*-knockdown signatures. Leading edge analysis: proportion of differentially expressed genes, which contribute to the enrichment core in (G) BM cells (FDR=1) and (H) 32D cells (FDR values are indicated). (I) Core-enriched upregulated (left) and downregulated (right) genes in HOXB4- Δ PRD₂₂₉₋₃₆₀ BM cells in comparison to HOXB4-wt, which are recurrent in several oncogenic signatures as reported in MSigDB database (Broad Institute).

Fig. S7. (A) Heat maps showing the ChIP-enrichment over gene body for HOXB4-wt (left) and HOXB4- Δ PRD₂₂₉₋₃₆₀ (right). (B) Overlap of targets identified for HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀ in this study with previously published data sets⁸⁻¹⁰. (C) numbers of ChIP targets in each study. (D) Venn diagram showing the overlap between the HOXB4-wt ChIP targets from previous reports⁸⁻¹⁰ and the targets bound by HOXB4- Δ PRD₂₂₉₋₃₆₀, but not by HOXB4-wt in the present study (unique HOXB4- Δ PRD₂₂₉₋₃₆₀ targets are reported in Table S6). (E) Expression of identified direct targets of HOXB4- Δ PRD₂₂₉₋₃₆₀ in normal and malignant human hematopoiesis as reported in human microarray collection in <https://gexs.stanford.edu> (human model: PMID: 21177505/GEO: 24006, Probesets: GART: 212378_at, NUP85: 218014_at, EIF2B3: 218488_at). (F) qPCR analysis after HOXB4-ChIP (monoclonal anti-FLAG M2) in 32D cells for selected direct targets, differentially regulated by HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀. Primers were designed based on the targets identified by the ChIPseq analysis performed using the anti-HOXB4 polyclonal antibody (N-18X). Genomic positions were confirmed using UCSC genome browser. SEM bars are indicated. NC1= negative genomic control 1.

Fig. S8. Motif enrichment analysis on HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀ ChIPseq targets. p values of enrichments are indicated. Analysis was performed using HOMER (JASPAR database). Output results for 50-200bp and 200bp regions were included. *known results (outputs per 200bp or 50-200bp intervals from previously described ChIPseq datasets); °de novo motifs (HOMER outputs per 50-200bp intervals).

Fig. S9. FLAG-IP in 293T cells overexpressing HOXB4- Δ PRD₂₂₉₋₃₆₀, HOXB4-wt or the empty MIY vector control, followed by mass spectrometry proteomics analysis. (A) Log of intensity ratios for each protein in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt (x axis) are plotted versus absolute intensities of single proteins (y axis). Data were normalized for relative amount of HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀. Red dots indicate proteins interacting with HOXB4-

Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt; green dots indicate proteins interacting with HOXB4-wt but not with HOXB4- Δ PRD₂₂₉₋₃₆₀, blue dots indicate proteins interacting with HOXB4- Δ PRD₂₂₉₋₃₆₀ but not with HOXB4-wt vs. (B) Log of intensity ratios for each protein in HOXB4-wt vs. MIY control (x axis) are plotted versus absolute intensities of single proteins (y axis). (C) Log of intensity ratios for each protein in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. MIY control (x axis) are plotted versus absolute intensities of single proteins (y axis). (D) Venn diagram showing differentially and commonly bound proteins. (E) Gene ontology analysis of proteins interacting with both HOXB4- Δ PRD₂₂₉₋₃₆₀ and with HOXB4-wt. Number of genes per category and p-values are indicated. (F) Gene ontology analysis of proteins interacting only with HOXB4-wt. Number of genes per category and p-values are indicated. (G) Gene ontology analysis of proteins interacting only with HOXB4- Δ PRD₂₂₉₋₃₆₀. Number of genes per category and p-values are indicated. (H) Validation of binding by Co-IP for selected proteins as indicated.

Supplemental Tables:

Table S1. CRU Assay – HOXB4

Construct	CRU frequency	p value	Injected cells						
			10K	50K	75K	100K	370K	666K	1300K
GFP	1 : 1,304,178 (1:5,174,254- 1:328,720)	-	0/3	0/3	0/1	0/3	1/1	0/2	1/1
HOXB4-wt	1 : 834 (1:2,240-1:311)	P=0.0001 vs. GFP	200 0/5	2K 5/5	20k 5/5	200k 5/5			
HOXB4- ΔPRD₂₂₉₋₃₆₀	1 : 575 (1:1,670-1:198)	P=0.0001 vs. GFP; P=0.6164 vs. HOXB4-wt	200 1/5	2K 5/5	20K 5/5	200K 3/3			

Table S2. Hematological characteristics of transplanted mice.

Mouse no.	Experimental Groups	Days post transplant	RB Cs x10 ⁹ /ml	WBCs x10 ⁶ /ml	Spleen size (mm)	Spleen weight (mg)	Blasts % PB	Blasts % BM	Blasts % Spleen
1	B4-wt	404	7.75	7.25	n.a.	n.a.	0	14.9	n.a.
2	B4-wt	416	6.8	7.75	n.a.	250	0	11.8	0
3	B4-wt	483	4.4	7	24x7	350	0.9	12.3	0
4	B4-wt	510	6.4	2.5	20x6	231	5	14	0
5	B4-wt	605	0.52	0.37	22x6	334	0	5	1
6*	B4-ΔPRD	310	5.9	19.8	20x8	500	n.a.	50	n.a
7*	B4-ΔPRD	126	n.a.	n.a.	n.a.	n.a.	n.a.	70	n.a.
8*	B4-ΔPRD	331	2.25	17.5	20x5	n.a.	89	93	23.5
9*	B4-ΔPRD	147	0.85	65	30x9	800	93	90	90
10*	B4-ΔPRD	244	2.8	24.3	27x7	750	36.8	30	30
11*	B4-ΔPRD	384	9.8	6.5	17x4	n.a.	27	80	78
12#	B4-ΔPRD-HD	n.a.	8.25	6	n.a.	n.a.	0	5.3	0
13#	B4-ΔPRD-HD	485	17.2 5	3.75	10x4	160	0	6.2	0
14#	B4-ΔPRD-HD	485	9.25	6.5	10x3	150	0	4.8	0
15	GFP	480	8.1	3.25	n.a.	150	0	7.5	0
16	GFP	485	9.95	9	n.a.	200	0	6	0
17	GFP	384	10.9	4	n.a.	200	0	6.30	0
18	B4-ΔPRD-PID	153	5.2	25	n.a.	475	10	50	20
19	B4-ΔPRD-PID	470	8.1	8	n.a.	200	5	22	n.a.
20	B4-ΔPRD-PID	310	2.3	16	n.a.	300	20	30	10

Table S3A. Integration sites analysis by LM-PCR in HOXB4-wt

Mouse gene symbol	Chromosome	Position	Location	Distance in kb
<i>Serinc3*</i>	2qH3	163,642,790- 163,643,105	Intron 1	
<i>Pdia6</i>	12qA1.3	19,243,666-19,243,880	3 prime	1961.8
<i>Bcl9l</i>	9qA5.2	44,494,370-44,494,510	5 prime UTR	
<i>AK144798</i>	2qC2	71,765,030-71,765,230	3 prime	3.1
<i>CD68</i>	11qB3	69,664,109-69,664,416	5 prime over TSS	
<i>F5</i>	1qH2.2	164,228,815- 164,228,852	3 prime	8.54
<i>Ccdc2</i>	4qD3	134,159,390- 134,159,937	Intron 3	
<i>5730507C01Rik</i>	12qA1.2	19,243,628-19,243,889	3 prime	729
<i>Tmem130</i>	5qG2	144,678,680- 144,679,249	3 prime	56.6

Table S3B: Integration sites analysis by LM-PCR in HOXB4-ΔPRD₂₂₉₋₃₆₀.

Mouse gene symbol	Chromosome	Position	Location	Distance in kb
<i>Col4a3bp-Hmgcr*</i>	13qD1	96,644,909-96,644,981	3 prime of both genes	6.3 and 4.7, respectively
<i>Hs6st3-Oxgr1</i>	14qE4	119,950,275-119,950,303	3 prime of both genes	80.4 and 69.3, respectively
<i>Dhx40 and Ypel2</i>	11qC	86,890,577-86,890,795	5 prime of both	82.9 and 45.6, respectively
<i>Camk2b*<i>T</i></i>	11qA1	6,006,251-6,006,408	Intron 3	
<i>Galnt14</i>	7qE3	111,304,761-111,304,785	3 prime	166.87
<i>Olfir370</i>	8qC2	83,531,903-83,531,957	3 prime	9.1
<i>Zfp954</i>	7qA1	7070610-7070647	5 prime	44
<i>Cdk5rap2</i>	4qC2	172,388,335-172,388,366	Intron 5	
<i>Camk2b*<i>T</i></i>	11qA1	6,011,359-6,011,522	Intron 3	
<i>Edaradd</i>	13qA1	12,494,932-12,495,170	Intron 4	
<i>AK144798</i>	2qC2	71,765,030-71,765,230	3 prime	3
<i>CD68</i>	11qB3	69,664,109-69,664,416	5 prime over TSS	
<i>Fgd2</i>	17qA3.3	29,395,292-29,395,638	3 prime	15.75
<i>Rreb1*</i>	13qA3.3	37,856,816-37,856,872	Intron 1	

Table S4. In vivo transplantation.

	Construct (plus Meis1)	No. cells injected	No. mice w AML/cohort	Median survival, days (range)	P value
5-FU BM	Hoxa9-wt	5x10 ⁵	4/4	57 (38-64)	n.s. (vs.scr)
	Hoxa9-scr	5x10 ⁵	4/4	42 (39-46)	--
	Hoxa9-Pro-in ₄₂₃	5x10 ⁵	4/4	73.5 (46-105)	0.016 (vs. Scr)
c-kit ⁺ BM	Hoxa9-wt	2.8 10 ⁶	5/5	39 (33-44)	-
	Hoxa9-Pro-in ₈₁₃	2.8 10 ⁶	5/5	50 (45-57)	0.0018 (vs. wt)

Table S5. CRU Assay – Hoxa9.

Arm	LSC frequency	+/- SEM	P value vs. wt
Hoxa9-wt	1:7226	1:12,933-4,037	-
Hoxa9-scramble	1:4656	1:7,914-2,739	0.5768
Hoxa9-Pro-in₄₂₃	1:55,937	1:102,945-30,394	0.0153

Table S6A: List of top genes differentially expressed in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt BM progenitor cells (p<0.05).

Genes downregulated (Δ PRD vs. wt)	Log Fold Change	Function	Associated diseases (ref.)	Genes upregulated (Δ PRD vs. wt)	Log Fold Change	Function	Associated diseases (ref.)
<i>Olr1</i>	-2.40	Lipid metabolism	Mut. In ovarian Ca. ¹¹	<i>Taf4b</i>	1.23	Member of TBP-associated factor	Cooperates with Oct4 in ESC proliferation ¹²
<i>Hbb-b1</i>	-2.36	hemoglobin, β adult major chain	na	<i>Car8</i>	1.27	Carbonic Anyhdrase	Neurological defects
<i>Marcks</i>	-2.32	Control IP3 in cancer cells	Mut. in gastric adenoca. ¹³	<i>Mmp27</i>	1.27	Matrix metallopeptidase ²⁷	Thyroid cancer, mutated in AML ¹⁴
<i>Ndr2</i>	-2.31	N-myc down-regulated gene ²	Inhibited in AML ¹⁵	<i>Aven</i>	1.27	Caspase activation inhibitor	Protect from apoptosis after NO exposure ¹⁶
<i>Hist2h3c2-ps</i>	-2.28	Histone cluster ²	na	<i>Slc6a9</i>	1.32	Neurotransmitter transposter	Mutated in AML ¹⁴
<i>Mogat2</i>	-2.24	Lipid metabolism	na	<i>Trmt61a</i>	1.36	tRNA methyltransferase	Mutated in sarcoma ¹⁷
<i>Car1</i>	-2.23	Carbonic anhydrase	Subacute thyroiditis	<i>Gpr34</i>	1.39	G protein coupled receptor	Upregulated in non-Hodgkin lymphomas ¹⁸
<i>Scd1</i>	-2.17	Lipid metabolism	Involved in tumor growth ¹⁹	<i>Gpr126</i>	1.42	G protein-coupled receptor	Regulates angiogenesis ²⁰

<i>Nos2</i>	-2.15	Nitric Oxide Synthase 2	Associated with virilizing ovarian tumor	<i>Gbp8</i>	1.46	IFN-induced GTPase	Involved in host defense ²¹
<i>Vcam1</i>	-1.97	Adhesion molecule	Decreased in AML serum ²²	<i>Slc7a11</i>	1.55	Cystein/glutamate exchanger	High expression correlates with poor outcome in bladder cancer ²³
<i>Gsta4</i>	-1.91	Glutathione S-transferase	Modulates chemoresistance in HL60 ²⁴	<i>Acot2</i>	1.58	Acyl-CoA Thioesterase	related to more aggressive phenotype in breast cancer cell lines ²⁵
<i>Rcn3</i>	-1.89	EF-Hand Calcium-binding protein	Mutated in gastric Adenoca. ¹³	<i>Mmp12</i>	1.61	Metalloproteinase12	Promote tumor progression in apoptotic B cell lymphoma ²⁶
<i>Cnn3</i>	-1.82	Calponin3, cytoskeleton associated protein	Translocation partner in MALT ²⁷	<i>Gm129</i>	1.65	Transcriptional repressor	Regulation of circadian clock ²⁸
<i>Itm2a</i>	-1.71	Membrane protein	Mutated in Bladder cancer ²⁹	<i>Tcrg-VI</i>	1.72	T Cell Receptor Gamma Variable 1	Mediates interaction of V γ 1 T cells with activated macrophages ³⁰
<i>Mras</i>	-1.70	GTPase	Stemness marker ³¹	<i>Fam129a</i>	1.76	Translation regulatory protein	Tumor marker in several cancers ³²
<i>B4gal16</i>	-1.68	Beta-1,4-Galactosyltransferase 6	Prostate ca. in situ	<i>Lhcgr</i>	2.00	lutinizing hormone/choriogonadotropin receptor	male secondary sexual character development
<i>Eno2</i>	-1.66	Neural enolase	Malignant granular cell myoblastoma ³³	<i>Ivl</i>	2.13	involucrin	Squamous cell carcinoma ³⁴

<i>5430435G22Rik</i>	-1.65	Ras-related protein Rab-7	Upregulated after PME induced differentiation in HL-60/NB4 ³⁵	<i>Akr1c18</i>	2.44	Aldo/keto reductase	Desensitizes HL-60 to differentiating agents ³⁶
<i>Il12a</i>	-1.64	Natural killer cell stimulatory factor 12a	Modulates response to chemotherapy ³⁷	<i>Lce6a</i>	2.99	Late cornified envelope	Expressed in ovarian and skin cancer (proteinatlas.org/ENSG00000235942-LCE6A/cancer/tissue)

Table S6B. List of top genes differentially expressed in HOXB4- Δ PRD₂₂₉₋₃₆₀ and HOXB4-wt BM progenitor cells in comparison to the GFP control (p<0.05) (see separate Excel file).

Table S6C: List of top genes differentially expressed in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt overexpressing 32D cells (p<0.05).

<i>Genes downregulated (ΔPRD vs. wt)</i>	<i>score</i>	Function	Associated diseases (ref.)	<i>Genes upregulated (ΔPRD vs. wt)</i>	<i>score</i>	Function	Associated diseases (ref.)
<i>Arhgap28</i>	-1.11	Rho GTPase activating protein	Upregulated in meningioma ³⁸	<i>2610305D13RIK</i>	1.98	Ssm1, KRAB-zinc-finger	DNA methylation in mouse embryogenesis ³⁹
<i>Capn2</i>	-1.11	Non-lysosomal cysteine protease	Upregulated in prostate cancer ⁴⁰	<i>Parp8</i>	1.77	Post-translational modification	Potential MLL-AF9 target ⁴¹
<i>Abca1</i>	-1.12	Cholesterol Efflux Regulatory Protein	Target of miR760 in chemoresistant breast cancer ⁴²	<i>Sgk3</i>	1.74	serum/glucocorticoid regulated kinase family, member 3	Target in glioblastoma ⁴³
<i>Padi2</i>	-1.12	post-translational deimination of proteins	overexpression induces apoptosis in activated Jurkat cells ⁴⁴	<i>Mecom</i>	1.74	MDS1 and Evi1 complex locus protein	Upregulated in Evi1 ⁺ AML ⁴⁵
<i>Mier3</i>	-1.13	Chromatin binding protein	Downregulated in colon cancer ⁴⁶	<i>Stat4</i>	1.67	Th1 development	Upregulated by HDAC1 ⁴⁷
<i>Hook1</i>	-1.14	Binds to microtubules and organelles	Splicing variant in MPN ⁴⁸	<i>IL18r1</i>	1.65	Proinflammatory cytokine	Overexpressed in epithelial cells of ovarian carcinoma ⁴⁹

<i>Fzd4</i>	-1.14	Frizzled 4, receptor for Wnt proteins	Downregulated in de novo AML ⁵⁰	<i>Tcfec</i>	1.64	Transcription factor EC	Enhances HSC activity ⁵¹
<i>Otub2</i>	-1.14	deubiquitylating enzyme	na	<i>Mmrn1</i>	1.64	Binds platelet factor V	Bleeding disorders
<i>Fyn</i>	-1.14	Src kinase	Involved in progression of CML ⁵²	<i>IL12rb2</i>	1.60	Interleukine 12 receptor, beta 2	Better outcome in lung adenocarcinoma ⁵³
<i>Pik3c2a</i>	-1.14	Kinase involved in proliferation and transformation	Tagert of miR-509-3p->decreased proliferation ⁵⁴	<i>Fgd4</i>	1.53	Regulation of actin cytoskeleton	Charcot-Marie-Tooth neuropathy ⁵⁵
<i>Klk1b22</i>	-1.15	Serine protease	Maybe involved in carcinogenesis	<i>Tnip3</i>	1.52	Binds to zinc finger protein TNIP3	Inhibits NFkB ⁵⁶
<i>GdpdD1</i>	-1.15	Phosphodiesterase	na	<i>3830403N18RIK</i>	1.51	Sinaptonemal complex protein	Spermatogenic failure
<i>4932412H11RIK</i>	-1.16	Leucine rich repeats and death domain containing 1	na	<i>Fam125b</i>	1.51	Multivescicular body subunit 12B	
<i>Fam164a</i>	-1.20	Zinc Finger, C2HC-Type Containing 1A		<i>Mela</i>	1.51	Melanoma antigen	Block anti-melanoma T-cell reponse
<i>Dusp10</i>	-1.20	MAP kinase phosphatase	Mediates anti cancer activities in prostate ⁵⁷	<i>Akr1c18</i>	1.47	Aldo/keto reductase	Desensitizes HL-60 to differentiating agents ³⁶

<i>Acer3</i>	-1.24	Alkaline ceramidase	Regulates cell proliferation ⁵⁸	<i>Mboat2</i>	1.46	Membrane bound O-acyltransferase	Regulated LTB4 synthesis in neutrophil
<i>CD55</i>	-1.27	Decay accelerating factor for complement	Pancreatic cancer ⁵⁹	<i>St8sia4</i>	1.46	Regulates T-cell proliferation	Near translocation breackpoint in T-ALL ⁶⁰
<i>Lgr4</i>	-1.28	G-protein coupled receptor 4	Prostate cancer ⁶¹	<i>B3gnt5</i>	1.45	Synthesis of glycolipids	Gliblastoma ⁶²
<i>Sccpdh</i> /// <i>LOC1000</i>	-1.29	Saccharopine Dehydrogenase	na	<i>Chn2</i> /// <i>9130019P1</i>	1.44	GTPase activating protein	Hepatosplenic T-cell lymphoma ⁶³
<i>Cdk15</i>	-1.42	Cyclin-dependent kinase-like 5	Neurological disorders	<i>Klhl6</i>	1.43	Involved in B lymphocyte function	Mutated in CLL ⁶⁴

Table S7A-B: HOXB4-wt and HOXB4-ΔPRD targets (ChIPseq) (separate excel file).

Table S8: Identified direct targets of HOXB4-ΔPRD₂₂₉₋₃₆₀ in 32D cells

<i>Directly upregulated targets of HOXB4-ΔPRD₂₂₉₋₃₆₀</i>	<i>Fold change vs. GFP ctrl</i>	Function	Associated disease	Directly downregulated targets of HOXB4-ΔPRD ₂₂₉₋₃₆₀	Fold change vs. GFP ctrl	Function	Associated disease
<i>Ccnb1ip1</i>	2.32	Cyclin B1 interacting protein	Involved in cell cycle	<i>Ccr2</i>	-2.22	C-C chemokine receptor 2	Mediates monocytes chemotaxis
<i>Csgalnact1</i>	2.15	Acetylgalactosa minyl-transferase	Better prognosis in myeloma ⁶⁵	<i>Ccr1</i>	-1.83	C-C chemokine receptor 1	Mediates monocytes chemotaxis
<i>Polr1b</i>	1.89	DNA-deendent RNA-polymerase	Susceptibility to t-AML ⁶⁶	<i>Clec4d</i>	-1.68	Activating receptor on myeloid cells	Antigen processing ⁶⁷
<i>Elovl6</i>	1.74	Fatty acid elongase 6	na	<i>Lilra6</i>	-1.67	Leukocyte immunoglobulin-like receptors	Regulates myelomonocytes activity
<i>Lmo2</i>	1.71	Hematopoietic development	Cause T-ALL and AML if ectopic ⁶⁸	<i>Arrdc3</i>	-1.54	Arrestin family member	Decreases breast cancer progression ⁶⁹
<i>Gart</i>	1.63	Purine biosynthesis	HCC proliferation ⁷⁰	<i>Mfsd6</i>	-1.48	Macrophage MHC Receptor 2	na

<i>Lrpprc</i>	1.48	Mitochondrial gene expression	Pancreas carcinoma proliferation ⁷¹	<i>Dpep2</i>	-1.45	Membrane bound dipeptidase 2	na
<i>Mnd1</i>	1.41	Meiotic recombination	Developmental disorders	<i>Slc6a12</i>	-1.37	Neurotransmitter transporter	na
<i>Eif2b3</i>	1.38	Translation initiation	Neurological disorders	<i>Slc26a3</i>	-1.31	Anion exchanger	Intestinal chloride absorption
<i>Tbc1d22b</i>	1.37	GTPase activating protein	na	<i>Mocos</i>	-1.27	Molybdenum cofactor sulfurase	Potential target of SIRT1 in melanoma ⁷²
<i>Aff3</i>	1.23	Lymphoid nuclear protein related to AF4	Downregulated in trastuzumab-resistant breast cancer ⁷³	<i>Frmd8</i>	-1.25	FERM domain containing b	Correlates with survival in non-lung tissue in adenocarcinoma ⁷⁴
<i>Ampd2</i>	1.22	AMP-deaminase 2	Increased in neoplastic liver tumor cells ⁷⁵	<i>Havcr2</i>	-1.23	Hepatitis A virus receptor	na
<i>Dis3</i>	1.21	RNA turnover	Mutated in del(3)myeloma and AML ⁷⁶	<i>Plxdc2</i>	-1.21	Tumor endothelial marker 7 related protein	Radioresistance related gene in laryngeal carcinoma ⁷⁷
<i>Rab38</i>	1.19	k-ras family member 38	Upregulated in melanoma adn CML ⁷⁸	<i>Rbpj</i>	-1.21	Regulator of Notch signaling	Impairs growth of lung cancer ⁷⁹
<i>Nup85</i>	1.18	Nuclear pore complex component	na	<i>Cd274</i>	-1.19	Programmed cell death 1 ligand 1	Primary testicular lymphoma ⁸⁰

<i>Gpr160</i>	1.15	G-protein coupled receptor	na	<i>Parp10</i>	-1.18	Adds ADP to histone and regulates transcription	Required for DNA damage response
<i>Gmbs</i>	1.14	GDP-mannose-4,6-dehydratase	If deficient, TRAIL-induced apoptosis resistance ⁸¹	<i>Cd300lf</i>	-1.1	Immunoreceptor on myeloid cells	Mediates caspase-independent cell death
<i>Rhoj</i>	1.14	Endothelial enriched Rho GTPase	Promotes tumor progression ⁸² and chemoresistance in melanoma ⁸³	<i>Lrrc51</i>	-1.06	O-methylation of catecholamine neurotransmitters	deafness
<i>Sclt1</i>	1.11	Sodium channel and clathrin linker	At breackpoints in myelofibrosis ⁸⁴	<i>Nedd9</i>	-1.03	Adhesion docking molecule	Stem cell nich integrity ⁸⁵
<i>Vps13c</i>	1.1	Vacuolar protein sorting associated protein	Mutated in pancreatic IPMN ⁸⁶	<i>Utn</i>	-1.02	Putative tumor suppressor	Mutated I multiple tumors ⁸⁷
<i>Fxn</i>	1.08	Mitochondrial frataxin	na	<i>Specc1</i>	-1	Nuclear structure protein 5	Translocation in juvenile myelomonocytic leukemia ⁸⁸
<i>Diexf</i>	1.07	Growth of digestive organs	na				
<i>Pip4k2b</i>	1.07	Phosphate kinase	Overexpressed in subset of breast cancers ⁸⁴				

<i>Pccb</i>	<i>1.06</i>	Catabolism of propionyl-CoA	acidemia				
<i>H2-T24</i>	<i>1.02</i>	Histocompatibility 2, T region locus 24	na				
<i>Ttl4</i>	<i>1.01</i>	Tubulin tyrosine ligase-like family member 4	Promotes growth in pancreatic cancer ⁸⁹				

Table S9. Identified direct targets of HOXB4-wt in 32D cells

Direct downregulated targets of HOXB4-wt	Fold change vs. GFP ctrl	Direct downregulated targets of HOXB4-wt	Fold change vs. GFP ctrl	Direct upregulated targets of HOXB4-wt	Fold change vs. GFP ctrl
<i>Pomt2</i>	-1.1	<i>Rasa4</i>	-1.59	<i>Fyn</i>	4.42
<i>Tmem116</i>	-1.1	<i>Aff3</i>	-1.62	<i>Cd79b</i>	3.36
<i>Sdccag8</i>	-1.1	<i>Rgs3</i>	-1.71	<i>Sgms1</i>	1.7
<i>Fcgr4</i>	-1.1	<i>Evi5</i>	-2.02	<i>Gpr125</i>	1.6
<i>Def8</i>	-1.12	<i>Zfp608</i>	-2.09	<i>H2-Dma</i>	1.27
<i>Snx29</i>	-1.12	<i>Trio</i>	-2.21	<i>Gm14005</i>	1.15
<i>Mtmr7</i>	-1.12	<i>Fam198b</i>	-2.52	<i>Rps6ka5</i>	1.14
<i>Clcn3</i>	-1.2	<i>Cdyl2</i>	-2.64	<i>Trim14</i>	1.14
<i>Pls1</i>	-1.21	<i>Slc24a3</i>	-2.72	<i>Rcsd1</i>	1.11
<i>Pdzd2</i>	-1.21	<i>Elovl7</i>	-2.86	<i>Ical1</i>	1.11
<i>Msra</i>	-1.23	<i>Luzp1</i>	-2.96	<i>Aim1</i>	1.07
<i>Ehbp1</i>	-1.26	<i>Pknx2</i>	-3.34	<i>Tspan3</i>	1.05
<i>Pik3r3</i>	-1.29	<i>Creb5</i>	-3.52	<i>Chchd3</i>	1.04
<i>Ccdc62</i>	-1.31	<i>Mycn</i>	-3.87	<i>Elf2</i>	1.04
<i>Serpinb6b</i>	-1.32	<i>Mecom</i>	-3.98	<i>Phc3</i>	1.03

<i>Acaa2</i>	-1.32	<i>Lair1</i>	-4.02	<i>Glb1</i>	1.02
<i>Pate2</i>	-1.37	<i>Rhoj</i>	-4.17	<i>Zfp35</i>	1.02
<i>Selp</i>	-1.5				
<i>BC055004</i>	-1.56				
<i>Slc2a9</i>	-1.58				

Table S10. (A) Proteins bound to HOXB4 and HOXB4- Δ PRD₂₂₉₋₃₆₀ as determined by mass spectrometry analyses after FLAG co-IP using intensity cut-off to avoid false positive results (see materials and methods)(separate excel file). **(B)** List of all identified proteins in the FLAG-Co-IP (separate excel file).

Legends Supplemental Tables:

Table S1. Limit dilution transplantation (CRU) assay for HOXB4. Frequencies of LT-HSC calculated at week 16 post transplantation. Numbers of injected cells and numbers of long-term engrafted mice per cohort are indicated. () indicate the 95% confidence intervals.

Table S2. Hematological characteristics of transplanted mice. RBCs and WBCs counts, spleen sizes and weights, and proportion of blasts morphologically determined are indicated. * these mice developed AML, and where sacrificed when moribund. # the engraftment was $\leq 15\%$. n.a. = not available.

Table S3. Integration sites analysis by LM-PCR in (A) HOXB4-wt and (B) HOXB4- Δ PRD₂₂₉₋₃₆₀ from primary cells obtained from transplanted mice. *indicates RISs (Retrovirus Integration Sites) as reported in the RTCGD database. † indicates samples obtained from mice transplanted with cells from the same transduction experiment.

Table S4. Summary of the in vivo transplantation experiments performed with the Hoxa9/Meis1 constructs. 5-FU-BM was successfully transduced, highly purified and transplanted with helper cells (5×10^5 /mouse) into lethally irradiated recipients. c-kit⁺ cells were isolated from steady state BM and injected without helper cells into sublethally irradiated recipients. Median survival of transplanted mice is indicated, P-values were calculated using univariate Log-rank (Mantel-Cox) test.

Table S5. Limit dilution transplantation assay for Hoxa9. Frequencies of LSC were calculated 12 weeks after transplantation.

Table S6. (A) Top 38 genes up- and downregulated in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt BM progenitor cells ($p < 0.05$). (B) Genes differentially expressed in HOXB4- Δ PRD₂₂₉₋₃₆₀ and HOXB4-wt BM progenitor cells in comparison to the GFP control ($p < 0.05$), as shown in Fig.

2B. (C) Top 40 genes up- and downregulated in HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt overexpressing 32D cells ($p < 0.05$).

Table S7. (A) Comparison of target genes in 32D cells overexpressing HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀ obtained by ChIP-seq. Peaks alignment was performed as described in Supplemental Materials and Methods. (B) List of unique targets of HOXB4- Δ PRD₂₂₉₋₃₆₀ in comparison to the HOXB4-wt targets identified in the present as well as in previous reports.

Table S8. List of identified direct targets of HOXB4- Δ PRD₂₂₉₋₃₆₀ in 32D cells. All the genes were bound by HOXB4- Δ PRD₂₂₉₋₃₆₀ in the ChIP-seq analysis and were differentially expressed in comparison to the control. In the right columns the differences in the expression level in comparison to the GFP control are indicated.

Table S9. List of identified direct targets of HOXB4-wt in 32D cells. All the genes were bound by HOXB4-wt in the ChIP-seq analysis and were differentially expressed in comparison to the control. In the right columns the differences in the expression level in comparison to the GFP control are indicated.

Table S10. Mass spectrometry analyses after FLAG co-IP performed on 293T derived cells transduced with HOXB4- Δ PRD₂₂₉₋₃₆₀, HOXB4-wt or the empty MIY vector control. (A) List of proteins co-immunoprecipitated only by HOXB4- Δ PRD₂₂₉₋₃₆₀, by both HOXB4-wt and HOXB4- Δ PRD₂₂₉₋₃₆₀, and only by HOXB4-wt. Binding cut-off in comparison to the control was set considering the lowest ratio (HOXB4-wt/MIY control) and its corresponding intensity. (B) Normalized ratio of proteins differentially bound by HOXB4- Δ PRD₂₂₉₋₃₆₀ vs. HOXB4-wt.

Supplemental Material and Methods

Cell lines

The viral conditioned medium (VCM) of ecotropic Phoenix transient packaging cells used to infect GP⁺E86 stable packaging cells and/or primary murine bone marrow cells. 293T and 32D cell lines were used for protein expression analysis, ChIP analyses, and co-IP experiments.

Mice

The mice strains used as bone marrow transplant donors and recipients were either C57BL/6-Ly-Pep3b (Pep3b) or the F1 hybrid of (C57BL/6Ly-Pep3b x C3HeJ)([PepC3] F1, Ly5.1/CD45.1 positive) and hybrid of (C57BL/6J x C3H/HeJ)([B6C3] F1, Ly5.1/CD45.1 negative), respectively. For Hoxa9 experiments, C57BL/6 mice (Taconic©) were used.

Mice and retroviral infection of primary bone marrow progenitor cells

Mice were bred and maintained according to institutional guidelines at the animal facility in Helmholtz Zentrum in Munich and the Animal Facility of the University of Ulm. Primary 5-FU-enriched BM progenitor cells were obtained and transduced as previously described⁸⁸. Alternatively, c-kit⁺ enriched BM progenitor cells were used.

In vitro and ex vivo proliferation and CFC assays

Freshly transduced BM cells were cultivated in DMEM supplemented with 15% FBS, 1% Pen/Strep, rmIL3, rmIL6, and rmSCF. CFC assay was performed by plating 500 and 1x10⁵ cells in methylcellulose supplemented with rmIL3, rmIL6, rhEpo (MethoCult M3434®, StemCell Technologies, BC, Canada), and the colony numbers were scored after 7-10 days. For immunostaining PE- or APC-conjugated antibodies were used: anti-Sca-1, -Gr-1, -Mac-1, -Ter119, -B220, -CD4, -CD8, -c-kit, -Flk-2, and -Thy1.1 (BD Bioscience and eBioscience, CA, USA).

BM transplantation and assessment of mice

5-FU-enriched BM progenitor cells were isolated and transduced in vitro. Alternatively, c-kit⁺ BM progenitor cells were enriched using CD117-Microbeads Mouse (MACS®, Miltenyi Biotec, Germany). Target cells were retrovirally transduced with MSCV/EGFP vectors overexpressing HOXB4 and Hoxa9 and their mutants. For Hoxa9 experiments we co-transduced with MSCV/YFP vector encoding *Meis1*; GFP⁺ and GFP⁺/YFP⁺ cells were sorted (SoloAria2, BD Bioscience) and injected intravenously into ≥ 8 weeks old syngeneic recipient mice irradiated with 850 cGy. 2×10^4 to 10^6 GFP⁺ cells were transplanted in combination with 3×10^5 - 2×10^6 GFP⁻ cells. For Hox9-wt and Hoxa9-Pro-in₈₁₃ experiments c-kit⁺ cells were expanded in vitro for 2 weeks and 2.8 Mio cells were injected into sublethally irradiated (600 cGy) syngeneic recipient mice. Engraftment and lineage differentiation were evaluated after 4, 8, 12, and 16 weeks post-transplantation by bleeding and/or BM aspiration from the femur. Peripheral blood (PB) smears and cytopins were stained with May-Grünwald-Giemsa stains and microscopically analyzed. Moribund mice were sacrificed and cells from BM, spleen and PB were analyzed for GFP and surface marker expression. RNA/DNA were purified following standard protocols. BM and spleen cells were injected into lethally irradiated secondary and tertiary recipient mice (1×10^6 cells/mouse). The organs were fixed with formaldehyde (10% v/v) and prepared for histological and immunohistochemical examinations. The classification of disease was performed following the criteria of the Bethesda proposal for the classification for non-lymphoid hematopoietic neoplasms in mice⁹⁰.

Competitive repopulating unit (CRU assay) and limit dilution transplantations of leukemic cells

BM cells from 5-FU primed donor mice were retrovirally transduced and sorted for GFP expression. Serial limiting dilutions (2×10^2 - 2×10^5 for HOXB4 constructs, 10^4 - 1.3×10^6 for GFP control) were injected into lethally irradiated (850 cGy) recipient mice together with 3×10^5

competitor BM cells. Engraftment was defined as positive, when greater than 1% donor-derived (GFP⁺) cells in both lymphoid and myeloid subpopulations were detected in the peripheral blood of mice after more than 16 weeks post-injection⁹¹. Multilineage hematopoietic reconstitution was evaluated by FACS analysis using the antibodies described above. The frequency of CRUs was estimated using the LDA software L-CalculTM (Stemsoft Software Inc., StemCell, BC, Canada). For quantification of leukemic stem cell frequencies leukemic cells of the different experimental groups were injected into cohorts of 5 lethally irradiated mice for each serial dilution (50, 500, 5000, 50.000 cells per mouse). The number of mice succumbed to leukemia were evaluated 12 weeks post transplant, the LSC frequency was determined by the LDA software L-CalculTM (Stemsoft Software Inc., StemCell, BC, Canada).

Expression profiling and Western blotting

Expression profile analysis was performed as follows: 400 ng of RNA were obtained from 32D cell lines and 5-FU BM cells, freshly transduced with pGFP, pHOXB4-wt, pHOXB4-ΔPRD retroviral vectors. RNA was purified using the Qiagen-RNA isolation kit (Qiagen, MD, USA). Sense strand of cDNA was synthesized, labeled and hybridized on GeneChip® Mouse Gene 1.0 ST Arrays (Affymetrix, CA, USA). Data analysis was performed using the limma software package (Bioconductor). Set Enrichment Analysis was performed using the GSEA Software⁹². The following primary antibodies were used for Western blotting and co-immunoprecipitation (Co-IP): goat polyclonal anti-HOXB4 (N-18, Santa Cruz, CA, USA), rat monoclonal anti-HOXB4 I12 (Iowa, USA), mouse monoclonal anti-FLAG M2 (Sigma, MO), rabbit polyclonal anti-actin HRP (I-19, Santa Cruz, CA, USA), goat polyclonal anti-Meis1/2 (C-17, Santa Cruz), rabbit polyclonal anti-Sin3A (AK-11, Santa Cruz), rabbit polyclonal anti-PCNA (FL-261, Santa Cruz).

Immunostaining and confocal laser scanning fluorescence microscopy

For intracellular localization study, 293T cells were grown on coverslips and transfected with pMSCV-IRES-GFP, pMSCV-HOXB4-IRES-GFP, and pMSCV-HOXB4-ΔPRD-IRES-GFP plasmids as previously described⁹³. After 24 h the cells were fixed with PBS, 2% paraformaldehyde for 10 min., permeabilized with PBS 0.1% Triton X for 10 min. and blocked with PBS 10% FCS for 1 h. The coverslips were incubated with monoclonal mouse anti-FLAG antibody (Sigma, MO, USA) or alternatively with polyclonal anti-HOXB4 antibody (Santa Cruz, CA, USA). Alexa 555- and Alexa 488-conjugated secondary antibodies (Invitrogen, CA, USA) were added and incubated for 1 h, followed by DAPI staining. Immunostained samples were analyzed in a confocal fluorescence laser scanning system (TCS-SP2 scanning system and DM IRB inverted microscope, Leica, Solms, Germany). For intracellular staining to detect HA-tagged Hoxa9 proteins rabbit polyclonal anti-HA antibody (GTX115044, GeneTex, USA) was used. As secondary antibody goat anti-rabbit IgG conjugated with Alexa Fluor®555 was used (Thermo Scientific).

Southern blot and LM-PCR

The retroviral integration analysis was performed using gDNA from BM, spleen and PB cells from transplanted mice. Southern blot was performed using standard protocols. Briefly, gDNA was digested with *Pst*I, which cuts the proviral DNA once. After electrophoresis and denaturation, DNA was transferred overnight on zeta-Probe GT membrane (BIO-RAD, Ca, USA) followed by UV-cross-linking. The membrane was then incubated at 62°C with a GFP probe labeled with α -³²P dCTP using Megaprime DNA labeling system (Amersham, PA, USA). Subsequently, the membrane was exposed to an x-ray film for 24 h to one week. The screening for retroviral integration sites were performed as previously described⁹⁴. The provirus flanking sequences from the genomic location were identified by BLAST search at the University of California Santa Cruz (UCSC) genome project website (<http://genome.ucsc.edu>).

ChIP and ChIP-seq

ChIP-seq was performed using the ChIP-IT® Express Kit (Active Motifs, CA, USA) according to the supplier's directions. *HOXB4-wt*, *HOXB4-ΔPRD* and *GFP* alone were retrovirally overexpressed in 32D cells. After crosslinking with 1% formaldehyde for 10 min, chromatin was fragmented using a S220 adaptive focused acoustics (AFA) system (Covaris, MA, USA). Samples were immunoprecipitated with a polyclonal antibody directed against HOXB4 (N-18X, Santa Cruz, TX, USA), which recognizes the HOXB4-ΔPRD mutant as well. Normal IgG (ab37373, Abcam, UK) was used as a negative control. Specific enrichment was monitored by real time-quantitative PCR using primers against the known HOXB4 target genes *RUNX1* and *HEMGN*. ChIP-seq targets were confirmed in independent experiments using anti-FLAG monoclonal (SIGMA) antibodies and performing qChIP-PCR using SYBR-AB technology with primers designed from the peaks obtained by the ChIPseq analysis. Primers sequences are available upon request. Library preparation was performed using the NEBNext DNA Library Preparation Kit (NEB, MA, USA) from 10 ng of ChIP-enriched DNA. Following End-Repair and dA-Tailing, adaptors were ligated to DNA fragments, which were subsequently size selected (175-225bp). DNA fragments were subjected to limited PCR amplification (18 cycles) and quality control was made by running the PCR products on a Bioanalyzer (Agilent Technologies, CA, USA). Finally, cluster generation and sequencing-by-synthesis (56bp) was performed using the Illumina HiScanSQ™ according to standard protocols of the manufacturer (Illumina, CA, USA). The raw sequencing data were processed using the Illumina CASAVA 1.8.2 software to extract DNA sequence data. Sequences were aligned to the murine mm9 reference genome using the BWA (Burrows-Wheeler Alignment) algorithm. Only tags uniquely aligning to the genome were considered for further analysis. HOXB4 peaks were detected using CisGenome v2 and resulting peak regions were tested for relative HOXB4 tag density (tag density at peak divided by total number of tags at all peaks). Motif enrichment analysis was performed using HOMER. Briefly, output files were obtained considering 50-

200bp and 200 bp overlapping intervals. P value was calculated automatically by the HOMER software.

FLAG Co-IP and Mass Spectrometry

FLAG Co-IP of tagged HOXB4 proteins was performed following the manufacturer's instructions. Nuclei lysate of stable transduced 293T-derived cells were prepared and after dounce homogenization nuclei were separated and lysed in IPH solution, incubated with Benzonase and exposed to FLAG-M2 affinity gel (SIGMA) overnight. After washing, co-immunoprecipitated proteins were eluted by incubating with 3xFLAG peptide (SIGMA) following manufacturer's instructions. For co-IP experiments western blot analyses were performed following standard procedures. For MS analysis equal amounts of proteins were then separated on agarose gel (10%), fixed and stained with Blue Comassie following standard procedures. Peptides were extracted following standard procedures and samples were measured using a LTQ Orbitrap Velos Pro system (Thermo Fisher Scientific, Bremen, Germany) online coupled to an U3000 RSLCnano (Thermo Fisher Scientific, Idstein, Germany) employing an Acclaim® PepMap™ analytical column (Thermo Fisher Scientific, Bremen, Germany) at a flow rate of 250 nl/min. Samples were preconcentrated on a C18 nano trap column (Thermo Fisher Scientific, Bremen, Germany) and subsequently separated using a binary solvent gradient. The online coupled mass spectrometer was equipped with a nanoelectrospray ion source and distal coated SilicaTips (FS360-20-10-D, New Objective, Woburn, USA) were used for ionization. XCalibur 2.2 SP1.48 (Thermo Fisher Scientific, Bremen, Germany) was used as the software platform for data-dependent MS/MS analyses. Full scans ranging from m/z 370 to 1,700 were acquired in the Orbitrap part of the instrument at a resolution setting of 30,000. Ten multiple-charged peptide ions were selected from each survey scan and MS/MS fragmentation

was performed in the linear ion trap (LIT). Resulting fragments were detected in the LIT using the rapid scan setting.

MS data analysis and statistics

Both, database search and quantitative analysis, were performed using MaxQuant Ver. 1.5.2.8 (www.maxquant.org)⁹⁵. For peptide identification, MS/MS spectra were correlated with the UniProt human reference proteome set (www.uniprot.org) employing the build-in Andromeda search engine⁹⁶. Carbamidomethylated cysteine was considered as a fixed modification along with oxidized methionine and acetylated protein N-termini as variable modifications. False Discovery rates were set on both, peptide and protein level, to 0.01. Label-free quantitation was enabled using a LFQ min ratio count of 1. To allow for visualization of proteins only found in a single sample, zero values were replaced with the minimum non-zero intensity found in the respective experiment. To limit the number of false-positive identifications of interaction partners, the intensity cut-off was based on the intensity of the protein with the lowest observed ratio, while the ratio cut-off was the inverse ratio of the protein having the second lowest ratio while also exceeding the intensity cut-off. These cut-offs were calculated individually for the HOXB4-WT and HOXB4-ΔPRD arms.

Statistical and sequences analysis

Evaluation of the data was performed using a non-paired *t*-test (GraphPad Prism 6 and Microsoft EXCEL). Differences between means were evaluated using Student's *t*-test, where *p* values <0.05 were considered statistically significant. The calculation of the competitive repopulating unit frequencies was performed using the L-CalcTM software (StemCells, BC, Canada). Survival curves were compared using the Log-rank (Mantel-Cox) test. Multiple sequences alignments were performed using CLUSTAL O (1.1.0). The phylogenetic tree was

generated using the maximum likelihood approach based on JTT matrix-based model. Branch lengths are measured in the number of amino acid substitutions per site.

Supplemental References:

1. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005;21(16):3433-3434.
2. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. *Science*. 2002;298(5593):601-604.
3. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell*. 2008;2(4):333-344.
4. McKinney-Freeman S, Cahan P, Li H, et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell*. 2012;11(5):701-714.
5. Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature medicine*. 2011;17(9):1086-1093.
6. Gal H, Amariglio N, Trakhtenbrot L, et al. Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia*. 2006;20(12):2147-2154.
7. Somervaille TC, Matheny CJ, Spencer GJ, et al. Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell*. 2009;4(2):129-140.
8. Lee HM, Zhang H, Schulz V, Tuck DP, Forget BG. Downstream targets of HOXB4 in a cell line model of primitive hematopoietic progenitor cells. *Blood*. 2010;116(5):720-730.
9. Oshima M, Endoh M, Endo TA, et al. Genome-wide analysis of target genes regulated by HoxB4 in hematopoietic stem and progenitor cells developing from embryonic stem cells. *Blood*. 2011;117(15):e142-150.
10. Fan R, Bonde S, Gao P, et al. Dynamic HoxB4-regulatory network during embryonic stem cell differentiation to hematopoietic cells. *Blood*. 2012;119(19):e139-147.
11. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011.
12. Bahat A, Kedmi R, Gazit K, Richardo-Lax I, Ainbinder E, Dikstein R. TAF4b and TAF4 differentially regulate mouse embryonic stem cells maintenance and proliferation. *Genes Cells*. 2013;18(3).
13. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202-209.
14. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine*. 2013;368(22):2059-2074.
15. Tschan MP, Shan D, Laedrach J, et al. NDRG1/2 expression is inhibited in primary acute myeloid leukemia. *Leukemia Research*. 2010;34(3).
16. Starikova EG, Tashireva LA, Novitsky VV, Ryazantseva NV. Nitric oxide donor NOC-5 increases XIAP and Aven level in Jurkat cells. *Cell Biol Int* 2014;38(7).
17. Barretina J, Taylor BS, Banerji S, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nature genetics*. 2010;42(8):715-721.

18. Ansell SM, Akasaka T, McPhail E, et al. t(X;14)(p11;q32) in MALT lymphoma involving GPR34 reveals a role for GPR34 in tumor cell growth. *Blood*. 2012;120(19).
19. Sun Y, He W, Luo M, et al. SREBP1 regulates tumorigenesis and prognosis of pancreatic cancer through targeting lipid metabolism. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2015.
20. Cui H, Wang Y, Huang H, et al. GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. *J Biol Chem* 2014;289(50).
21. Degrandi D, Konermann C, Beuter-Gunia C, et al. Extensive characterization of IFN-induced GTPases mGBP1 to mGBP10 involved in host defense. *Journal of immunology*. 2007;179(11).
22. Horacek JM, Kupsa T, Vasatova M, Jebavy L, Zak P. Biochip array technology and evaluation of serum levels of multiple cytokines and adhesion molecules in patients with newly diagnosed acute myeloid leukemia. *Experimental Oncology*. 2014;36(1).
23. Drayton RM, Dudziec E, Peter S, et al. Reduced expression of miRNA-27a modulates cisplatin resistance in bladder cancer by targeting the cystine/glutamate exchanger SLC7A11. *Clinical Cancer Research*. 2014;20(7).
24. Sharma R, Ellis B, Sharma A. Role of alpha class glutathione transferases (GSTs) in chemoprevention: GSTA1 and A4 overexpressing human leukemia (HL60) cells resist sulforaphane and curcumin induced toxicity. *Phytotherapy Research* 2011;25(4).
25. Maloberti PM, Duarte AB, Orlando UD, et al. Functional interaction between acyl-CoA synthetase 4, lipoxygenases and cyclooxygenase-2 in the aggressive phenotype of breast cancer cells. *PLoS One*. 2010;5(11).
26. Ford CA, Petrova S, Pound JD, et al. Oncogenic properties of apoptotic tumor cells in aggressive B cell lymphoma. *Current Biology*. 2015;25(5).
27. Vinatzer U, Gollinger M, Müllauer L, Raderer M, Chott A, Streubel B. Mucosa-associated lymphoid tissue lymphoma: novel translocations including rearrangements of ODZ2, JMJD2C, and CNN3. *Clinical Cancer Research*. 2008;14(20).
28. Goriki A, Hatanaka F, Myung J, et al. A novel protein, CHRONO, functions as a core component of the mammalian circadian clock. *PLoS Biology*. 2014;12(4).
29. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315-322.
30. Dalton JE, Howell G, Pearson J, Scott P, Carding SR. Fas-Fas ligand interactions are essential for the binding to and killing of activated macrophages by gamma delta T cells. *Journal of Immunology* 2004;173(6).
31. Mathieu ME, Faucheux C, Saucourt C, et al. MRAS GTPase is a novel stemness marker that impacts mouse embryonic stem cell plasticity and Xenopus embryonic cell fate. *Development*. 2013;140(16):3311-3322.
32. Matsumoto F, Fujii H, Abe M, et al. A novel tumor marker, Niban, is expressed in subsets of thyroid tumors and Hashimoto's thyroiditis. *Human Pathology* 2006;73(12).

33. Wang J, Zhu XZ, Zhang RY. Malignant granular cell tumor: a clinicopathologic analysis of 10 cases with review of literature. *Zhonghua Bing Li Xue Za Zhi*. 2004;33(6).
34. Brambilla C, Laffaire J, Lantuejoul S, et al. Lung squamous cell carcinomas with basaloid histology represent a specific molecular entity. *Clinical Cancer Research*. 2014;20(22).
35. Yang M, Chen T, Han C, Li N, Wan T, Cao X. Rab7b, a novel lysosome-associated small GTPase, is involved in monocytic differentiation of human acute promyelocytic leukemia cells. *Biochemical and Biophysical Research Communications*. 2004;318(3).
36. Desmond JC, Mountford JC, Drayson MT, et al. The aldo-keto reductase AKR1C3 is a novel suppressor of cell differentiation that provides a plausible target for the non-cyclooxygenase-dependent antineoplastic actions of nonsteroidal anti-inflammatory drugs. *Cancer Research*. 2003;63(2).
37. Ruffell B, Chang-Strachan D, Chan V, et al. Macrophage IL-10 blocks CD8+ T cell-dependent responses to chemotherapy by suppressing IL-12 expression in intratumoral dendritic cells. *Cancer Cell*. 2014;26(5).
38. Fèvre-Montange M, Champier J, Durand A, et al. Microarray gene expression profiling in meningiomas: differential expression according to grade or histopathological subtype. *International journal of oncology*. 2009;35(6).
39. Ratnam S, Engler P, Bozek G, et al. Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. *Development*. 2014;141(10).
40. Liu T, Mendes DE, Berkman CE. Prolonged androgen deprivation leads to overexpression of calpain 2: implications for prostate cancer progression. *International journal of oncology*. 2014;44(2).
41. Fleischmann KK1 PP, Schmid I, Roscher AA. RNAi-mediated silencing of MLL-AF9 reveals leukemia-associated downstream targets and processes. *Molecular cancer therapeutics*. 2014.
42. Lv J, Fu Z, Shi M, et al. Systematic analysis of gene expression pattern in has-miR-760 overexpressed resistance of the MCF-7 human breast cancer cell to doxorubicin. *Biomedicine & pharmacotherapy*. 2015.
43. Liu H, Li C, Shen C, et al. MiR-212-3p inhibits glioblastoma cell proliferation by targeting SGK3. *Journal of neuro-oncology*. 2015.
44. Hsu PC, Liao YF, Lin CL, Lin WH, Liu GY, Hung HC. Vimentin is involved in peptidylarginine deiminase 2-induced apoptosis of activated Jurkat cells. *Molecules and cells*. 2014;37(5).
45. Lavalley VP, Gendron P, Lemieux S, D'Angelo G, Hebert J, Sauvageau G. EVI1-rearranged acute myeloid leukemias are characterized by distinct molecular alterations. *Blood*. 2015;125(1):140-143.
46. Pitule P, Vycital O, Bruha J, et al. Differential expression and prognostic role of selected genes in colorectal cancer patients. *Anticancer Research*. 2013;33(11).
47. Litvinov IV, Cordeiro B, Fredholm S, et al. Analysis of STAT4 expression in cutaneous T-cell lymphoma (CTCL) patients and patient-derived cell lines. *Cell Cycle*. 2014;13(18).

48. Spinelli R, Pirola A, Redaelli S, et al. Identification of novel point mutations in splicing sites integrating whole-exome and RNA-seq data in myeloproliferative diseases. *Molecular genetics & genomic medicine*. 2013;1(4):246-259.
49. Medina L, Rabinovich A, Piura B, Dyomin V, Levy RS, Huleihel M. Expression of IL-18, IL-18 binding protein, and IL-18 receptor by normal and cancerous human ovarian tissues: possible implication of IL-18 in the pathogenesis of ovarian carcinoma. *Mediators of Inflammation*. 2014.
50. Zang S, Liu N, Wang H, et al. Wnt signaling is involved in 6-benzylthioinosine-induced AML cell differentiation. *BMC cancer*. 2014.
51. Deneault E, Wilhelm BT, Bergeron A, Barabé F, Sauvageau G. Identification of non-cell-autonomous networks from engineered feeder cells that enhance murine hematopoietic stem cell activity. *Experimental Hematology*. 2013;41(5).
52. Singh MM, Howard A, Irwin ME, et al. Expression and activity of Fyn mediate proliferation and blastic features of chronic myelogenous leukemia. *Plos One*. 2012;7(12).
53. Suzuki K, Kadota K, Sima CS, et al. Clinical impact of immune microenvironment in stage I lung adenocarcinoma: tumor interleukin-12 receptor β 2 (IL-12R β 2), IL-7R, and stromal FoxP3/CD3 ratio are independent predictors of recurrence. *Journal of clinical oncology*. 2013;31(4).
54. Yoon S, Han E, Choi YC, et al. Inhibition of cell proliferation and migration by miR-509-3p that targets CDK2, Rac1, and PIK3C2A. *Molecules and cells*. 2014;37(4).
55. Tazir M, Bellatache M, Nouioua S, Vallat JM. Autosomal recessive Charcot-Marie-Tooth disease: from genes to phenotypes. *Journal of the peripheral nervous system* 2013;18(2).
56. Wullaert A, Verstrepen L, Van Huffel S, et al. LIND/ABIN-3 is a novel lipopolysaccharide-inducible inhibitor of NF-kappaB activation. *the journal of biological chemistry*. 2007;282(1).
57. Krishnan AV, Feldman D. Molecular pathways mediating the anti-inflammatory effects of calcitriol: implications for prostate cancer chemoprevention and treatment. *Endocrine-related cancer*. 2010;17(1):R19-38.
58. Hu W, Xu R, Sun W, et al. Alkaline ceramidase 3 (ACER3) hydrolyzes unsaturated long-chain ceramides, and its down-regulation inhibits both cell proliferation and apoptosis. *The Journal of biological chemistry*. 2010;285(11):7964-7976.
59. He Z, Wu H, Jiao Y, Zheng J. Expression and prognostic value of CD97 and its ligand CD55 in pancreatic cancer. *Oncology letters*. 2015;9(2):793-797.
60. Kjeldsen E, Roug AS. A novel unbalanced de novo translocation der(5)t(4;5)(q26;q21.1) in adult T-cell precursor lymphoblastic leukemia. *Molecular cytogenetics*. 2012;5(1):21.
61. Liang F, Yue J, Wang J, et al. GPCR48/LGR4 promotes tumorigenesis of prostate cancer via PI3K/Akt signaling pathway. *Medical oncology*. 2015;32(3):49.
62. Etcheverry A, Aubry M, de Tayrac M, et al. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics*. 2010;11:701.

63. Finalet Ferreira J, Rouhigharabaei L, Urbankova H, et al. Integrative genomic and transcriptomic analysis identified candidate genes implicated in the pathogenesis of hepatosplenic T-cell lymphoma. *PLoS One*. 2014;9(7):e102977.
64. Sutton LA, Ljungstrom V, Mansouri L, et al. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica*. 2015;100(3):370-376.
65. Agenelli L, Forcato M, Ferrari F, et al. The reconstruction of transcriptional networks reveals critical genes with implications for clinical outcome of multiple myeloma. *Clin Cancer Research*. 2011;17(23).
66. Cahan P, Graubert T. Integrated genomics of susceptibility to alkylator-induced leukemia in mice. *BMC Genomics*. 2010;11.
67. Graham LM, Gupta V, Schafer G, et al. The C-type lectin receptor CLECSF8 (CLEC4D) is expressed by myeloid cells and triggers cellular activation through Syk kinase. *The Journal of biological chemistry*. 2012;287(31):25964-25974.
68. Calero-Nieto F, Joshi A, Bonadies N, et al. HOX-mediated LMO2 expression in embryonic mesoderm is recapitulated in acute leukaemias. *Oncogene*. 2013;32(48).
69. Draheim KM, Chen HB, Tao Q, Moore N, Roche M, Lyle S. ARRDC3 suppresses breast cancer progression by negatively regulating integrin beta4. *Oncogene*. 2010;29(36):5032-5047.
70. Cong X, Lu C, Huang X, et al. Increased expression of glycinamide ribonucleotide transformylase is associated with a poor prognosis in hepatocellular carcinoma, and it promotes liver cancer cell proliferation. *Hum Pathol*. 2014;45(7).
71. Zhou J, Zhang F, Hou X, Zhang N. Downregulation of LRPPRC induces apoptosis in prostate cancer cells through the mitochondria-mediated pathway. *Cancer Biother Radiopharm*. 2014;29(9).
72. Singh CK, George J, Nihal M, Sabat G, Kumar R, Ahmad N. Novel downstream molecular targets of SIRT1 in melanoma: a quantitative proteomics approach. *Oncotarget*. 2014;5(7):1987-1999.
73. von der Heyde S, Wagner S, Czerny A, et al. mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer. *PLoS One*. 2015;10(2).
74. Galvan A, Frullanti E, Anderlini M, et al. Gene expression signature of non-involved lung tissue associated with survival in lung adenocarcinoma patients. *Carcinogenesis*. 2013;34(12):2767-2773.
75. Szydlowska M, Chodorowski Z, Rybakowska I, Nagel-Starczynowska G, Kaletha K. Full-size form of human liver AMP-deaminase? *Mol Cell Biochem*. 2004.
76. Walker BA, Wardell CP, Melchor L, et al. Intracлонаl heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. *Blood*. 2012;120(5).
77. Kim JS, Kim SY, Lee M, Kim SH, Kim SM, Kim EJ. Radioresistance in a human laryngeal squamous cell carcinoma cell line is associated with DNA methylation changes and topoisomerase II alpha. *Cancer biology & therapy*. 2015:0.

78. Biernacki MA, Marina O, Zhang W, et al. Efficacious immune therapy in chronic myelogenous leukemia (CML) recognizes antigens that are expressed on CML progenitor cells. *Cancer Research*. 2010;70(3).
79. Lv Q, Shen R, Wang J. RBPJ inhibition impairs the growth of lung cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2015.
80. Twa DD, Mottok A, Chan FC, et al. Recurrent genomic rearrangements in primary testicular lymphoma. *The Journal of pathology*. 2015.
81. Moriwaki K, Shinzaki S, Miyoshi E. GDP-mannose-4,6-dehydratase (GMDS) deficiency renders colon cancer cells resistant to tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor- and CD95-mediated apoptosis by inhibiting complex II formation. *J Biol Chem*. 2011;286(50).
82. Kim C, Yang H, Fukushima Y, et al. Vascular RhoJ is an effective and selective target for tumor angiogenesis and vascular disruption. *Cancer Cell* 2014;25(1).
83. Ho H, Aruri J, Kapadia R, Mehr H, White MA, Ganesan AK. RhoJ regulates melanoma chemoresistance by suppressing pathways that sense DNA damage. *Cancer Research* 2012;72(21).
84. Luoh SW, Venkatesan N, Tripathi R. Overexpression of the amplified Pip4k2beta gene from 17q11-12 in breast cancer cells confers proliferation advantage. *Oncogene*. 2004;23(7).
85. Peng H, Park JK, Katsnelson J, et al. microRNA-103/107 Family Regulates Multiple Epithelial Stem Cell Characteristics. *Stem Cells*. 2015.
86. Furukawa T, Kuboki Y, Tanji E, et al. Whole-exome sequencing uncovers frequent GNAS mutations in intraductal papillary mucinous neoplasms of the pancreas. *Sci Rep* 2011;1.
87. Li Y, Huang J, Zhao YL, et al. UTRN on chromosome 6q24 is mutated in multiple tumors. *Oncogene*. 2007;26(42):6220-6228.
88. Morerio C, Acquila M, Rosanda C, et al. HCMOGT-1 is a novel fusion partner to PDGFRB in juvenile myelomonocytic leukemia with t(5;17)(q33;p11.2). *Cancer research*. 2004;64(8):2649-2651.
89. Kashiwaya K, Nakagawa H, Hosokawa M, et al. Involvement of the tubulin tyrosine ligase-like family member 4 polyglutamylase in PELP1 polyglutamylation and chromatin remodeling in pancreatic cancer cells. *Cancer Research*. 2010;70(10).
90. Kogan SC, Ward JM, Anver MR, et al. Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood*. 2002;100(1):238-245.
91. Antonchuk J, Sauvageau G, Humphries RK. HOXB4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation. *Exp Hematol*. 2001;29(9):1125-1134.
92. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(43):15545-15550.

93. Greif PA, Tizazu B, Krause A, Kremmer E, Bohlander SK. The leukemogenic CALM/AF10 fusion protein alters the subcellular localization of the lymphoid regulator Ikaros. *Oncogene*. 2008;27(20):2886-2896.
94. 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.
95. 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.
96. 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.

Alignment of HOXB cluster

HOXB1 GYPYPQHPYPYGN-----QTASFAPAYAD---L-----LSEDKETPC---PSEPN-TPT
HOXB2 -----R-----AEDGPALPPPPPPPLPAAPPA
HOXB3 TSNSSNG-----GGPSKSGPPKCGPGTNSTLTk
HOXB4 -----PPCAQNPPLHPSPSHSACKEP
HOXB5 -----SSPS-----LARAQPEPMATSTAAPEGQTP
HOXB6 -----FGETEEQKCS
HOXB7 -----E---QRDSDLAAESNF
HOXB8 -----K-----LAAASGLGEEAEGSEQSPSPST
HOXB9 -----G-----QAAVKAEPFLGAPGELLKQGT
HOXB13 AAYPAETPTAGEEYPSRPTTEFAFYPGYPGTYOPMASYLDVSVVOTLGAPGEPRHDSLLPV

Alignment of oncogenic HOXA proteins

HOXB4 AVSSSPPPPCAQN-----PLHPSPSHSA
HOXA7 DLAGG-ACDKTDEG-----ALHG---AA
HOXA10 DLPALAGSADAARKERALDSPPPPTLACGSGGGSQGDEAHASSSAE
HOXA9 GIKPEPLSARRGDGP-----TLDTHTLSTLDYACGSP

Scale chr17: 46,652,500 | 46,653,000 | 46,653,500 | 1 kb | 46,654,000 | 46,654,500 | hg19 | 46,655,000 | 46,655,500 | 46,656,000 | 46,656,500 |

HOXB4 UCSC Genes (RefSeq, GenBank, CCDC, RefSeq, RefSeq & Comparative Genomes)

100 vertebrates Basewise Conservation by PhyloP

Proline-rich region

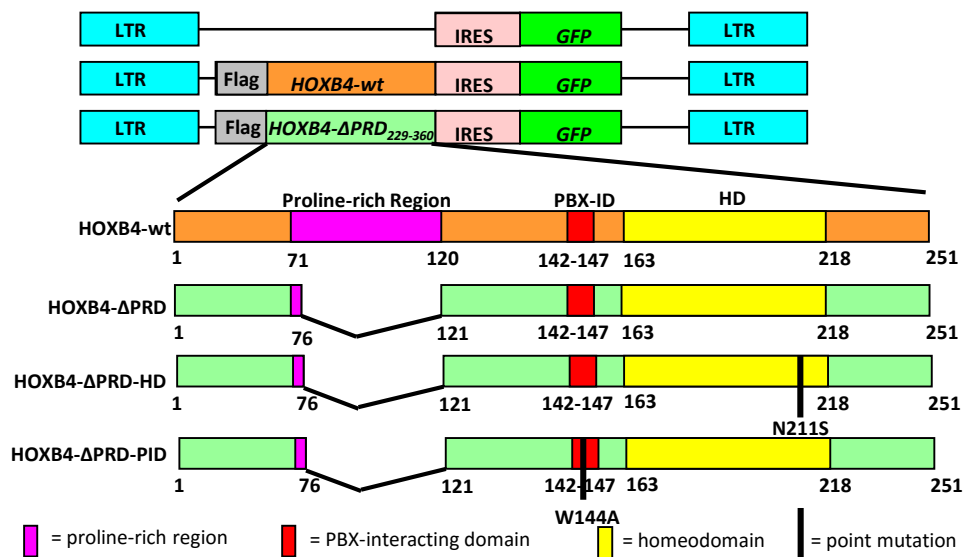
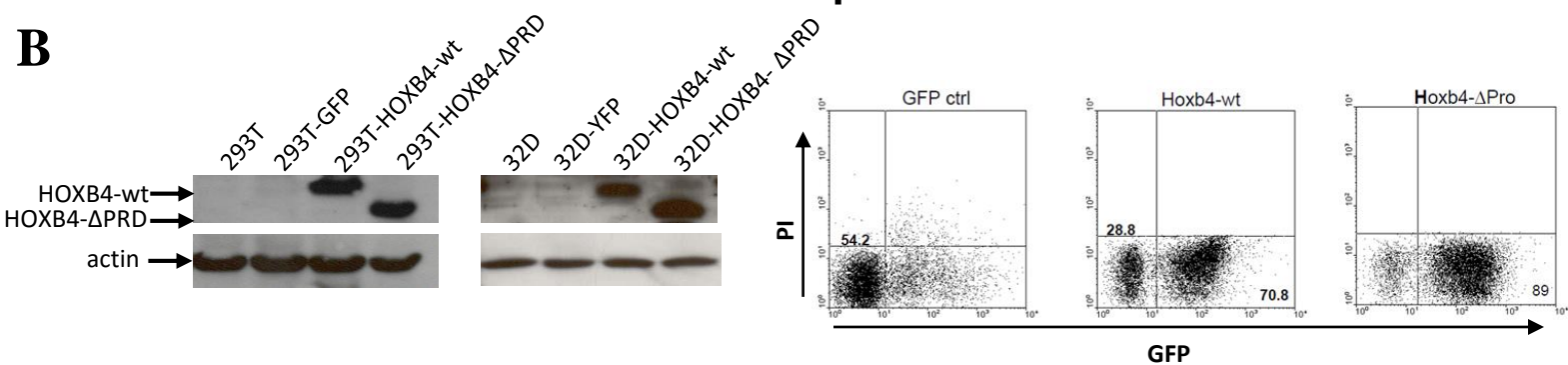
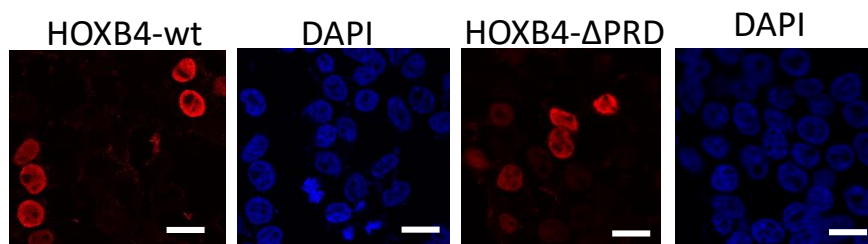
Fig. S2**A****B****C**

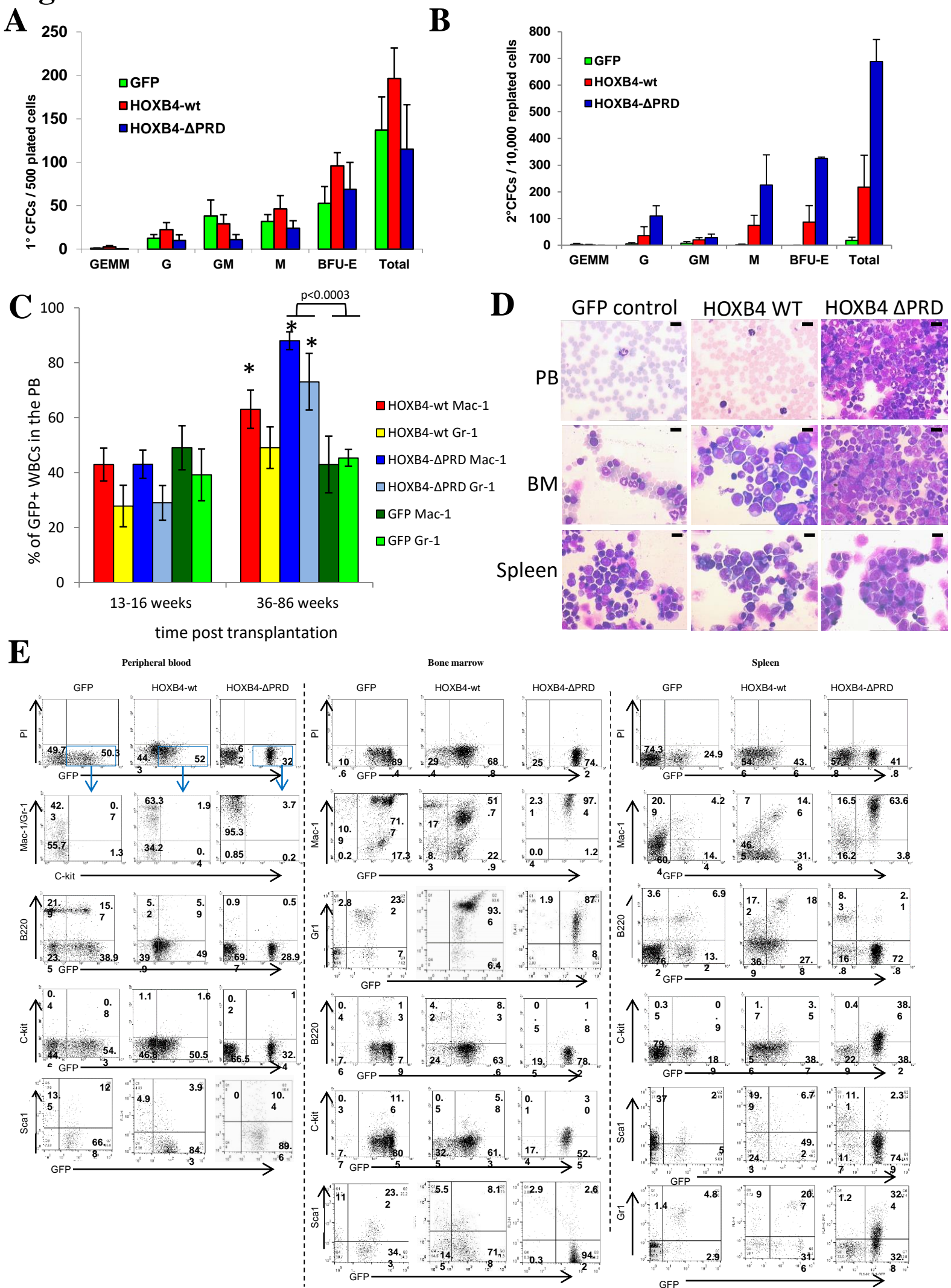
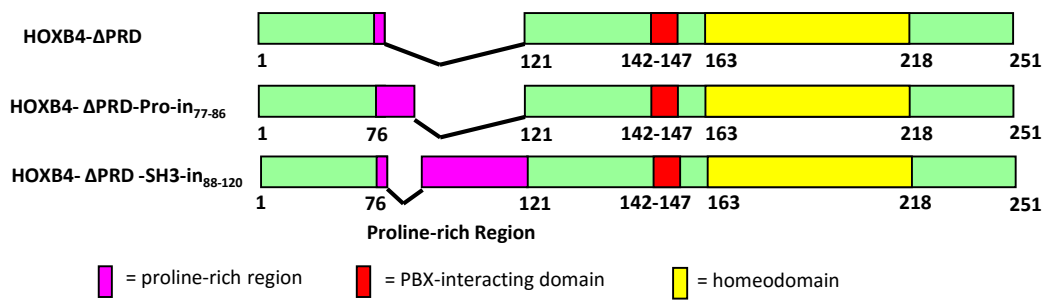
Fig. S3

Fig. S4

A



B

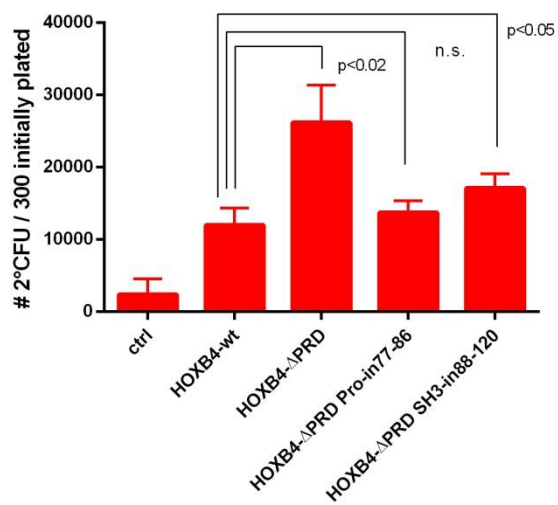


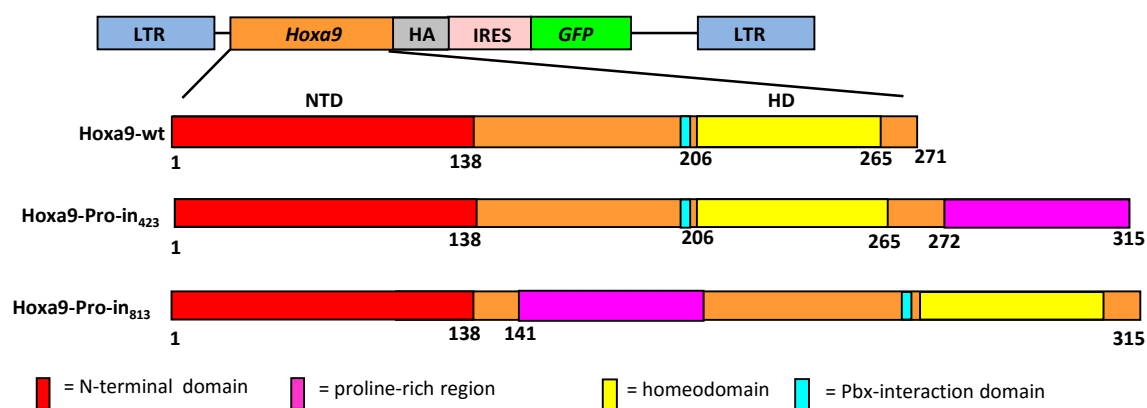
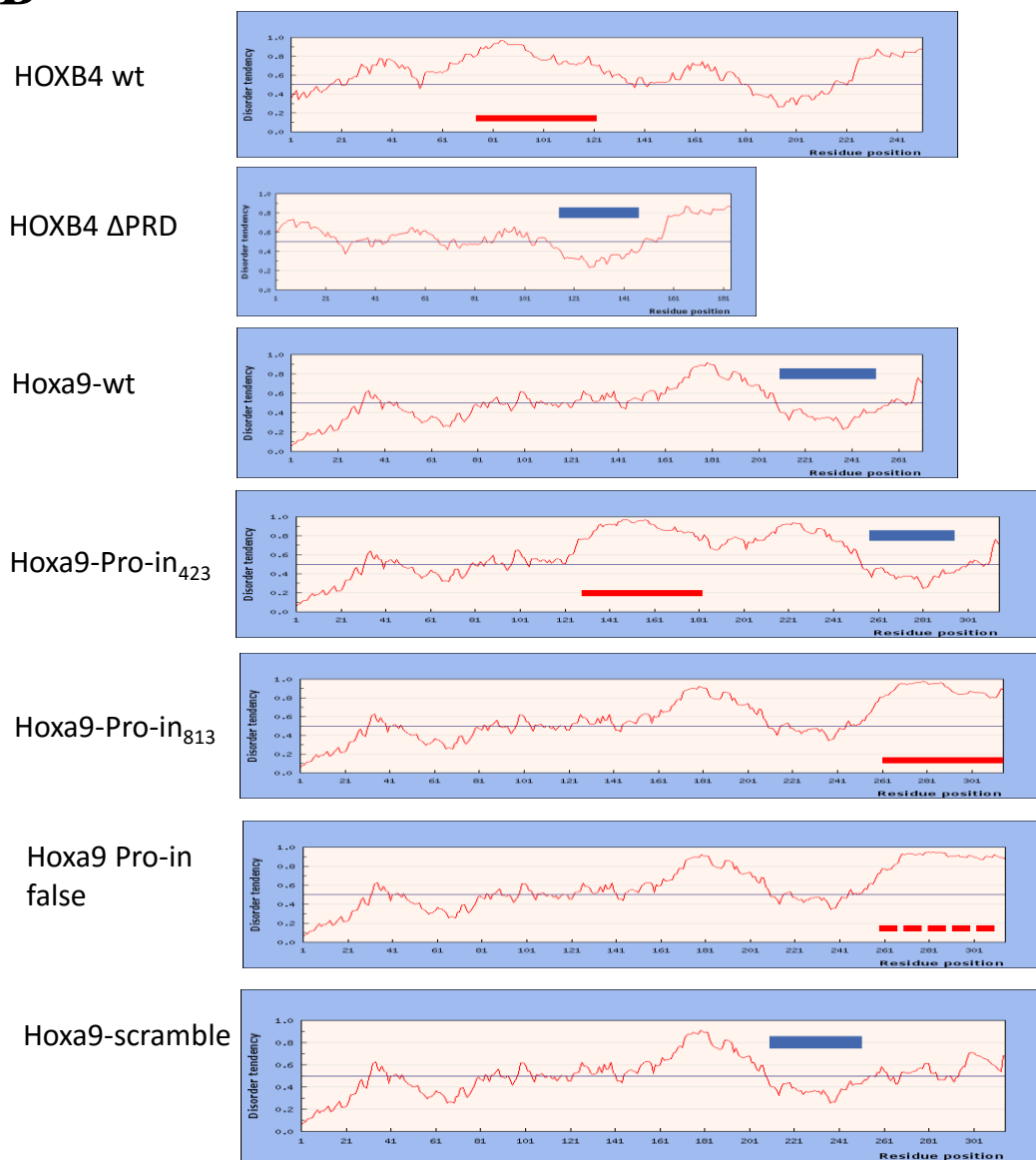
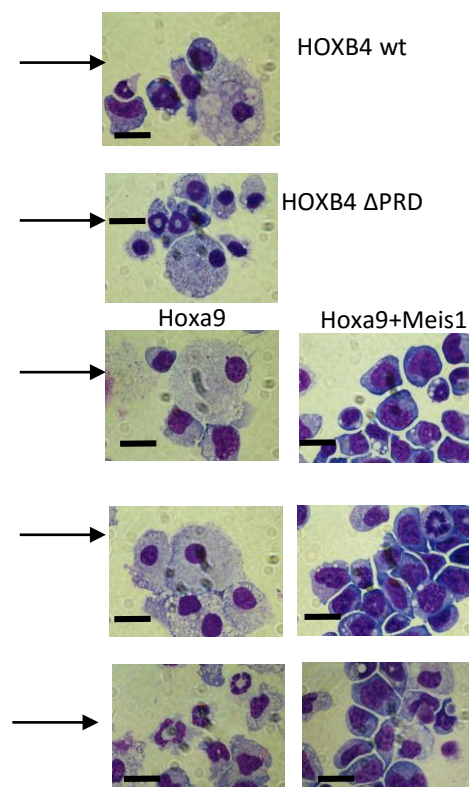
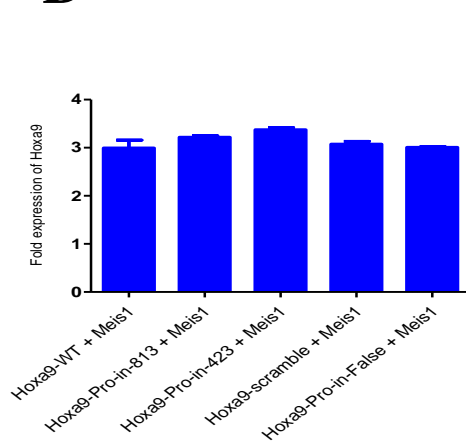
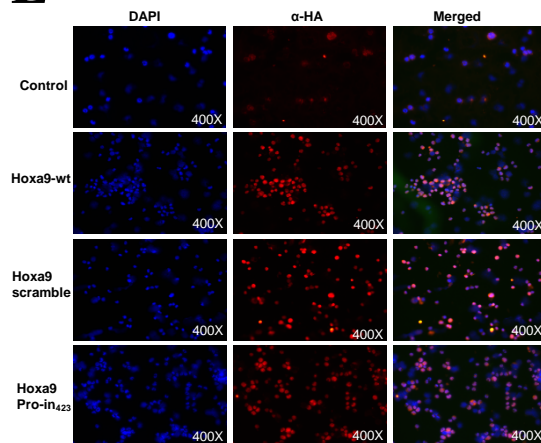
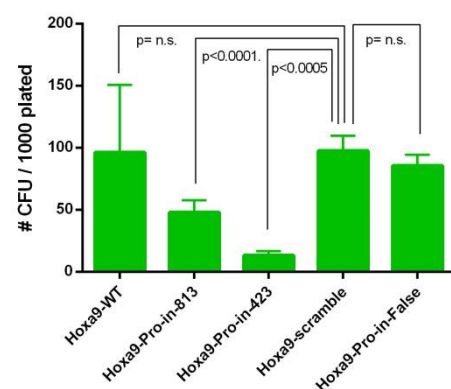
Fig. S5**A****B****C****D****E****F**

Fig. S6

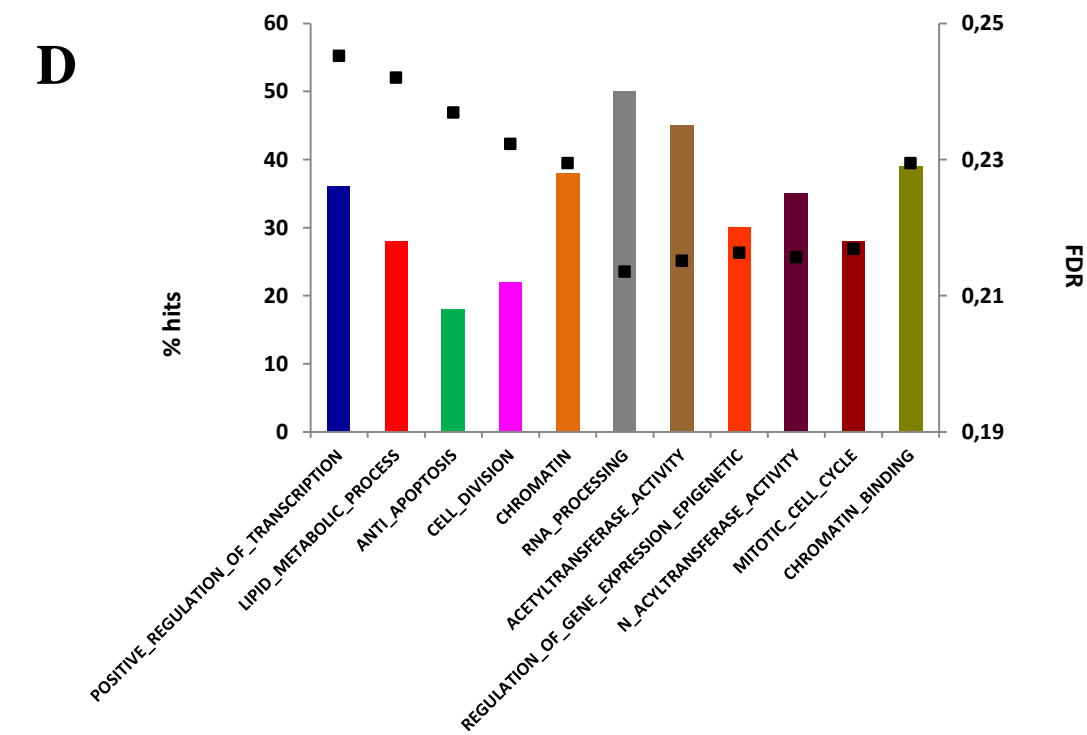
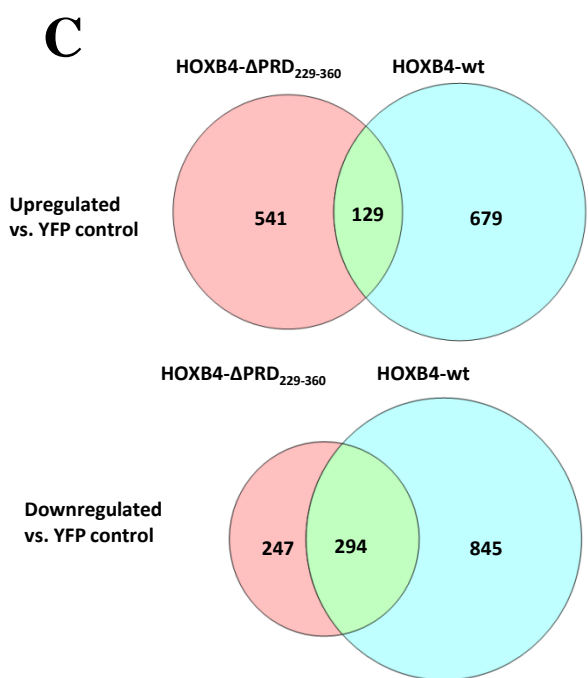
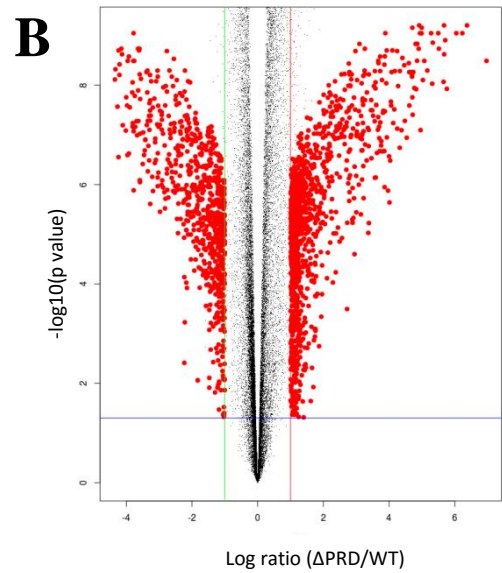
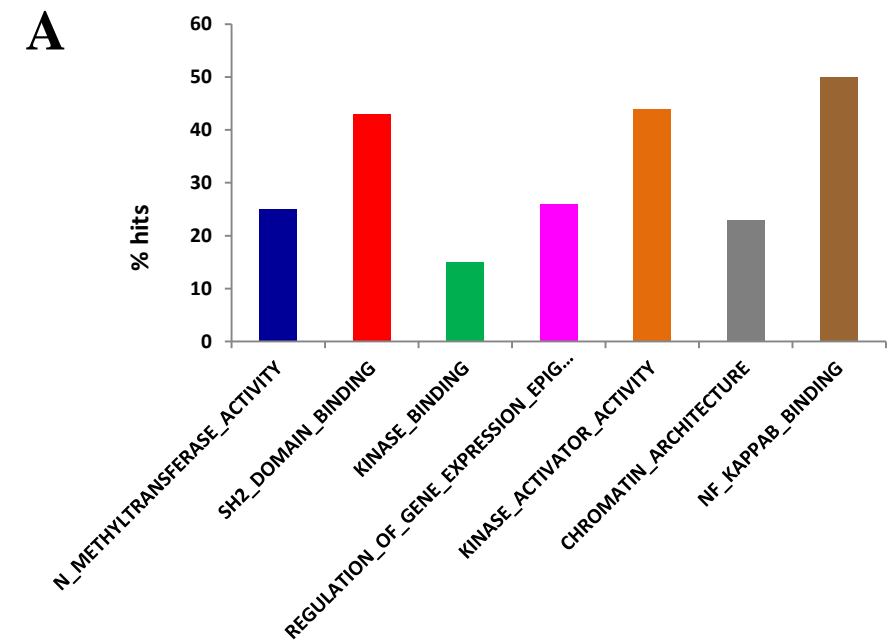
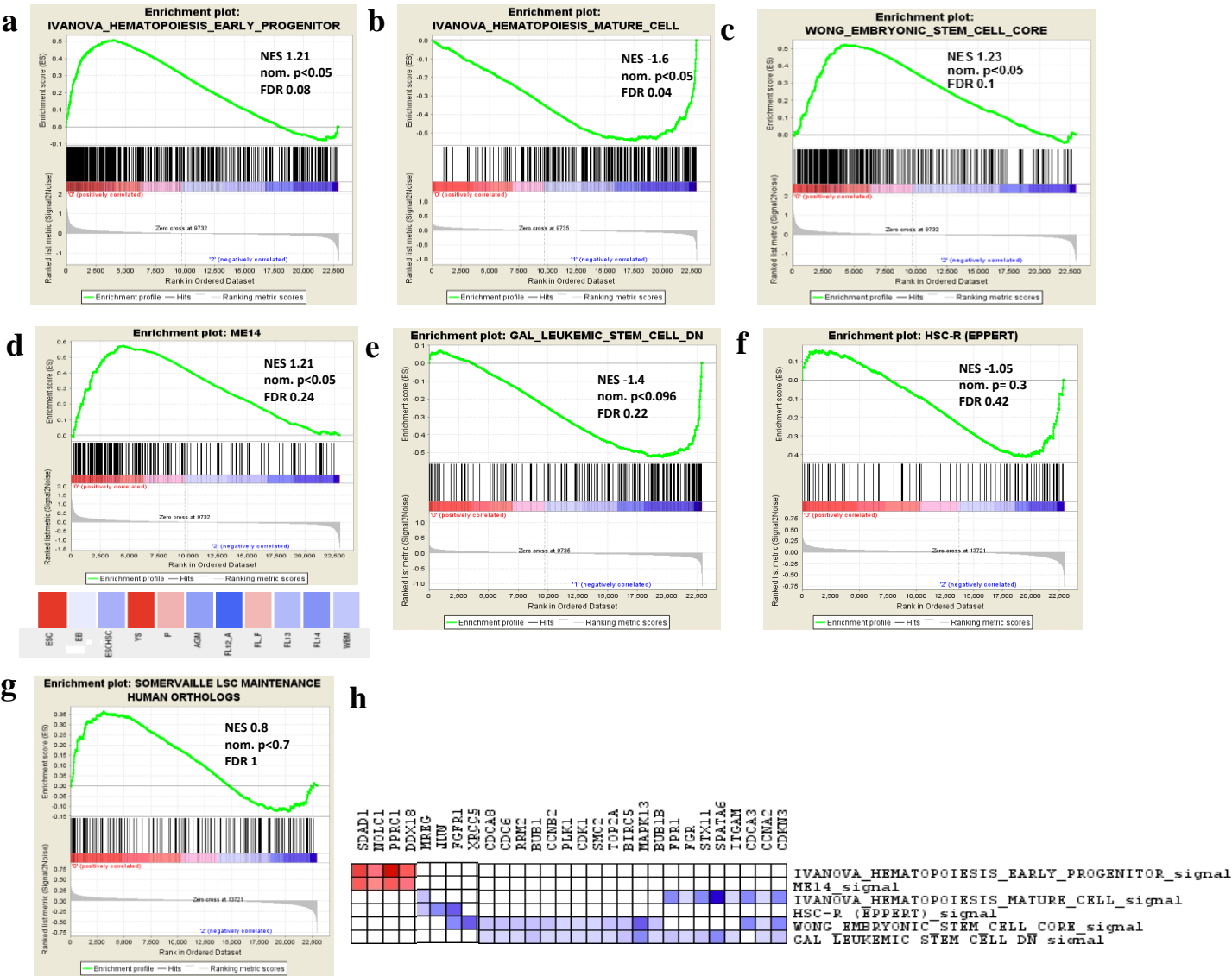


Fig. S6

E



F

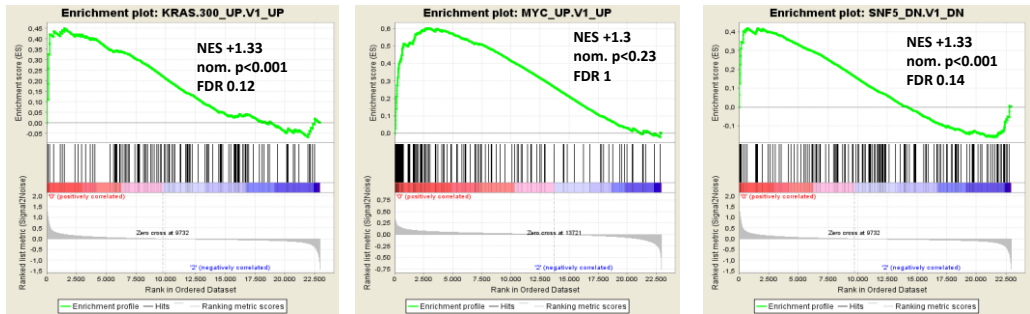
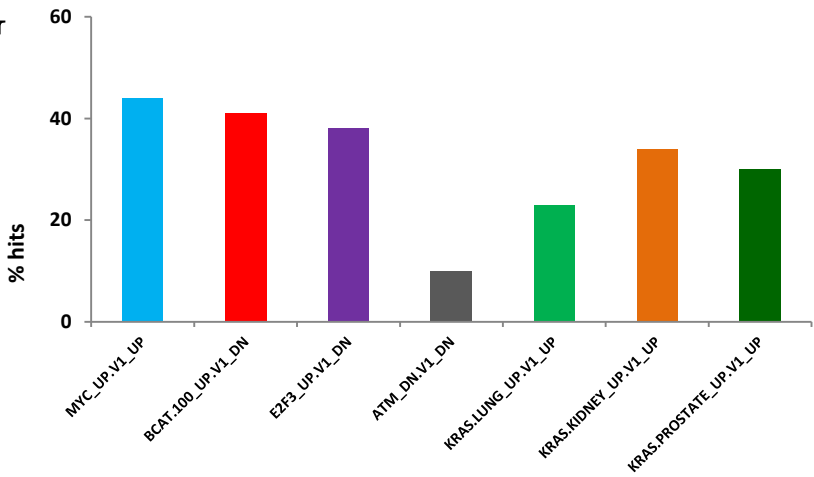
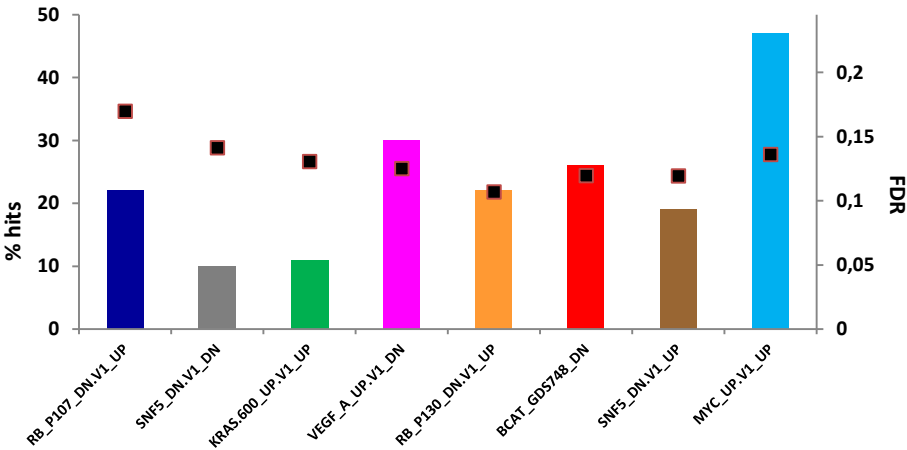


Fig. S6

G



H



I

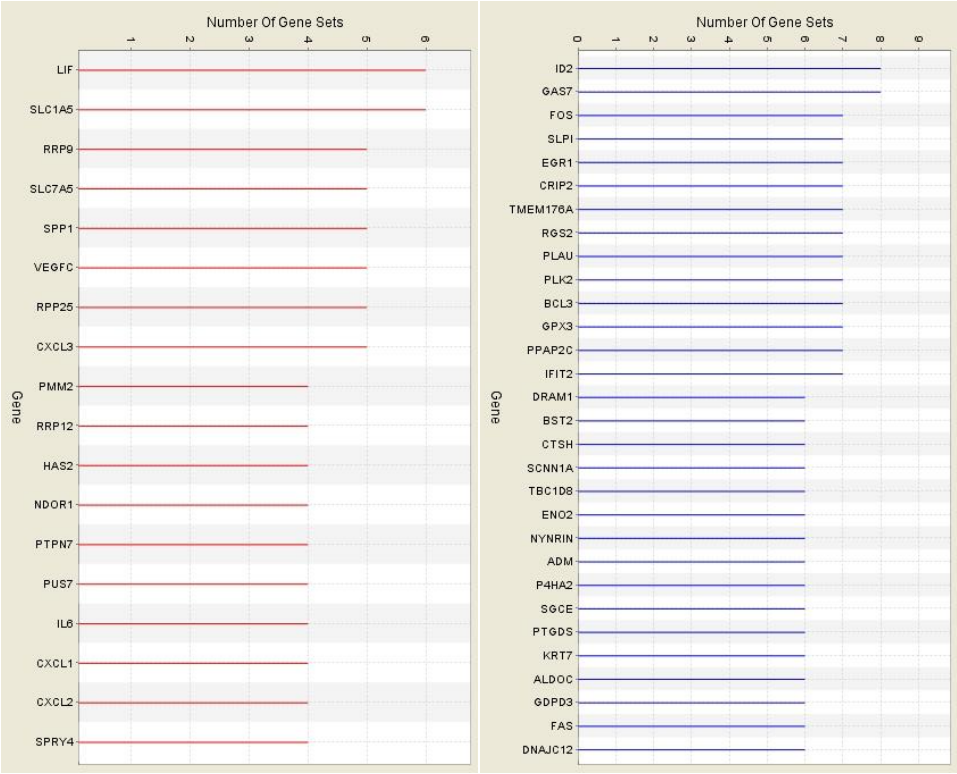
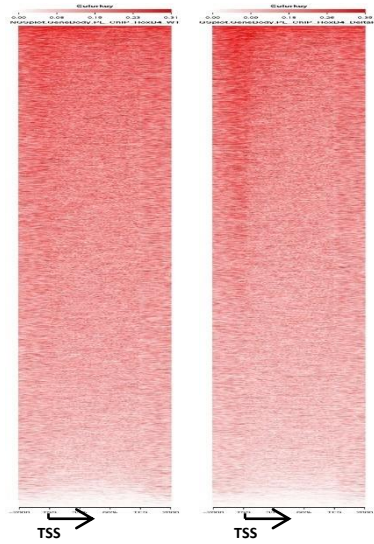
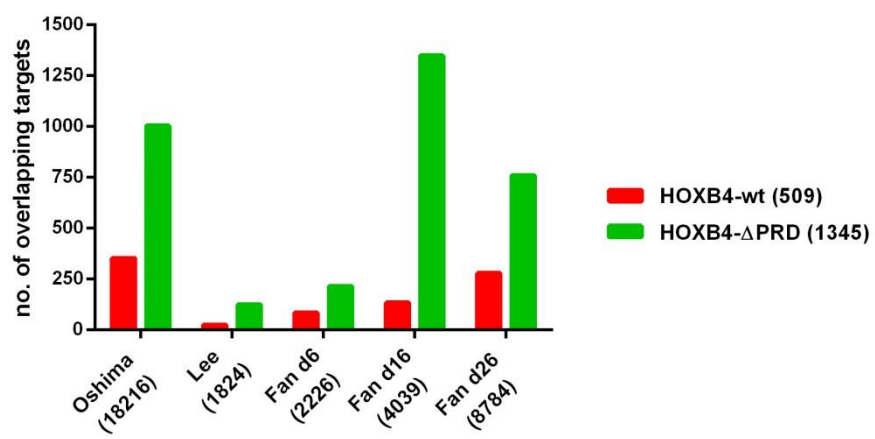


Fig. S7

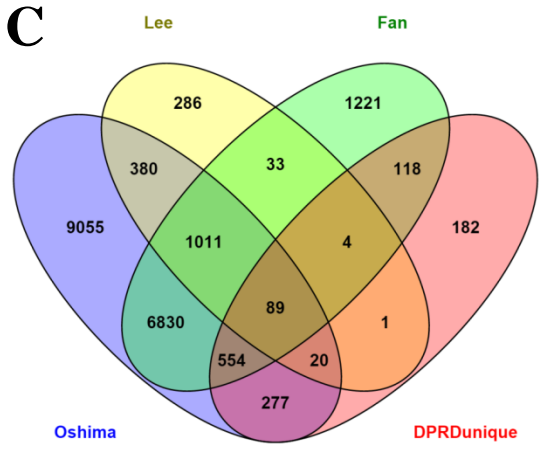
A



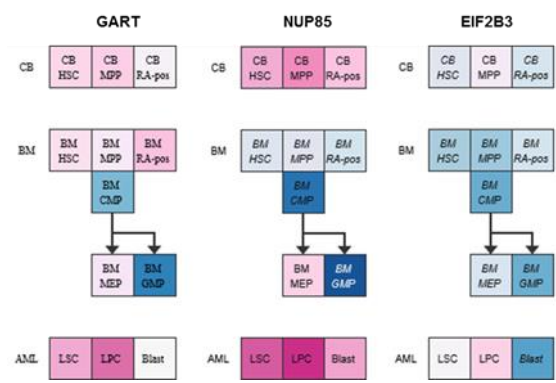
B



C



D



E

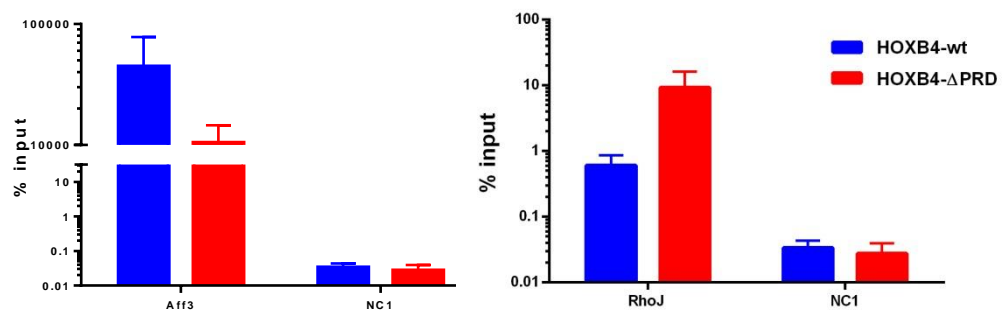


Fig. S8









	Motif	Protein	p value
HOXB4-WT		Myb	1e-12
		Rara	1e-12
		YY1°	1e-9
		USF1*	1e-4
HOXB4-ΔPRD		MafA	1e-12
		RUNX2	1e-12
		CRE*	1e-2
		STAT5*	1e-2

Fig. S9

