Supplementary Information

PU.1 is Essential for MLL Leukemia Partially via Crosstalk with the MEIS/HOX Pathway

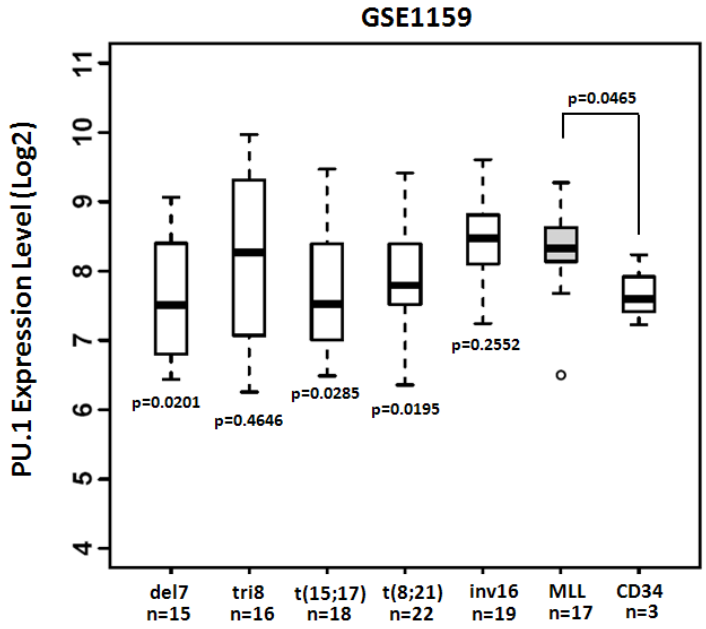
Running title: PU.1 is essential for MLL Leukemia

Jing Zhou[1]*, Jun Wu[1]*, Bo Li[1,3,4]*, Dan Liu[1,3], Jian Yu[2], Xiaomei Yan[4], Si Zheng[1,3], Jiguang Wang[1,5], Lixia Zhang[1,3], Lin Zhang[1], Fuhong He[1], Qiang Li[1], Aili Chen[1,3,4], Yue Zhang[4], Xinghui Zhao[4], Yinghui Guan[6], Xiaohong Zhao[1], Jin Yan[1], Jun Ni[7], Marcelo A. Nobrega[8], Bob Löwenberg[9], RDelwel[9], Peter J.M. Valk[9], Ashish Kumar[12], Lu Xie[2], Daniel G. Tenen[10,11], Gang Huang[4#] and Qian-fei Wang[1#]

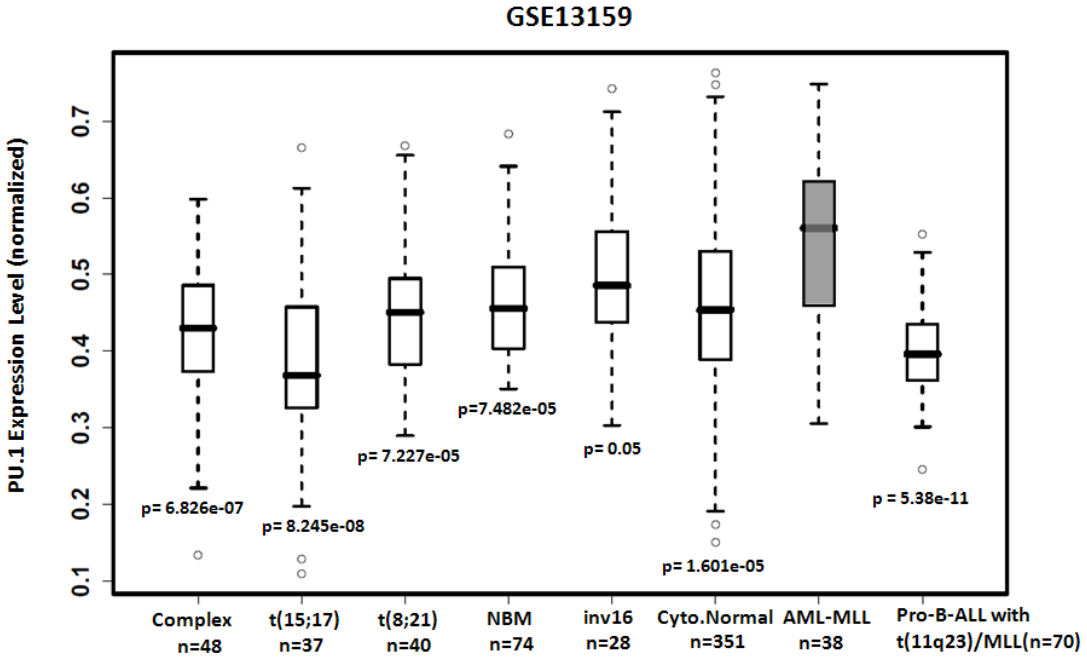
This file contains Supplementary Figures related to main Figures 1, 2, 3, 4, 5, and 6. It also provides all of the primer sequences used in the study, as well as supplementary Methods and Materials information.

Figure S1 (related to Fig.1)

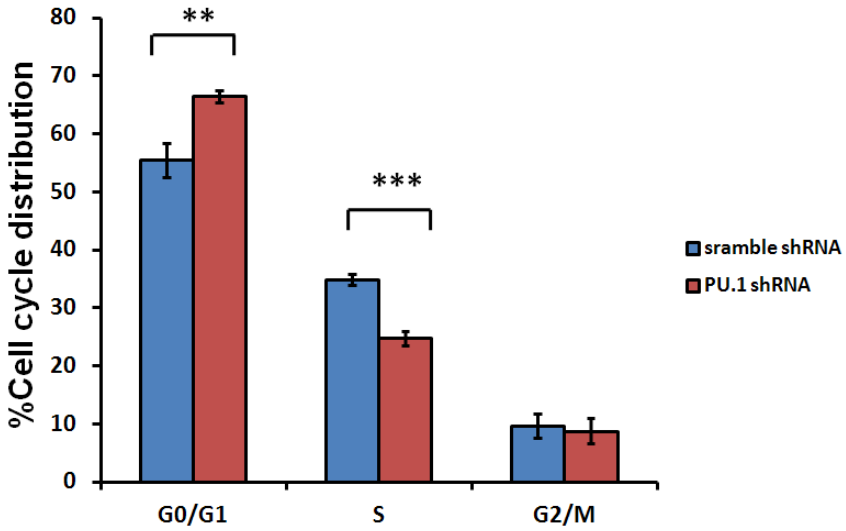
A



B



C



D

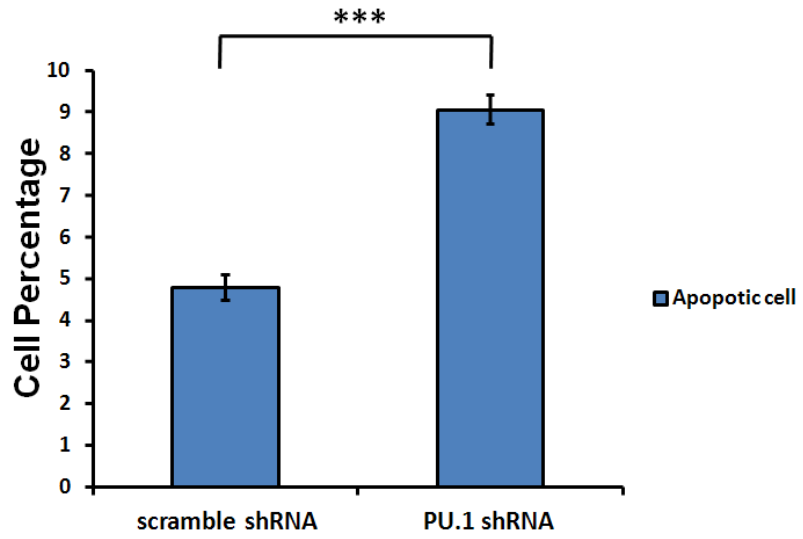


Figure S1.

(A) Expression level of PU.1 in different cytogenetic subgroups of AML in GSE1159. Data are shown as a box plot with whiskers. Wilcoxon rank sum test was performed for testing expression difference between corresponding groups, respectively; p values are given.

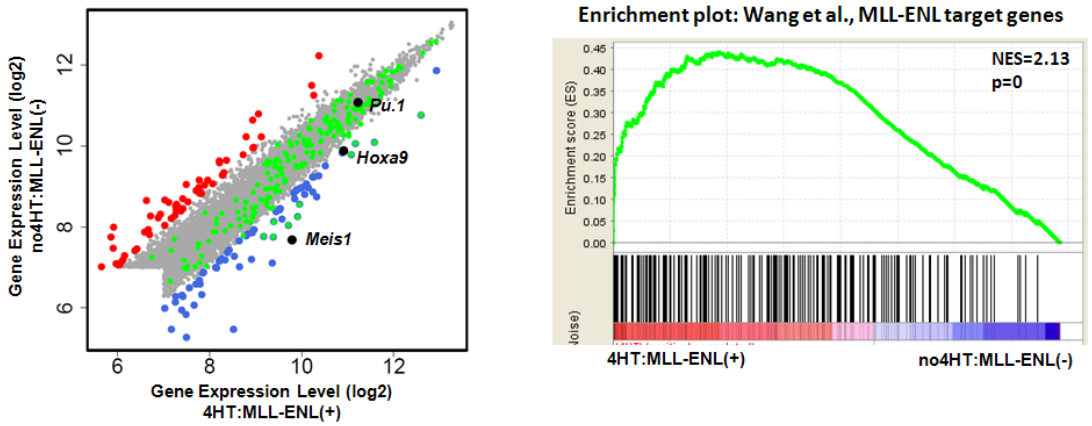
(B) Box plot showing the comparison of PU.1 expression levels between AML with t(11q23)/MLL, Pro-B-ALL with t(11q23)/MLL, cytogenetically normal AML and other cytogenetic subgroups of AML in GSE13159. Wilcoxon rank sum test was performed for testing expression differences between corresponding groups, respectively; p values are given

(C) Percentage of csh3 cell in G1, S and G2 phase with and without PU.1 specific shRNA treatment. Each bar represents the mean ± SD of three independent experiments. (t-test, G0/G1: p=0.001841, S: p=0.000169) (**: p<0.01, ***: p<0.001).

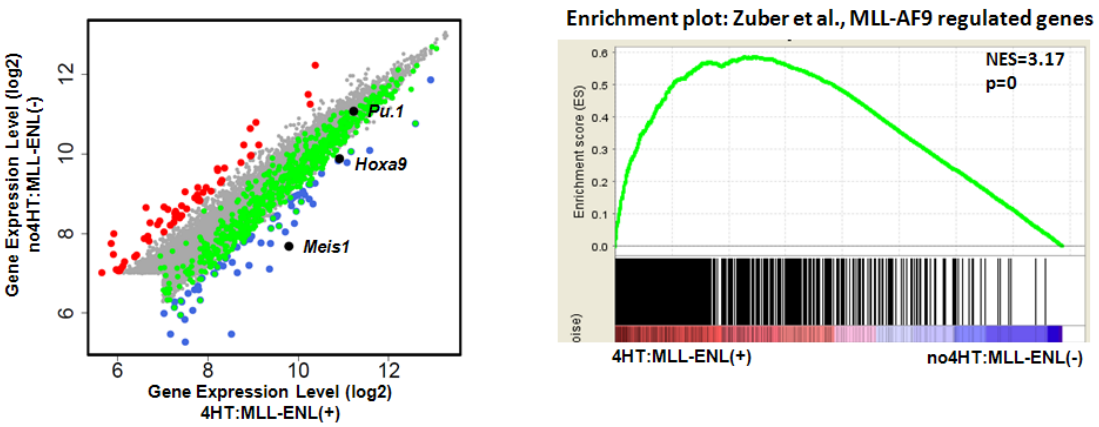
(D) Percentage of apoptotic cells in csh3 with and without PU.1 specific shRNA treatment. Each bar represents the mean ± SD of three independent experiments. (t-test, $p=4.46*10^{-5}$) (***: p<0.001).

Figure S2 (related to Fig.2)

A



B



C

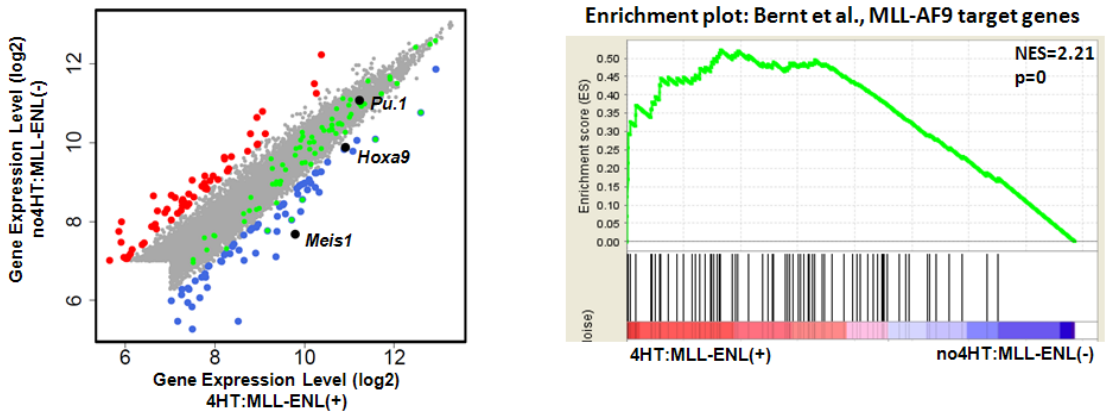


Figure S2.

Gene expression levels of three different gene sets of known MLL fusion targets (A: Wang et al., 233 MLL-ENL targets; B: Zuber et al., 936 MLL-AF9 regulated genes;

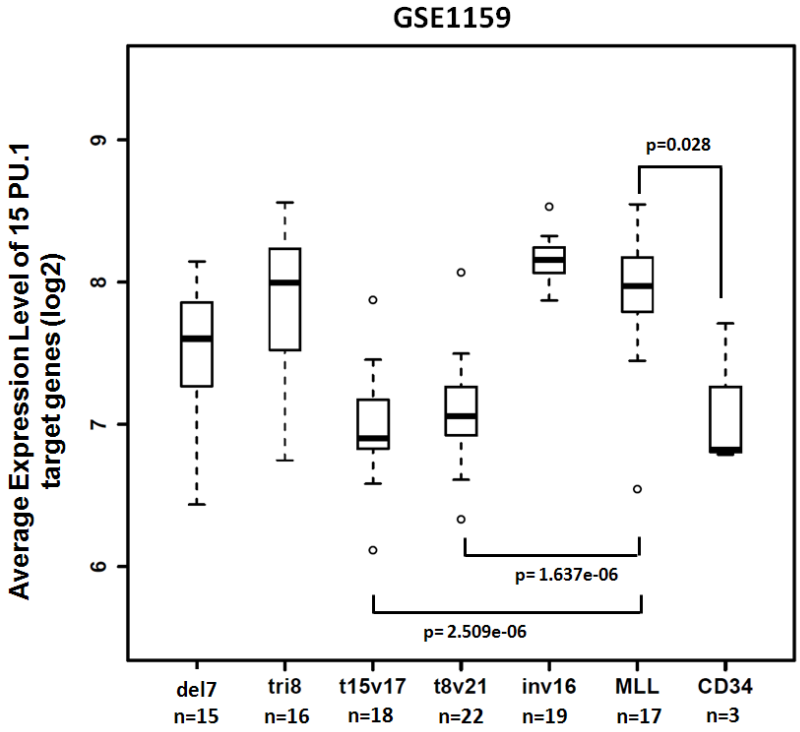
C: Bernt et al., 139 MLL-AF9 targets) in MLL-ENL induced (MLL-ENL(+)) and inactivated (MLL-ENL(-)) conditions.

Left panels in A-C: Scatter plot showing gene expression levels in MLL-ENL induced (MLL-ENL(+)) and inactivated (MLL-ENL(-)) conditions. All 7858 expressed genes detected by mouse exon array are displayed[1]. Significantly down-regulated genes upon inactivation of MLL-ENL (FDR<0.05) are shown as blue dots. Significantly up-regulated genes upon inactivation of MLL-ENL (FDR<0.05) are shown as red dots. Genes whose expression levels are not significantly changed between the two conditions are shown as grey dots. Genes in each gene set of MLL fusion targets are shown as green dots. PU.1, Meis1 and Hoxa9 are indicated by black dots.

Right panels in A-C: Enrichment plots of GSEA are shown for each gene set of MLL fusion targets. The normalized enrichment scores (NES) and p-values are given.

Figure S3 (related to Fig.3)

A



B

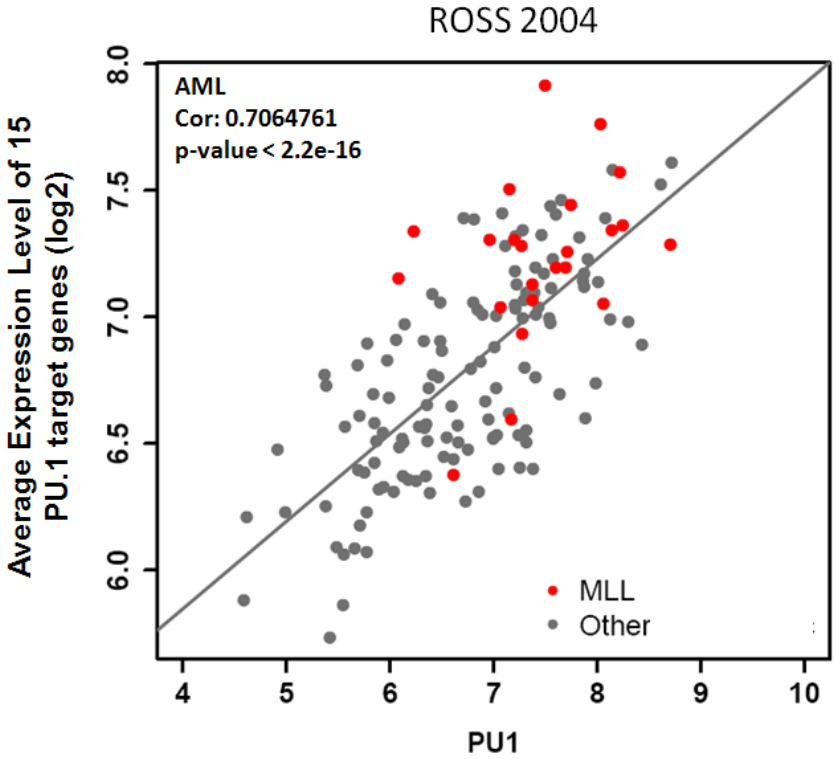


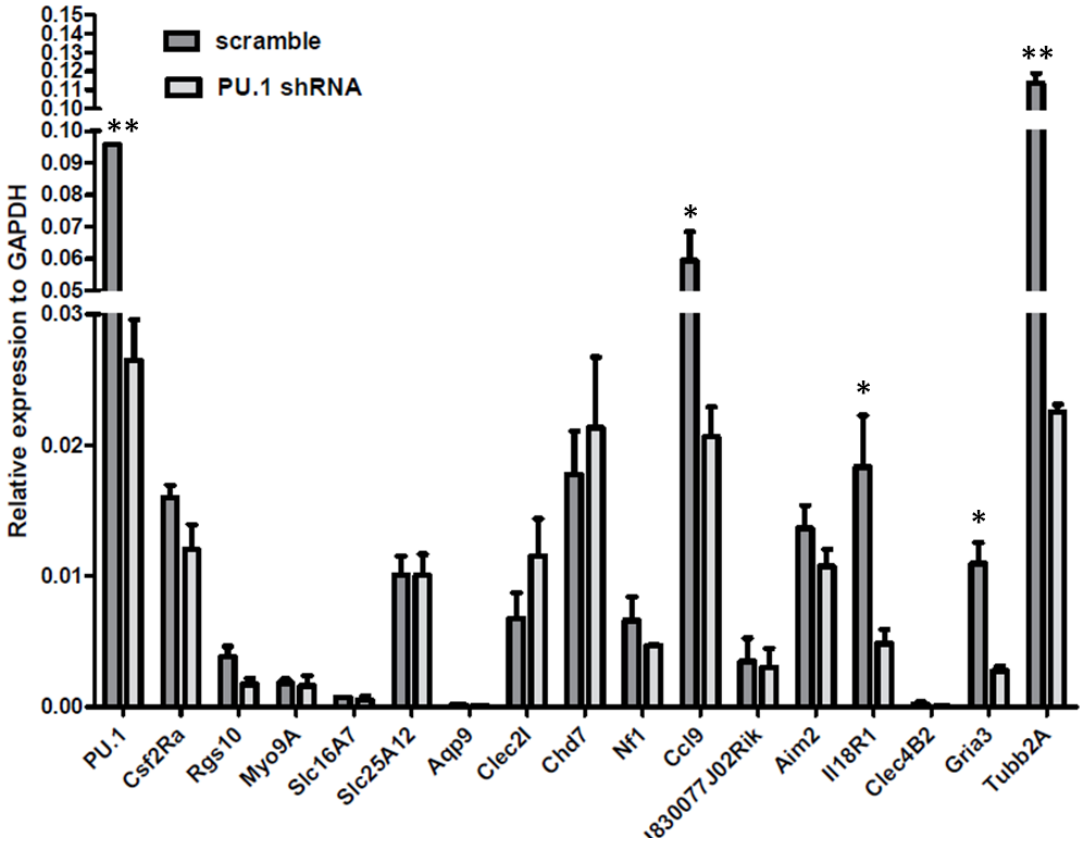
Figure S3.

(A) The average expression level of 15 PU.1 core targets in MLL compared with distinct cytogenetic sub-groups (t(8;21), t(15;17), tri8, inv16 and del7) and CD34 cells in GSE1159. Data are shown as a box plot with whiskers. Wilcoxon rank sum test was performed for testing the expression difference between corresponding groups, respectively; p values are given.

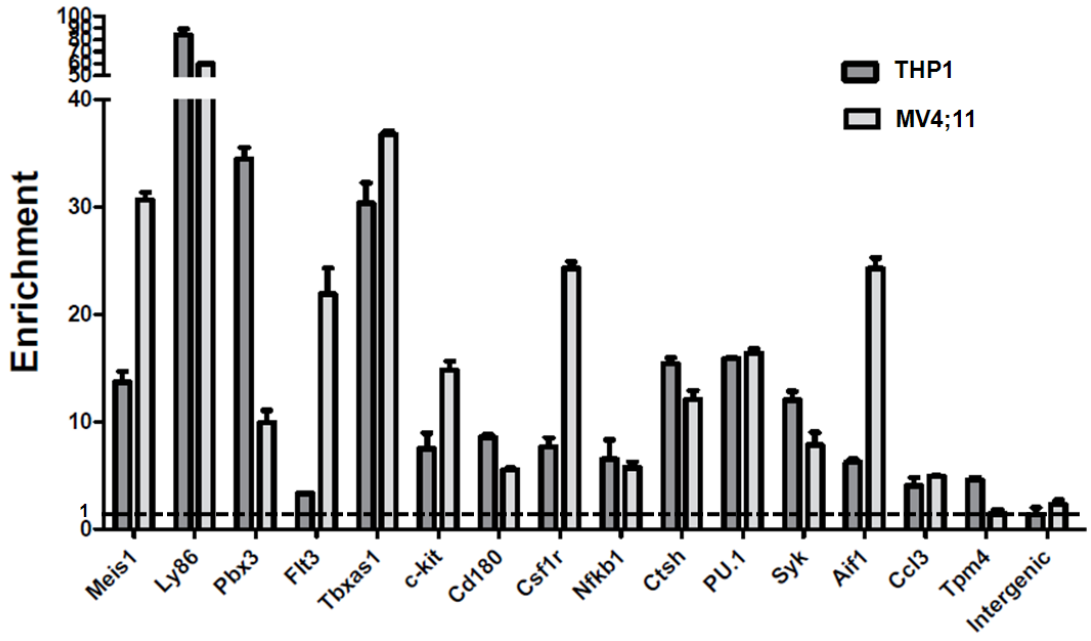
(B) Relationship between expression level of PU.1 and its 15 core targets in a large cohort of 150 AML patients (Ross et al., 2004). Red dots indicate MLL patients; the remaining AMLs are indicated by grey dots. The Pearson correlation coefficient between PU.1 and the average expression level of 15 core targets is 0.7064791 ($p<2.2e-16$).

Figure S4 (related to Fig. 4 and Fig. 5)

A



B



C

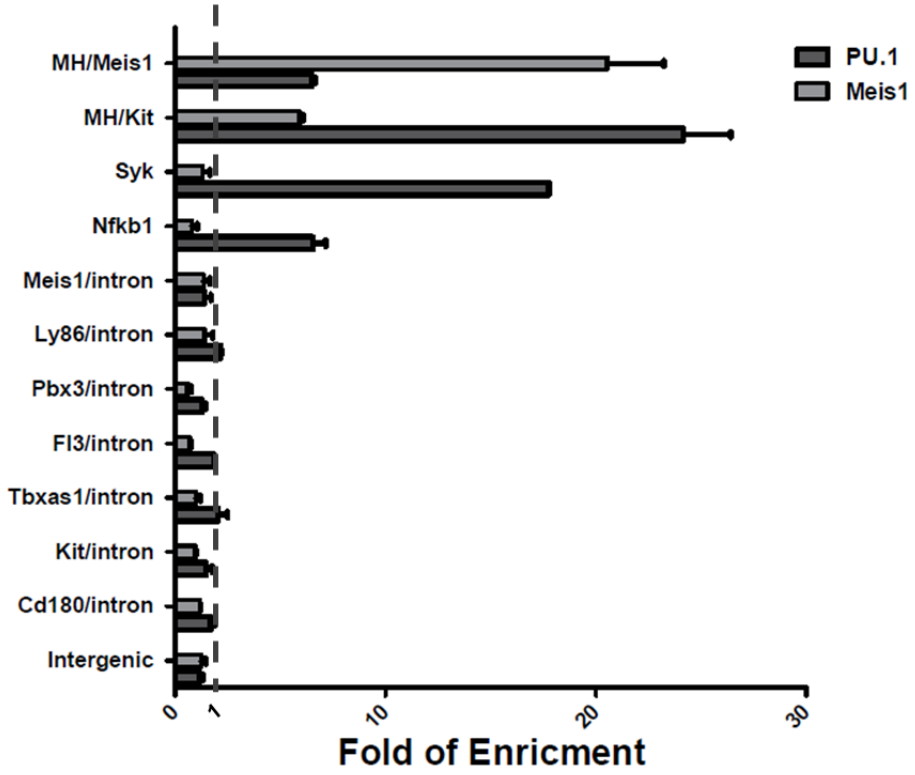


Figure S4.

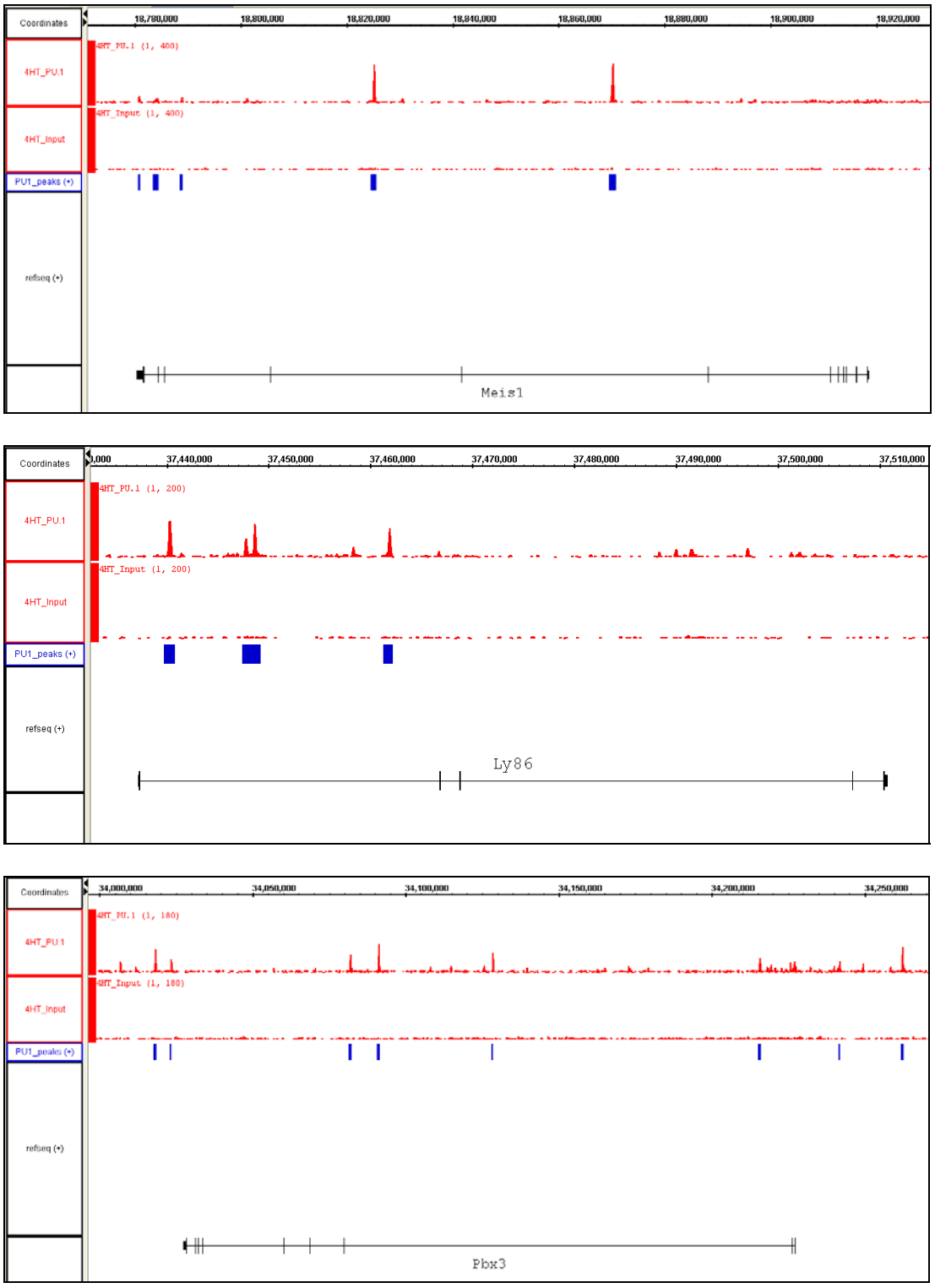
(A)Effects of PU.1 knockdown on additional target genes expression in MLL-ENL cells. The MLL-ENL cells were transduced with PU.1 shRNA lentivirus and scramble control lentivirus. The expression of PU.1 and its additional target genes were examined through real-time PCR 48h after transduction. The relative expression was normalized to GAPDH. Bar charts are mean ± SD. Genes with significantly decreased levels of expression after PU.1 knockdown are indicated (t-test, *: $p<0.05$, **: $p<0.01$).

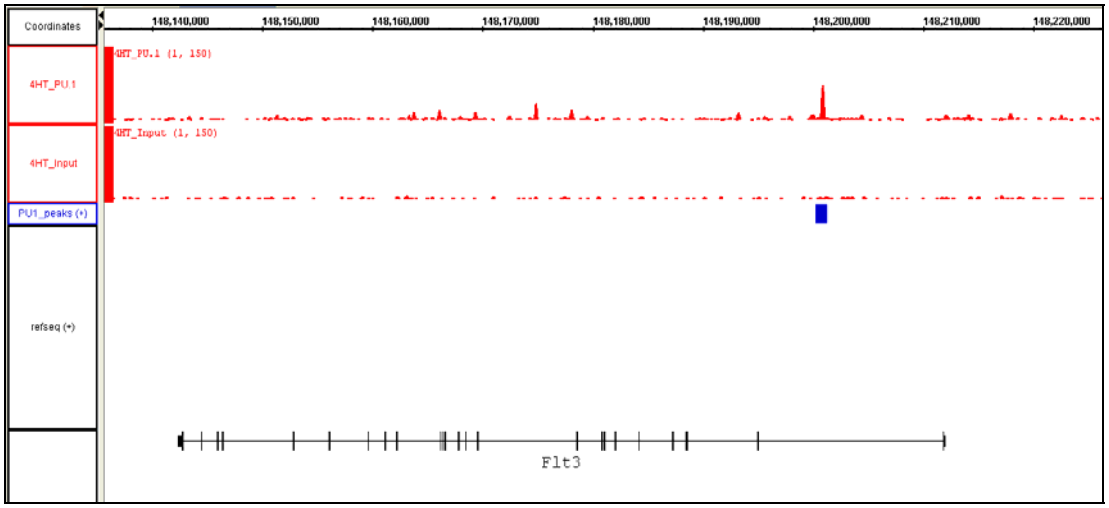
(B)PU.1 binding at gene regions of PU.1 15 core target genes in THP1 and MV4;11. The amounts of input DNA and ChIP DNA were

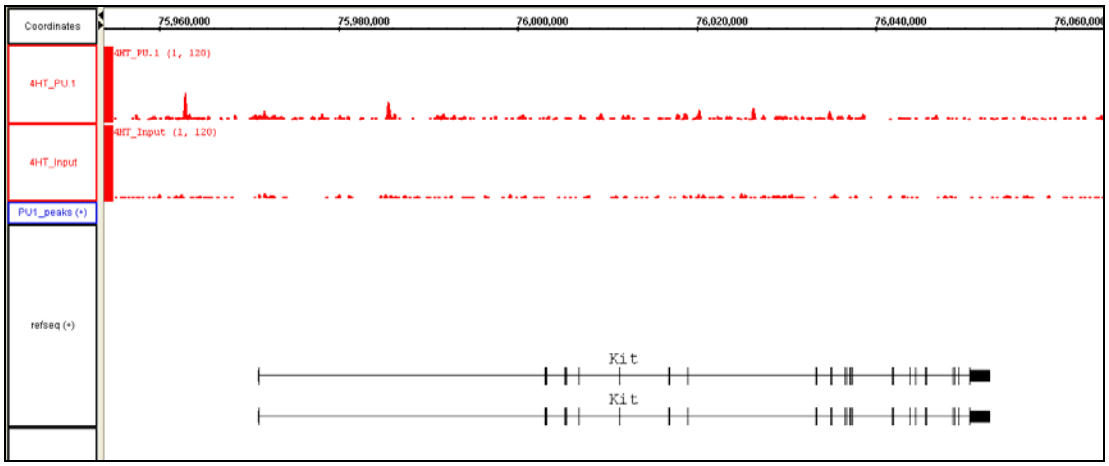
normalized, and the data are shown as the relative enrichment ratio of precipitated DNA to input DNA. Intergenic region was used as the negative control. Bar charts are mean ± SD. The dotted line indicates the level of no enrichment over input.

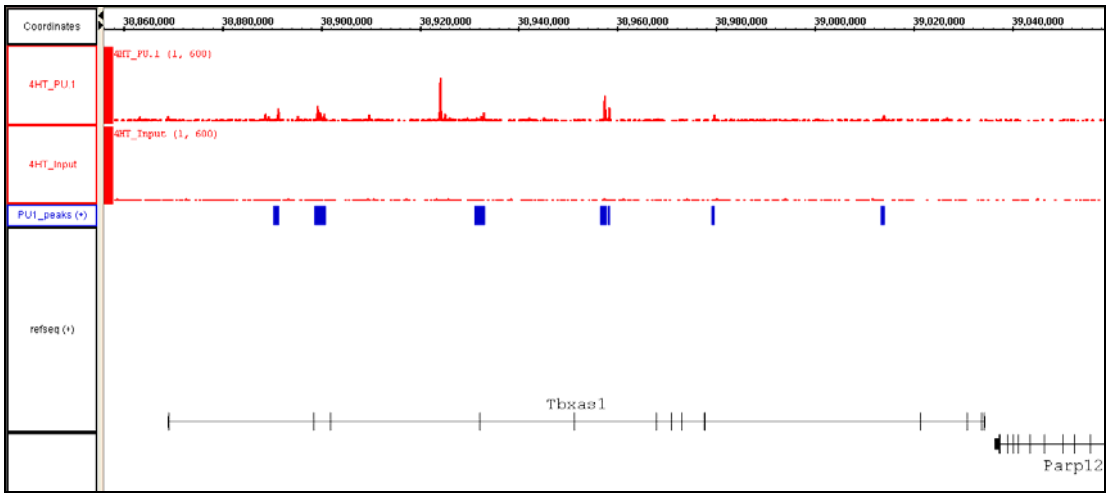
(C) PU.1 and MEIS1 binding at gene regions without predicted PU.1/MEIS1 binding sites of PU.1/MEIS1 co-target genes. PU.1 and MEIS1 binding was examined using ChIPqPCR. The amounts of input DNA and ChIP DNA were normalized, and the data are shown as the relative enrichment ratio of precipitated DNA to input DNA. Intergenic region was used as the negative control. Bar charts are mean ± SD. MH/Mesi1, MH/Kit, Syk and Nfkb1 are positive controls. The dotted line indicates the level of no enrichment over input.

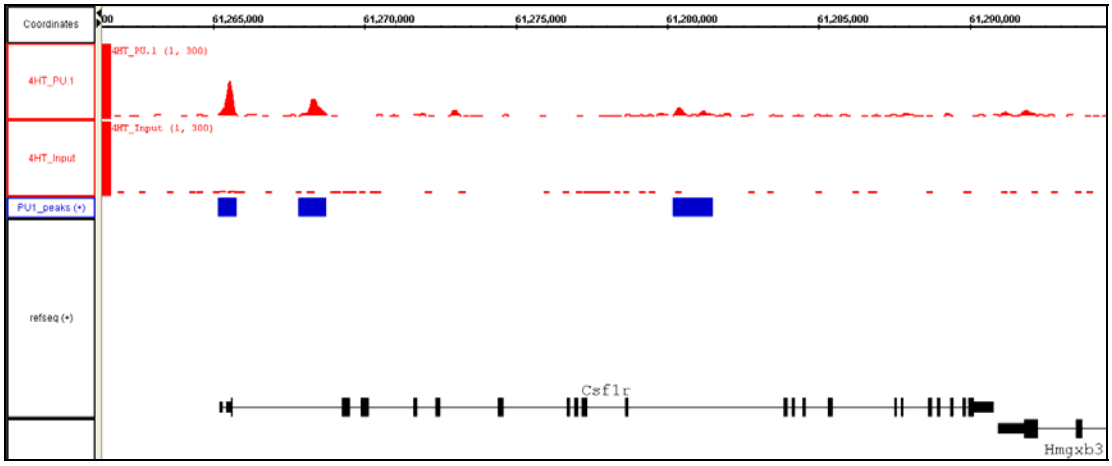
Figure S5 (related to Fig. 5)

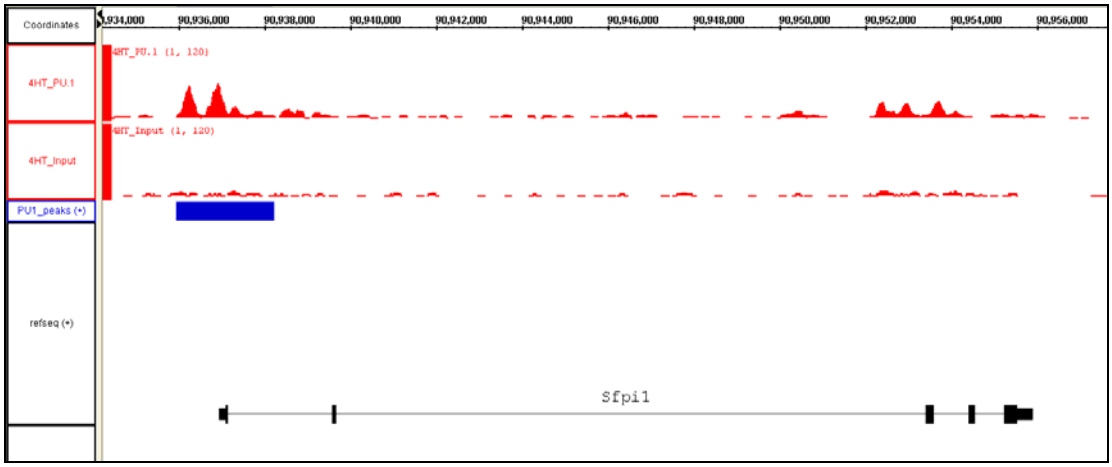


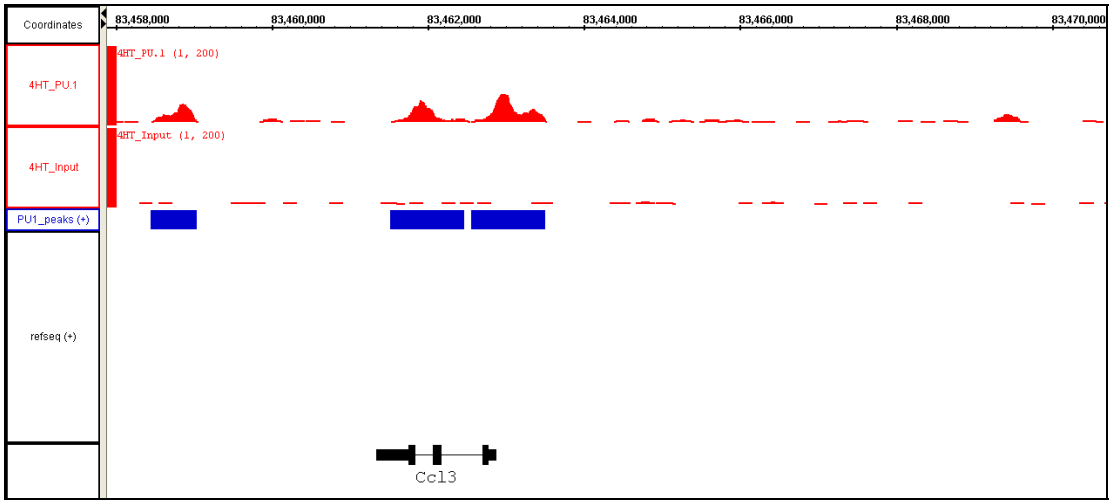

Coordinates
148,140,000
148,150,000
148,160,000
148,170,000
148,180,000
148,190,000
148,200,000
148,210,000
148,220,000
4HT_PU.1
4HT_PU.1 (1, 150)
4HT_Input
4HT_Input (1, 150)
PU1_peaks (+)
refseq (+)
Flt3

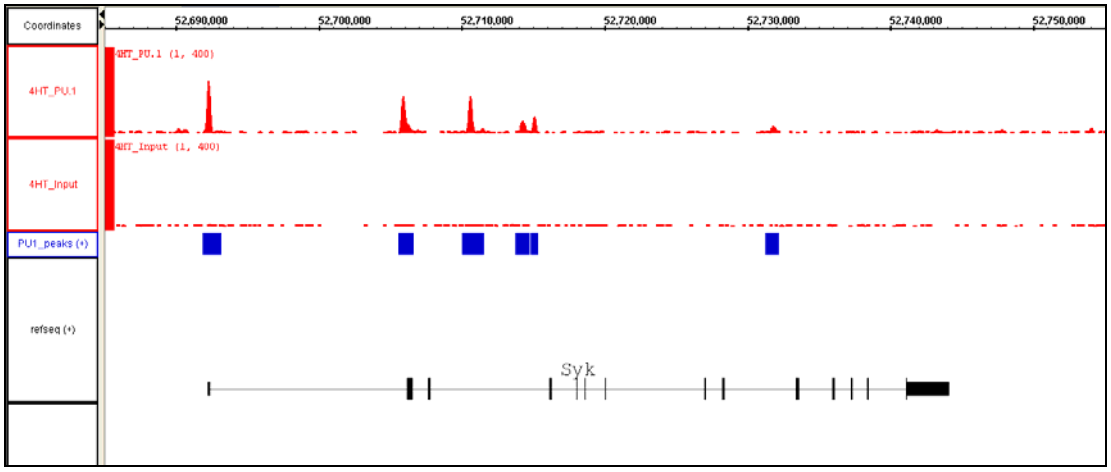

Coordinates
75,960,000
75,980,000
76,000,000
76,020,000
76,040,000
76,060,000
4HT_PU.1
4HT_PU.1 (1, 120)
4HT_Input
4HT_Input (1, 120)
PU1_peaks (+)
refseq (+)
Kit
Kit

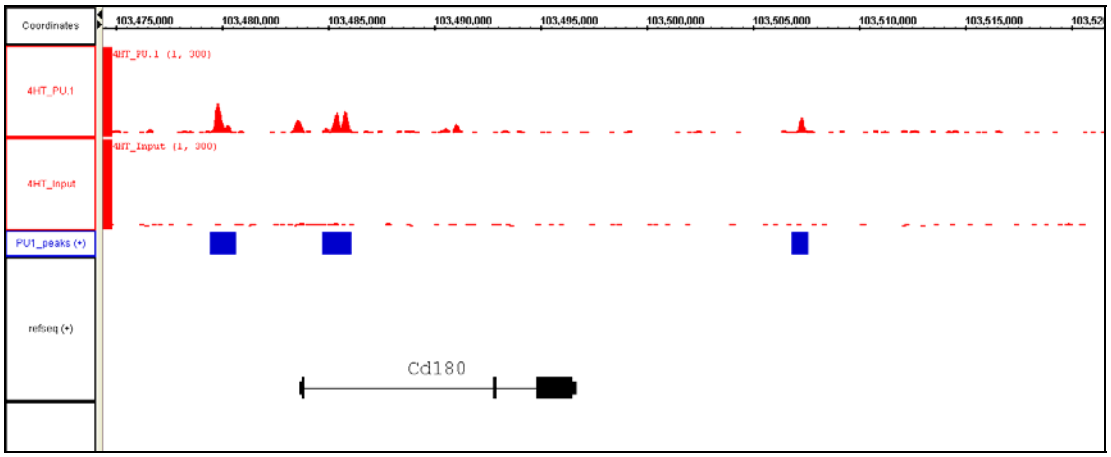

Coordinates
38,860,000
38,880,000
38,900,000
38,920,000
38,940,000
38,960,000
38,980,000
39,000,000
39,020,000
39,040,000
4HT_PU.1
4HT_PU.1 (1, 600)
4HT_Input
4HT_Input (1, 600)
PU1_peaks (+)
refseq (+)
Tbxas1
Parp12

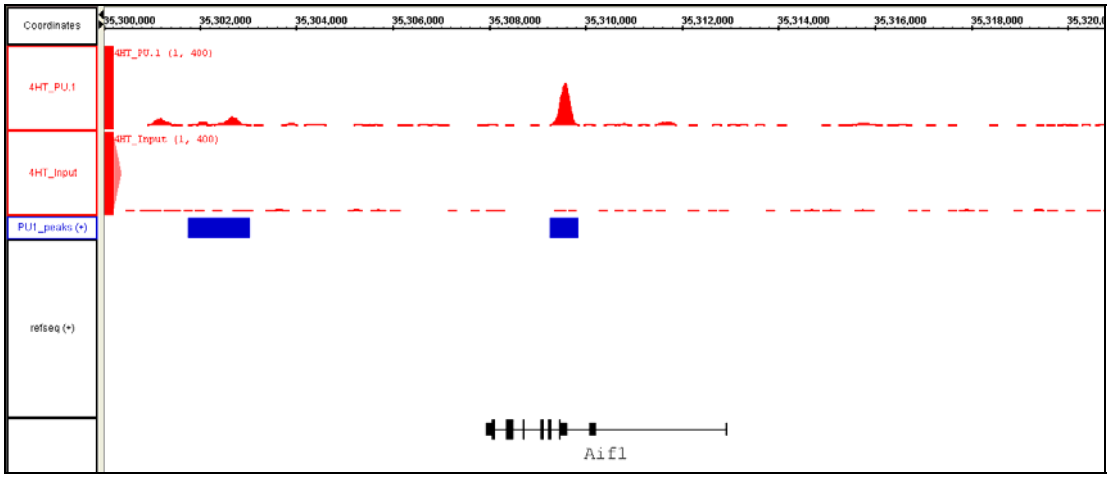

Coordinates
61,265,000
61,270,000
61,275,000
61,280,000
61,285,000
61,290,000
4HT_PU.1
4HT_PU.1 (1, 300)
4HT_Input
4HT_Input (1, 300)
PU1_peaks (+)
refseq (+)
Csf1r
Hmgxb3


Coordinates
90,934,000
90,936,000
90,938,000
90,940,000
90,942,000
90,944,000
90,946,000
90,948,000
90,950,000
90,952,000
90,954,000
90,956,000
4HT_PU.1
4HT_PU.1 (1, 120)
4HT_Input
4HT_Input (1, 120)
PU1_peaks (+)
refseq (+)
Sfpi1


Coordinates
83,458,000
83,460,000
83,462,000
83,464,000
83,466,000
83,468,000
83,470,000
4HT_PU.1
4HT_PU.1 (1, 200)
4HT_Input
4HT_Input (1, 200)
PU1_peaks (+)
refseq (+)
Ccl3


Coordinates
52,690,000
52,700,000
52,710,000
52,720,000
52,730,000
52,740,000
52,750,000
4HT_PU.1
4HT_PU.1 (1, 400)
4HT_Input
4HT_Input (1, 400)
PU1_peaks (+)
refseq (+)
Syk


Coordinates
103,475,000
103,480,000
103,485,000
103,490,000
103,495,000
103,500,000
103,505,000
103,510,000
103,515,000
4HT_PU.1
4HT_PU.1 (1, 300)
4HT_Input
4HT_Input (1, 300)
PU1_peaks (+)
refseq (+)
Cd180


Coordinates
35,300,000
35,302,000
35,304,000
35,306,000
35,308,000
35,310,000
35,312,000
35,314,000
35,316,000
35,318,000
4HT_PU.1
4HT_PU.1 (1, 400)
4HT_Input
4HT_Input (1, 400)
PU1_peaks (+)
refseq (+)
Aif1

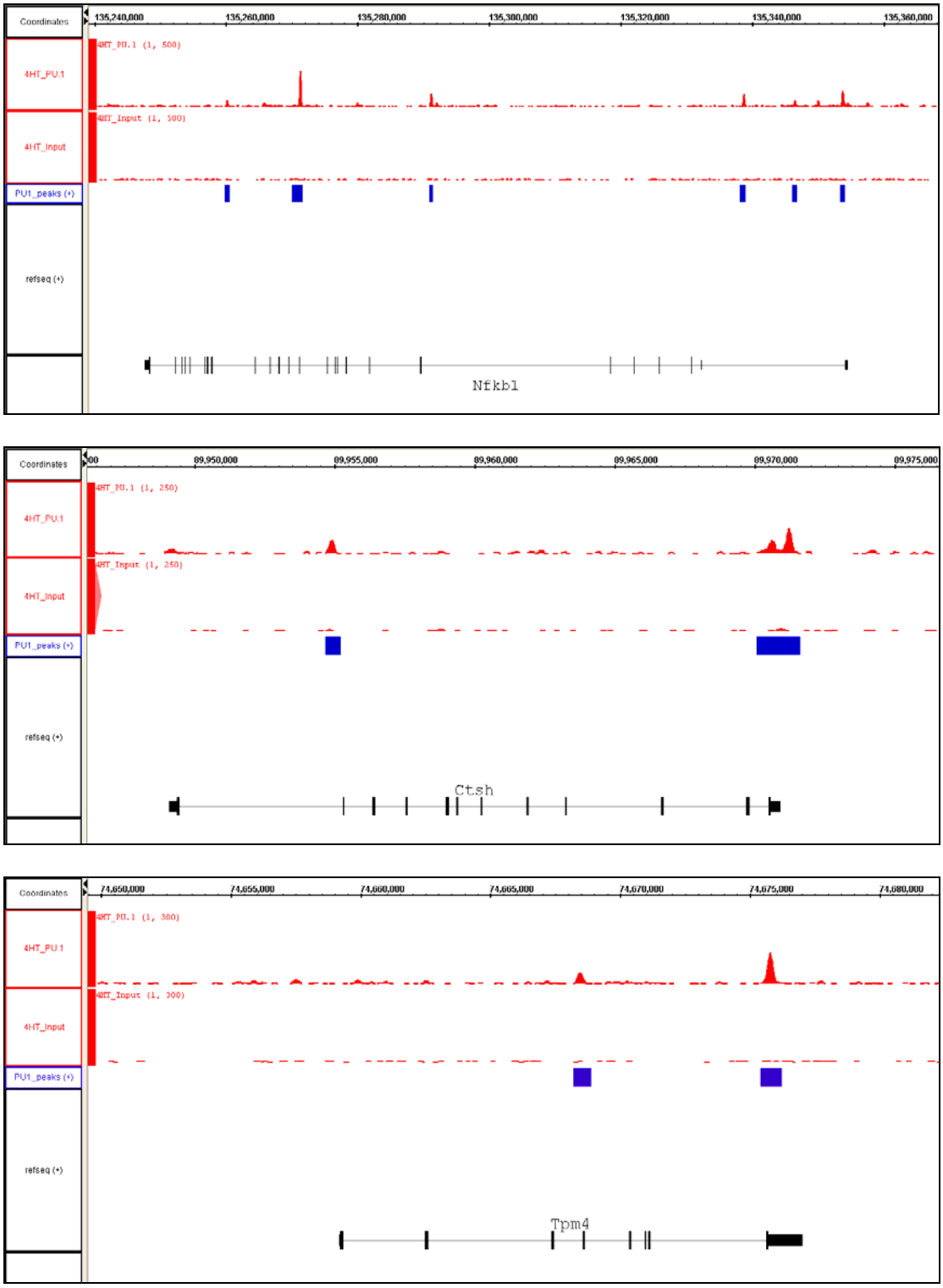


Figure S5.

PU.1 binding at gene loci of 15 PU.1 targets in MLL-ENL cells detected by ChIP-sequencing.

Figure S6 (related to Fig. 5)

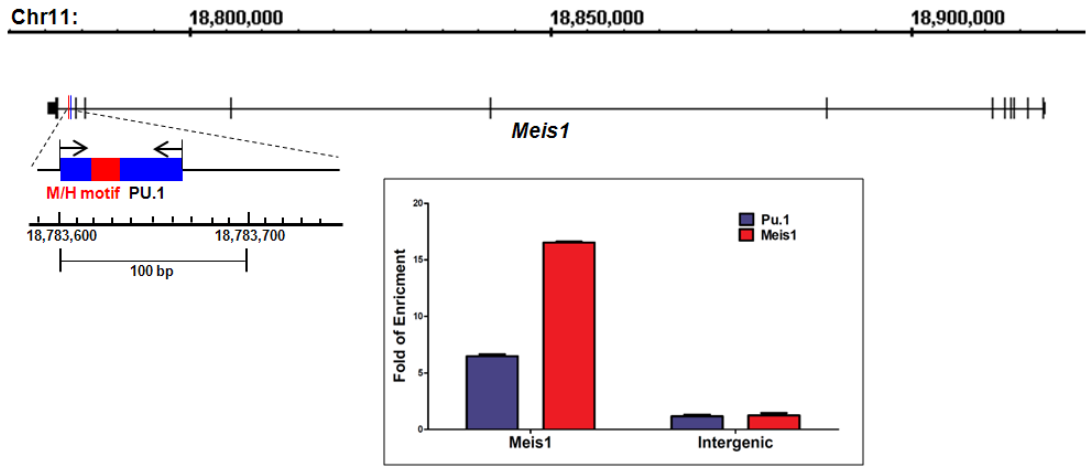
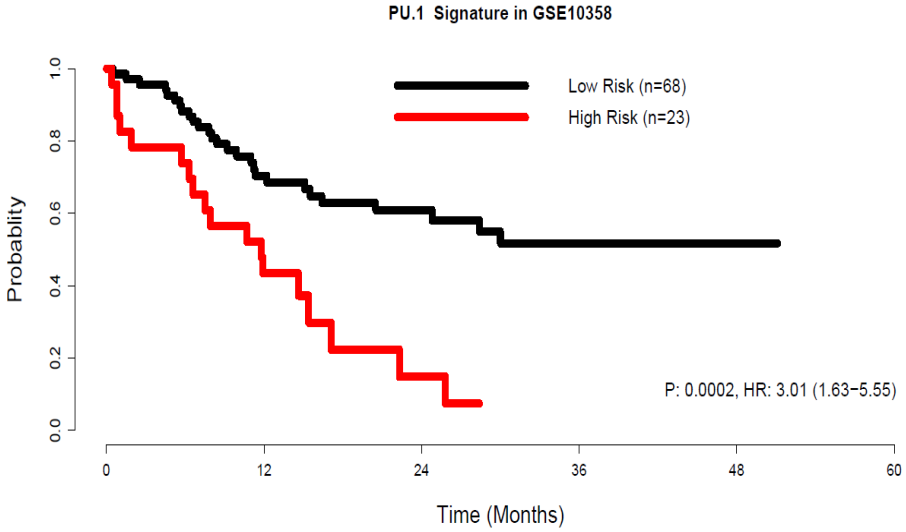


Figure S6. PU.1 and MEIS1 co-binding at the Meis1 locus. The DNA element bound by both PU.1 and MEIS1 is enlarged from the gene structure (chromosome coordinates from mm9 are indicated). Positions of the conserved MEIS1/HOXA9 motif (http://genome.ucsc.edu/) and the PU.1 binding signal detected by ChIP-qPCR were shown as red and blue rectangles, respectively. The paired arrows above the DNA element represent the primers used in detecting the binding of both MEIS1 and PU.1 The enrichments levels of MEIS1 and PU.1 detected by ChIP-qPCR are shown in the bar plot. Bar charts are mean ±SD.

Figure S7 (related to Fig.6)

A



B

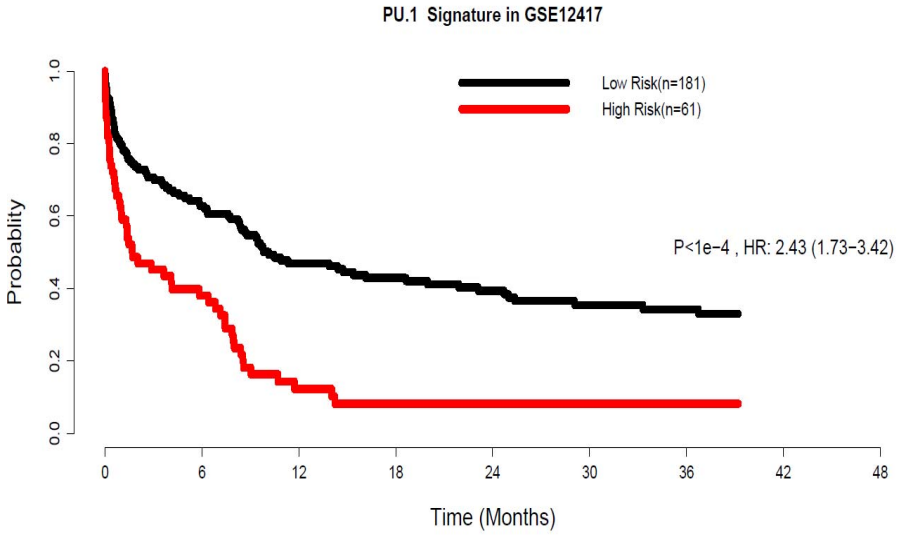


Figure S7. (A-B) Overall survival analysis of PU.1 15-gene signature in AML patients of GSE10358 (A, p=0.0002) and GSE12417 (B, p<1e-4).

Table S1. All datasets used in this study.

Datasets_ID	Data_type	Sample/Cell_type	Sample_Num
GSE1159	mRNA-Array	AML	285
GSE6891	mRNA-Array	AML	260
GSE10358	mRNA-Array	AML	279
GSE12417	mRNA-Array	AML	163
GSE13159	mRNA-Array	AML,ALL,NBM	2096
ROSS 2004	mRNA-Array	AML	150
GSE24794	ChIP-chip	MLL/H3K79me2 detection in a MLL-ENL fusion inducible system	10
GSE22178	PU.1 ChIP-seq	HPC-7 (a murine hematopoietic progenitor cell line)	1
GSE21314	PU.1 ChIP-seq	Macrophage (mouse)	1
In-House Data	mRNA-Array	LSK (mRUNX URE KI/KI mice)	2
In-House Data	PU.1 ChIP-seq	csh3 (MLL-ENL, mouse)	1

Table S2 (related to Fig.4 and Fig. 5). Primer sequences

ChIP-qPCR Primer in Fig. 5C:	
mMS/MS F	GCAAGGAGAGGAAAGAGCATGT
mMS/MS R	CATGCATCAACTGCGGTTAGA
mMS/LY86 F	CAGTCCAGCTTTCTATTGACTTTGG
mMS/LY86 R	GCACCCACAGGGCTTAGAGTT
mMS/pbx3 F	TGGGCTGGCAGCTTCACT
mMS/pbx3 R	TTCCCTTGCACAACAAATAGGA
mMS/KT F	GCCTCCCTTTGCTATCTCTTTG
mMS/KT R	TGCGGCGTGGGCTTAA
mMS/flt3 F	CGTAGGATTAGGATATGGGCACAT
mMS/flt3 R	GGGCAGCACCTCCTACACA
mMS/TBX1-1 F	CAAGGCAGACTCTCTTCTGGAAA
mMS/TBX1-1 R	TTGACAGGAAACATTTGCATAAGC
mMS/cd180 F	TTATCTTCTTTGAGGTGCCAGATTT
mMS/cd180 R	AACCGATGTTGGCTTTGGAA
pu/syk F	TGGCGGAGAAGGGTCTTTC
pu/sykR	AGGCCCCTGCTTGAAGCT
pu/nfkb1 F	GCTGGTATGGCAGGGAATGT
pu/nfkb1 R	TTTTCCCACCTTGGCTTCAC
ChIP-qPCR Primer in Fig. 4C:	
pu on ms F	CAACAGCAGCTCTCCAGGAGTT
pu on ms R	CGGCCCAGGAGCAGTTAA
pu on ly86 F	AACCAAACCGCCATGAACTC
pu on ly86 R	CATATCTTGCTCTGGGCATGTG
pu on pbx3 F	CATAACCAGAAAGCAAGGGTAGATG
pu on pbx3 R	CCCTTTCCTCGGTTCCTCAA
pu on flt F	TCTGTCAAGAGACTGTGGTTTCCT
pu on flt R	CTTAGCCGCCCGGTACTGT
pu on kit-F	GTGAATGGAAAGGTTGTAGGACAA
pu on kit R	CCTAGGACCCACACGGTAGAAG
pu on tbx F	AAACAAACAGGCTGGGTGTCTAC
pu on tbx R	TGAGAACCACTGTGCTGGACTCT
pu on cd180 F	GCCCTTCTGAGGCCACAA
pu on cd180 R	GCTCAGTACCTCCGCTGGAA
pu on csf F	GCAACAGACAGGAACGTGTTCA
pu on csf R	GGGCTCCCAGCTGCTAGTT
pu on PU F	CGGCCCATGACACAATACAC
pu on PU R	GGGCCAGCACCTCTCTGA

pu on nfkb1 F	CCTCGCCGCTCTGGAA
pu on nfkb1 R	TGCTTGTCAACCAACACAACATAA
pu on aif F	GCTGTGTGGAGCTAGGTGTGTT
pu on aif R	CTGCCTCAGCCTCCCAGAT
pu on syk F	GCTTACAGGCACATGTGAAGATG
pu on syk R	TGGTTCAGTGGGTGAGAACACTT
pu on ctsh F	CCAAGTGTGAGGGCCTGAGT
pu on ctsh R	TGCCCCACATCTGACTTTTTATG
pu on ccl3 F	CGGCCCATGACACAATACAC
pu on ccl3 R	GGGCCAGCACCTCTCTGA
pu on tpm4 F	TGGCTATGGGACCTGAGATGT
pu on tpm4 R	GACCCATCTCCACGGATACTCT
mMs1D130 F	CCAGCCCTTGAAACAATAGCA
mMs1D130 R	AATGGAGCATTTCTCTCCGCT
Expression Primer:	
mMeis1 F	AGTGCAGCCCATGATAGACC
mMeis1 R	TGCTGACCGTCCATTACAAA
mLy86 F	TGGAAGTAGTCTACCAGAGCTGTGA
mLy86 R	GGAACACTGGTCAATGGAAAGG
mPbx3 F	TGC CAA AAG CAT ATT GTC CA
mPbx3 R	CAG CGT CTT GTG TGA GAT CAA
mFlt3 F	AACTGGGCGTCATCATTTTC
mFlt3 R	GTGAACAGAGAGGCCTGGAG
c-Kit F	ATT GTG CTG GAT GGA TGG AT
c-Kit R	GAT CTG CTC TGC GTC CTG TT
mTbxas1 F	GAGGCTTCTGAAAGAGGTGG
mTbxas1 R	ATGTCCAGATACGGCAGACC
mCsf1r F	TTGCCTTCGTATCTCTCGATG
mCsf1r R	CTCTGCTGGTGCTACTGCTG
mAIF1 F	GAGGATCTGCCGTCCAAACTT
mAIF1 R	GATATCTCCATTTCCATTCAGATCAA
mPu.1 F	AGCGATGGAGAAAGCCATAG
mPu.1 R	TGCAGCTCTGTGAAGTGGTT
mNfkb1 F	GAACGATAACCTTTGCAGGC
mNfkb1 R	CATCACACGGAGGGCTTC
mSyk F	TCCTTTCAACGTTCCATGCT
mSyk R	TCTGCACCCCTTCAGAGTTC
mCtsh F	CCTGGCTGCTGAGTACTGG
mCtsh R	TCCACCGAGCTGTACGTCTT
mCD180 F	CATGAAAGGCTTCCTGCTTC
mCD180 R	GTGCCACCAGAAGACGATGT
mCCL3 F	TCCCAGCCAGGTGTCATTTT
mCCL3 R	TTGGAGTCAGCGCAGATCTG

mTPM4 F	TCTAAACTGGAGAAGACAATCGATGA
mTPM4 R	AAGCCCACATTCTCTTCTTTGG
mPgk F	GGGACTGTCATCCTGCTGGA
mPgk R	CATCAATTTTGGCCGGCTC
mGAPDH F	AGGTCGGTGTGAACGGATTTG
mGAPDH R	TGTAGACCATGTAGTTGAGGTCA

Supplementary Materials and Methods

Survival Analysis

Kaplan-Meier method was used to estimate overall survival (OS), and log-rank test was used to compare prognostic difference among different groups. To evaluate the potential effect of our gene list on patient outcome, a risk score generated from Cox proportional hazards regression was used to associate the expression level with prognosis. The 75% quantile of the risk score was used to classify patients into high or low risk group, in which the high risk group indicated poorer survival for patients[2]. To investigate the prognostic ability across all patients, the hazard ratio and 95% confidence interval in each dataset were combined using a random-effects model in the meta-analysis[3], and a unified p-value was also generated.

To address a recent concern that many published gene signatures are not significantly better prognostic predictors than random ones with the same size[4], we used 10,000-time gene sampling to generate a Monte Carlo p-value[5]. In each sampling, a random signature with matched genes in each dataset was extracted to calculate a p-value by meta-analysis, representing the prognostic ability of any 40-gene signature by chance:

$$p = (r + 1) / (n + 1)$$

Here n is the simulation time, and r is the number that p-value generated

from random meta-analysis, which is smaller than the observed p-value obtained from real meta-analysis.

Supplemental References

1. Wang Q-f, Wu G, Mi S, He F, Wu J, Dong J*, et al.* MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. *Blood* 2011; **117**(25)**:** 6895-6905.

2. Chen H-Y, Yu S-L, Chen C-H, Chang G-C, Chen C-Y, Yuan A*, et al.* A five-gene signature and clinical outcome in non–small-cell lung cancer. *New England Journal of Medicine* 2007; **356**(1)**:** 11-20.

3. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled clinical trials* 1986; **7**(3)**:** 177-188.

4. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS computational biology* 2011; **7**(10)**:** e1002240.

5. North BV, Curtis D, Sham PC. A note on the calculation of empirical P values from Monte Carlo procedures. *American journal of human genetics* 2002; **71**(2)**:** 439.