

Supplementary Table S1. Genomic Loci of provirus Integration in leukemia clone

Genomic loci representation of provirus integration “YourSeq” in upper and lower panels in the leukemic clone identified in two primary mice and one secondary mouse (Figure 4). The scale is adjusted in each panel to include the entire coding region of the gene where the provirus integration was identified, *Meis1* (upper panel) and *Zfp606* (lower panel).

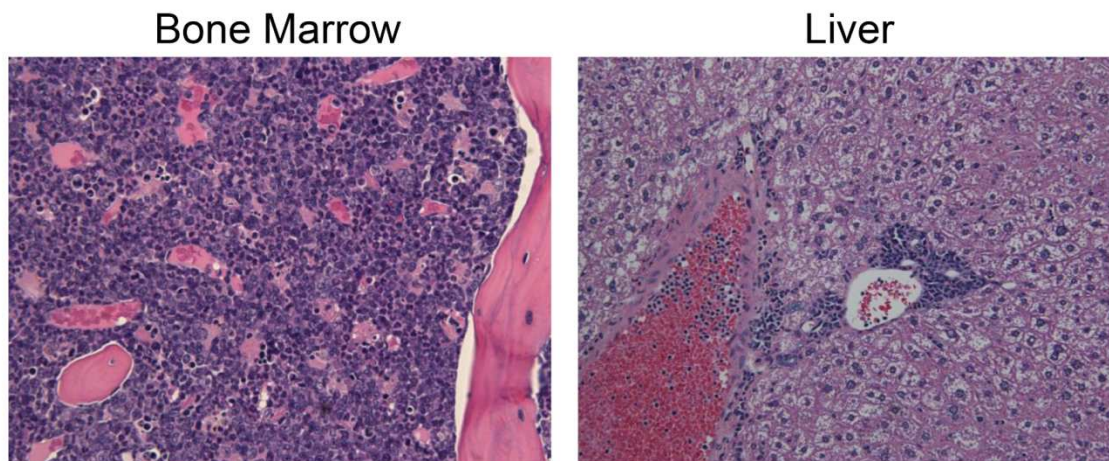
Mouse Number	Condition	Clonality	Major Hits	Sequence
11	Leukemia	Clonal	Meis 1	AATACCCTGCTCTGGATTTGACTGGCACCT TCACA
			ZFP606	TAAAATTTTGTITTTATTTTGTITTTATTTGTT TTGTTTTGTTTTATTTTCAGTGACTTTTATATT TTTTGAAGGAGCTATAAA
14	Leukemia	Clonal	Meis 1	GTATACCCTGCTCTGGATTTGACTGGCACC TTCACACAGTGCCATACATTTACACAGACA CACATGCATAAATA
			ZFP606	TAAAATTTTGTITTTATTTTGTITTTATTTGTT TTGTTTTGTTTTATTTTCAGTGACTTTATA
15	Leukemia	Clonal	Meis 1	GATACCCTGCTCTGGATTTGACTGGCACCT TCACACAGTGCCATACATTTACACAGACAC ACATGCATAAAATAAAAAAATAAATAT
			ZFP606	AATTTTGTITTTATTTTGTITTTATTTGTTTGT TTGTTTTATTTTCAGTGACTTTATATTTTTTG AAGGA
126	Myeloproliferative disorder	Clonal	Diap2	TTCCACCCAGAAAGAGACTATGTAATAAGC CATAGAAATAAGAAGTGGTTAATTGTGGGG CAATGGGCTATGCACAGACAGCCTGGTCT CCAGTCGAACTTAGGTCTTGAACCCAGTG GGACCCAATGGGTGATAATTTCCACCTACA TGGGACAGAAGGCATTTCGATCATGTCTCCT GAACCCCTGGCTTCTGTGCAAGTTACTGTC CCCACAGCCCCACAGGAGAGTTCGTGTGG CTAACAGTCACATAGACAATGTCCAAGCT TCTGGCATTCTGGCTAGACTCCTCCCACA GTTACCTAGCAAC
			ZFP946	ATAGTCCTTATAAATTAACCTCAGAGATACA ACTGTTGAATCATCAAAGCTAAACAGACAT AGAAGTGGAAACATCAGAGAA

Supplementary Table S2. Gene expression microarray was performed and 229 genes and 48 genes were differentially expressed between *HOXC6* and GMPs or *HOXC6* and CMPs, respectively (at a p-value of <0.05 after adjustment for multiple testing). Genes are listed from left upper corner to the right lower corner in the height of their regulation.

HOXC6 vs GMP				HOXC6 vs CMP	
HMG2	HSPA14	DUSP2	BLOC1S2	HAT1	SGSH
CBFA2T3H	ABCF2	TFRC	P2RY1	GLRX1	LOC630729
PGD	A230056E14RIK	FYN	PSMB2	LINGO2	S100A4
MED21	GJA1	1110014O20RIK	AUTS2	TSC22D1	A130086G11RIK
GOT2	4631426J05RIK	MBOAT2	TAF9	SH3GLB1	ITGAL
VPS13C	ZBTB38	USP48	2010004M13RIK	TOR1A	SOX21
HTRA2	MTHFD1	ATP5A1	GIN51	MSI2	FOSL2
APOE	DMKN	CALM2	D12ERTD647E	PIK3R1	GM2A
GNG2	SCL000896.1_70	DUSP6	GIN52	CRIM1	CSF2RB2
NGFRAP1	EG434401	EIF6	BC055368	GLRX	UTP18
MUC13	RBM47	TSPO	E430003D02RIK	TXNIP	ATP6AP1
H1F0	ANGPT1	MMP8	LOC635086	PTPRC	PTGS1
SKP2	SEPP1	TSPAN6	CKAP4	5730536A07RIK	5430426F23RIK
LOC544884	4732423C17RIK	TTK	2310043N10RIK	TSPAN13	RPS27L
THY1	DOCK11	LOC672274	MET	ANXA5	MANSC1
LHFPL2	KCNG2	IL15	MGL1	MSI2H	PFAS
NME2	STX3	ZFP787	ITGA6	ANXA3	LYZS
6620401M08RIK	EXOSC3	LOC100039636	SERPIN6B	LOC100041500	LOC383981
NME1	MCPT8	ST3GAL1	LOC100046513	IRF8	TMEM147
1110046J11RIK	LOC229810	GP5	PPIC	TMEM165	APP
SIPA1L1	NAB2	OLFR553	PTTG1IP	PECAM1	RRP1B
CD33	TNFSF12-TNFSF13	LRRCS9	DNAJB1	SLC18A2	SLC24A3
AA388235	RNPEP	HIRIP3	EG620119	APAF1	MC5R
IRAK1BP1	PYCR1	ASPH	NGK7	ELA2	FGL2
GCAP27	SORD	RBMS1	KCTD12	2700094K13RIK	B4GALT6
HSD17B10	PRTN3	LOC674195	SLC45A3	1110012O05RIK	RUNX2
LOC671523	KIT	D130026O08RIK	IER3	SLC14A1	PAPSS2
RCC1	HPRT1	DTX3L	LOC224163	MGST1	SRGN
ADAP1	LOC270665	MFGE8	DSTN	LOC667395	FNDC3B
HSPA1L	CTLA2A	PKIB	ACOT7		GCAP26
CLECSF9	BEST1	DDIT4	LOC383077		AATK
LOC666036	DAPP1	HHEX	ALAS1		MSH6
CENTD3	LOC638935	LAT2	GPC3		LYZ2
OSBPL5	HK3	FBNP1L	UBE1L		UPF1
KLF6	TRP53INP1	DTWD2	CDK10		LY78
GM1964	HBP1	H2-OB	RTEL1		MS4A3
IKZF2	CD27	A230072B04RIK	IMPDH2		LOC384418
PAIP2	MLLT3	PHKA2	PRKCA		ATP6V1D
YTHDC2	SOAT1	1110020P15RIK	EGR1		TMEM38B
GAS7	SELP	LOC100045864	HSP90AA1		IFIT2
LOC547343	PAOX	ASS1	CYCS		CCND2
LOC382264	SELL	PAFAH2	4921515J06RIK		GRN
FBNP1	FADS1	BC003498	PADI2		MARCKS
RPE	TMEM9	CCL9	KIF1B		8030467N07RIK
RNF6	CSNK2B	SOX4	LOC100041835		PGCP

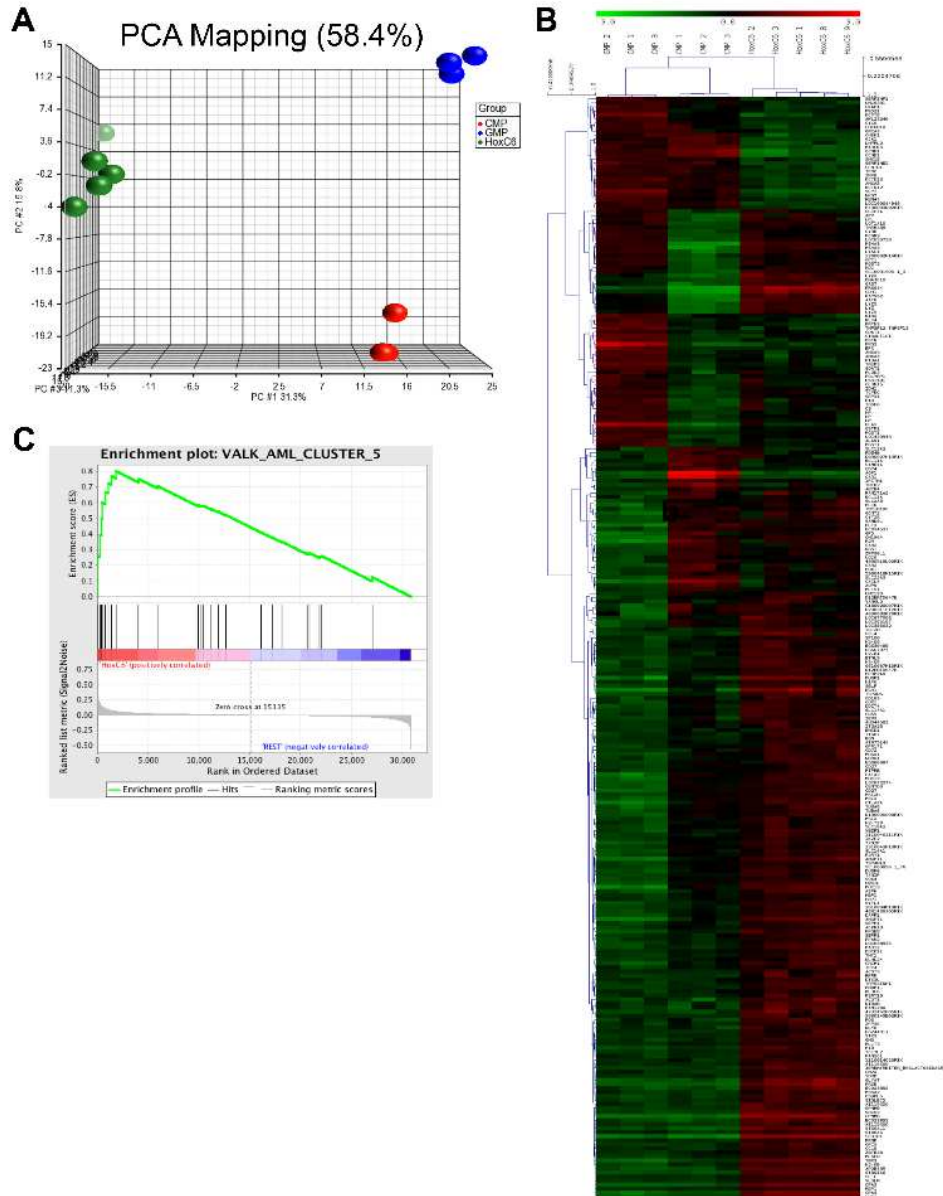
RGS1	STAT3	ITGB2	FRMD4A	AI447904
CNIH4	ITPR2	PSMA7	UCL5	PRNP
CCT8	PHF17	6330407118RIK	SLPI	
LOC666559	TUBA8	NOLA2	TREM3	
PARP6	PSMD14	NOLA3	CTSG	

ACCEPTED MANUSCRIPT



Supplementary Figure S1. Myeloproliferative disorder induced by *HOXC6* overexpression.

Histology of BM and liver from a *HOXC6* expressing primary mouse that was diagnosed with a myeloproliferative disease due to absence of excessive blasts in analysed tissues. Furthermore some infiltrations of hematopoietic cells into the liver were found (Hematoxylin-Eosin staining; x20).



Supplementary Figure S2. *HOXC6* expressing progenitors show clearly distinct gene expression and can be identified as AML gene signature. (A) Principle component analysis of the cDNA microarray data shows independent clustering of GFP control CMP samples (red dots), GFP control GMP samples (blue dots), and GMP/CMP-like *HOXC6* samples (green dots). (B) Raw normalized heatmap of relative expression levels of greater than \log_2 -fold regulated genes for the GMP, CMP and *HOXC6* overexpressing cells. (C) GSEA plots further identified

AML gene expression signatures were enriched in the HOXC6 expressing cells (NES 2.02, FDR=0.0).

ACCEPTED MANUSCRIPT