Table S1. Functional categorization of recurrently mutated genes in AML

DNA methylation	count
DNMT3A	43
IDH1	16
IDH2	17
TET1	2
TET2	15

Chromatin modifiers	count
ASXL1	3
ASXL2	1
ASXL3	1
BRPF1	1
CBX7	1
EED	2
EZH2	3
HDAC2	1
HDAC3	1
JMJD1C	1
MLL2	1
MLL3	1
MTA2	1
PRDM16	1
PRDM9	1
RBBP4	1
SAP130	1
SCML2	1
SUDS3	1
SUZ12	3
ZBTB33	2
ZBTB7B	1

Myeloid TFs	count
CBFB	2
CEBPA	4
ETV3	1
ETV6	2
GATA2	2
GLI1	1
MYB	1
RUNX1	16

Spliceosome	count
CSTF2T	1
DDX1	1
DDX23	1
DHX32	1
HNRNPK	2
METTL3	1
PRPF3	1
PRPF8	2
RBMX	1
SF3B1	1
SNRNP200	1
SRRM2	1
TRA2B	2
U2AF1	7
U2AF1L4	1
U2AF2	1

Cohesins	count
RAD21	5
SMC1A	7
SMC3	7
SMC5	1
STAG2	5

Tumor suppressors	count
PHF6	5
TP53	15
WT1	10

Kinases	count
ABL1	1
DYRK4	1
EPHA2	1
EPHA3	1
FLT3	49
JAK3	1
KIT	7
MST1R	1
PDGFRB	1
WEE1	1

Table S2. Statistical analysis of recurrent mutations in HOX expression groups

	Counts				Percent				
	NOHOX	HOXA	HOXB	HOXAB	NOHOX	HOXA	HOXB	HOXAB	Adjusted P-value*
NPM1	**0**	**1**	**0**	**47**	**0.0%**	**2.4%**	**0.0%**	**58.8%**	**7.74E-15**
DNMT3A	**3**	**6**	**4**	**30**	**8.6%**	**14.3%**	**18.2%**	**37.5%**	**0.039255112**
FLT3	7	6	3	33	20.0%	14.3%	13.6%	41.3%	0.052671238
IDH1	1	3	1	11	2.9%	7.1%	4.5%	13.8%	1
IDH2	1	7	1	8	2.9%	16.7%	4.5%	10.0%	1
TET2	3	3	3	6	8.6%	7.1%	13.6%	7.5%	1
RUNX1	0	5	5	6	0.0%	11.9%	22.7%	7.5%	0.50087433
TP53	0	6	2	7	0.0%	14.3%	9.1%	8.8%	1
NRAS	1	1	2	9	2.9%	2.4%	9.1%	11.3%	1
WT1	1	1	2	6	2.9%	2.4%	9.1%	7.5%	1
PTPN11	0	1	1	6	0.0%	2.4%	4.5%	7.5%	1
KIT	2	2	3	0	5.7%	4.8%	13.6%	0.0%	0.50087433
KRAS	0	2	1	4	0.0%	4.8%	4.5%	5.0%	1
MT.CO2	2	3	0	3	5.7%	7.1%	0.0%	3.8%	1
TTN	1	1	2	3	2.9%	2.4%	9.1%	3.8%	1
U2AF1	0	2	2	3	0.0%	4.8%	9.1%	3.8%	1
SMC1A	1	2	0	4	2.9%	4.8%	0.0%	5.0%	1
SMC3	1	1	1	4	2.9%	2.4%	4.5%	5.0%	1
MT.CYB	0	1	1	3	0.0%	2.4%	4.5%	3.8%	1
PHF6	0	1	3	1	0.0%	2.4%	13.6%	1.3%	0.218412682
FAM5C	1	0	1	3	2.9%	0.0%	4.5%	3.8%	1
MUC16	0	4	0	1	0.0%	9.5%	0.0%	1.3%	0.500393596
RAD21	1	0	0	4	2.9%	0.0%	0.0%	5.0%	1
STAG2	0	1	0	4	0.0%	2.4%	0.0%	5.0%	1
Total	35	42	22	80					

* Chi-squared test, bonferroni correction for multiple comparisons

Supplementary Figure Legends

Figure S1. Comparison of HOX gene expression between RNA-sequencing, Affymetrix U133+2 microarray, and custom Nanostring nCounter array. Panels A, B, and C show expression values for all genes from each of the four HOX clusters from RNA-seq, microarray and Nanostring platforms, respectively. All expression values are shown on a linear scale. Panels D and E show direct comparison of HOXA (in red) and HOXB (in blue) gene expression values RNA-seq and microarray expression values with Nanostring expression, respectively. Panel F shows the expression of *MEIS1* versus the total expression of all HOXA and HOXB genes (as a sum of the individual FPKM values), in red and blue respectively. Filled points show AML samples and open points show normal bone marrow cell populations.

Figure S2. RNA-seq-based HOXA and HOXB gene expression across AMLs with different recurrent mutations. All expression values are log2 (FPKM+1). A) HOX expression across AMLs with different recurrent gene fusions, including *MLLX* translocations, t(15;17)/*PML-RARA*, t(8;21)/*RUNX1-RUNX1T1*, and inv(16)/*CBFB-MYH11*. B) Comparison of HOXA and HOXB expression between *MLLX* translocations and *MLL* partial tandem duplications (PTD). C) Comparison of HOXA and HOXB expression between normal karyotype AMLs with and without NPMc mutations.

Figure S3. RNA-seq expression levels of selected HOXA genes (panel A) and HOXB genes (panel B) from AML samples with distinct HOX expression patterns and normal bone marrow cell populations. Only genes with evidence of expression in Figures 1 and S1 are shown. The AML samples shown include all those with the HOXA only pattern (in panel A, n=42), the HOXB only pattern (in panel B, n=22) and the HOXA and HOXB stem cell-like pattern (shown in both panel A and panel B, n=80). P-values reflect nonparametric significance testing (Wilcox rank sum test) between AMLs with the indicated expression patterns, and between all AML samples and normal $CD34^+$ cells, and are not corrected for multiple comparisons.

Figure S4. A) HOXC and HOXD expression in normal bone marrow cells from RNA-seq. Each plot shows RNA-seq expression of 29 total samples for the indicated HOXC/HOXD gene in log2 FPKM+1, including $CD34^+$ cells (n=20), promyelocytes (n=3), monocytes (n=3), and neutrophils (n=3). B) HOXA-HOXD expression of hematopoietic stem/progenitor cells, bulk $CD34^+$ cells, and more mature myeloid populations from the Affymetrix U133+2 microarray platform.

Figure S5. Heatmap showing unsupervised hierarchical clustering of microarray-based HOXA and HOXB expression values for survival analysis. Data are from GSE10358 (ref. 44) and are log2 transformed. Mutation and cytogenetic data are indicated in the matrix below the heatmap. The observed HOX expression phenotypes (no HOX, HOXA only, HOXB only, and HOXA/HOXB) are also indicated in the matrix. Note that only normal karyotype and intermediate risk samples (n=162) were used for the survival analysis shown in Figure 1D.

Figure S6. HOXA and HOXB transcript isoform and antisense noncoding transcript analysis. A) Scatter plot of HOXA (in red) and HOXB (in blue) antisense expression vs. HOX coding expression, each expressed as the sum FPKM of all antisense noncoding and coding transcripts, respectively. Data from normal $CD34^+$ cells are also shown in open circles. B) Unsupervised heatmap of HOXA and HOXB antisense noncoding expression, showing mutation-specific clustering that mirrors the patterns seen with coding HOX gene expression.

Figure S7. Illumina 450K methylation array data from the HOXA and HOXB loci from primary AML samples included in the TCGA AML project (ref. 32). A) Methylation values from 493 CpGs at the HOXA locus from AML samples with PML-RARA (red, n=18), RUNX1-RUNX1T1 (blue, n=7), MLLX translocations (green, n=11), and normal karyotype AMLs with the NPMc mutation (purple, n=53). Note the region highlighted in gray, which is identical to the highlighted region in Figure 4A. B) Methylation values from 306 CpGs at the HOXB locus. Samples are identical to panel A.

Figure S1

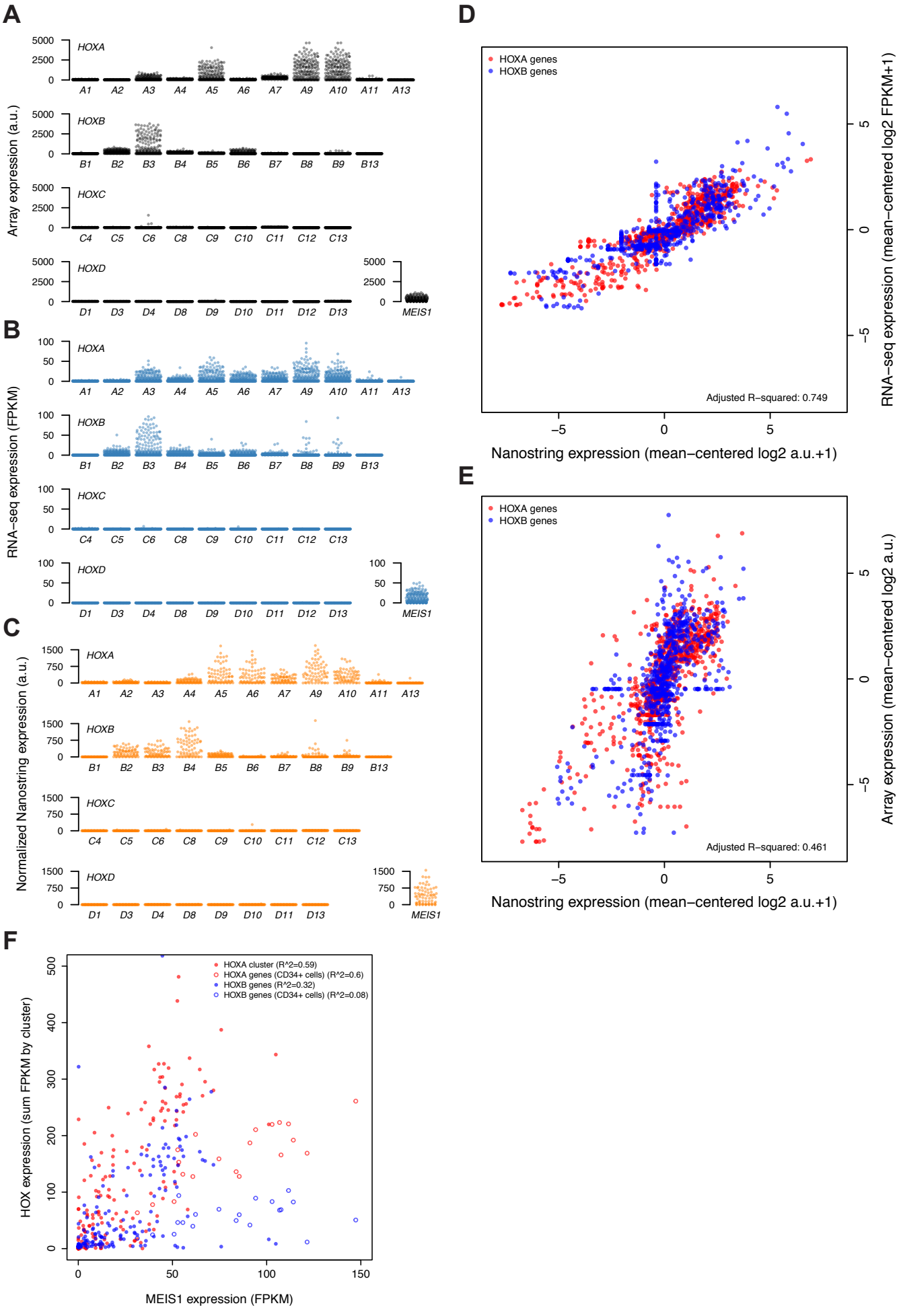

A
HOXA
HOXB
HOXC
HOXD
MEIS1
Array expression (a.u.)
B
RNA−seq expression (FPKM)
C
Normalized Nanostring expression (a.u.)
D
HOXA genes
HOXB genes
Adjusted R−squared: 0.749
Nanostring expression (mean−centered log2 a.u.+1)
RNA−seq expression (mean−centered log2 FPKM+1)
E
HOXA genes
HOXB genes
Adjusted R−squared: 0.461
Nanostring expression (mean−centered log2 a.u.+1)
Array expression (mean−centered log2 a.u.)
F
HOXA cluster (R^2=0.59)
HOXA genes (CD34+ cells) (R^2=0.6)
HOXB genes (R^2=0.32)
HOXB genes (CD34+ cells) (R^2=0.08)
HOX expression (sum FPKM by cluster)
MEIS1 expression (FPKM)

Figure S2

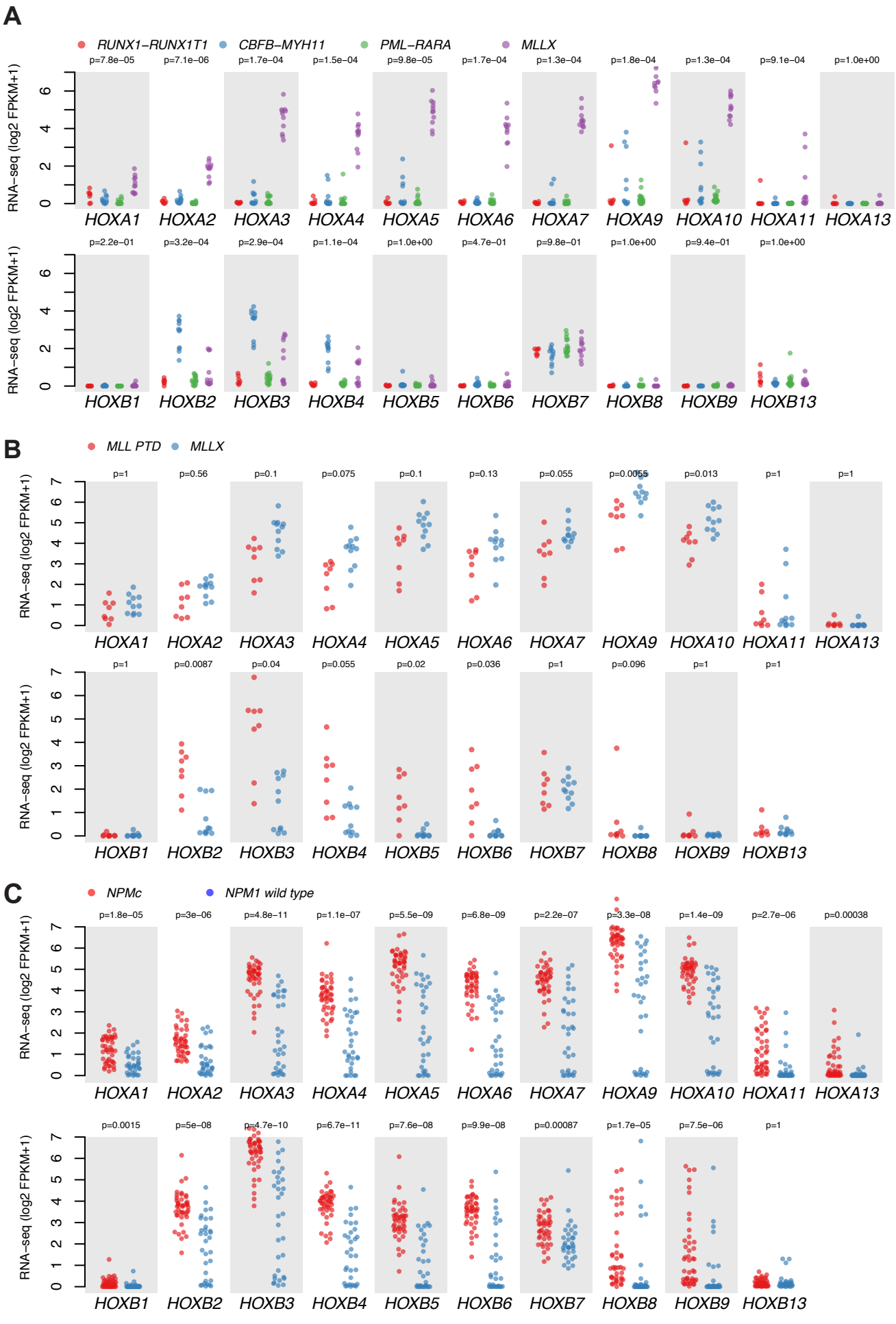

A
RUNX1–RUNX1T1
CBFB–MYH11
PML–RARA
MLLX
RNA–seq (log2 FPKM+1)
p=7.8e–05
p=7.1e–06
p=1.7e–04
p=1.5e–04
p=9.8e–05
p=1.7e–04
p=1.3e–04
p=1.8e–04
p=1.3e–04
p=9.1e–04
p=1.0e+00
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
p=2.2e–01
p=3.2e–04
p=2.9e–04
p=1.1e–04
p=1.0e+00
p=4.7e–01
p=9.8e–01
p=1.0e+00
p=9.4e–01
p=1.0e+00
HOXB1
HOXB2
HOXB3
HOXB4
HOXB5
HOXB6
HOXB7
HOXB8
HOXB9
HOXB13
B
MLL PTD
MLLX
p=1
p=0.56
p=0.1
p=0.075
p=0.1
p=0.13
p=0.055
p=0.0055
p=0.013
p=1
p=1
p=1
p=0.0087
p=0.04
p=0.055
p=0.02
p=0.036
p=1
p=0.096
p=1
p=1
C
NPMc
NPM1 wild type
p=1.8e–05
p=3e–06
p=4.8e–11
p=1.1e–07
p=5.5e–09
p=6.8e–09
p=2.2e–07
p=3.3e–08
p=1.4e–09
p=2.7e–06
p=0.00038
p=0.0015
p=5e–08
p=4.7e–10
p=6.7e–11
p=7.6e–08
p=9.9e–08
p=0.00087
p=1.7e–05
p=7.5e–06
p=1

Figure S3

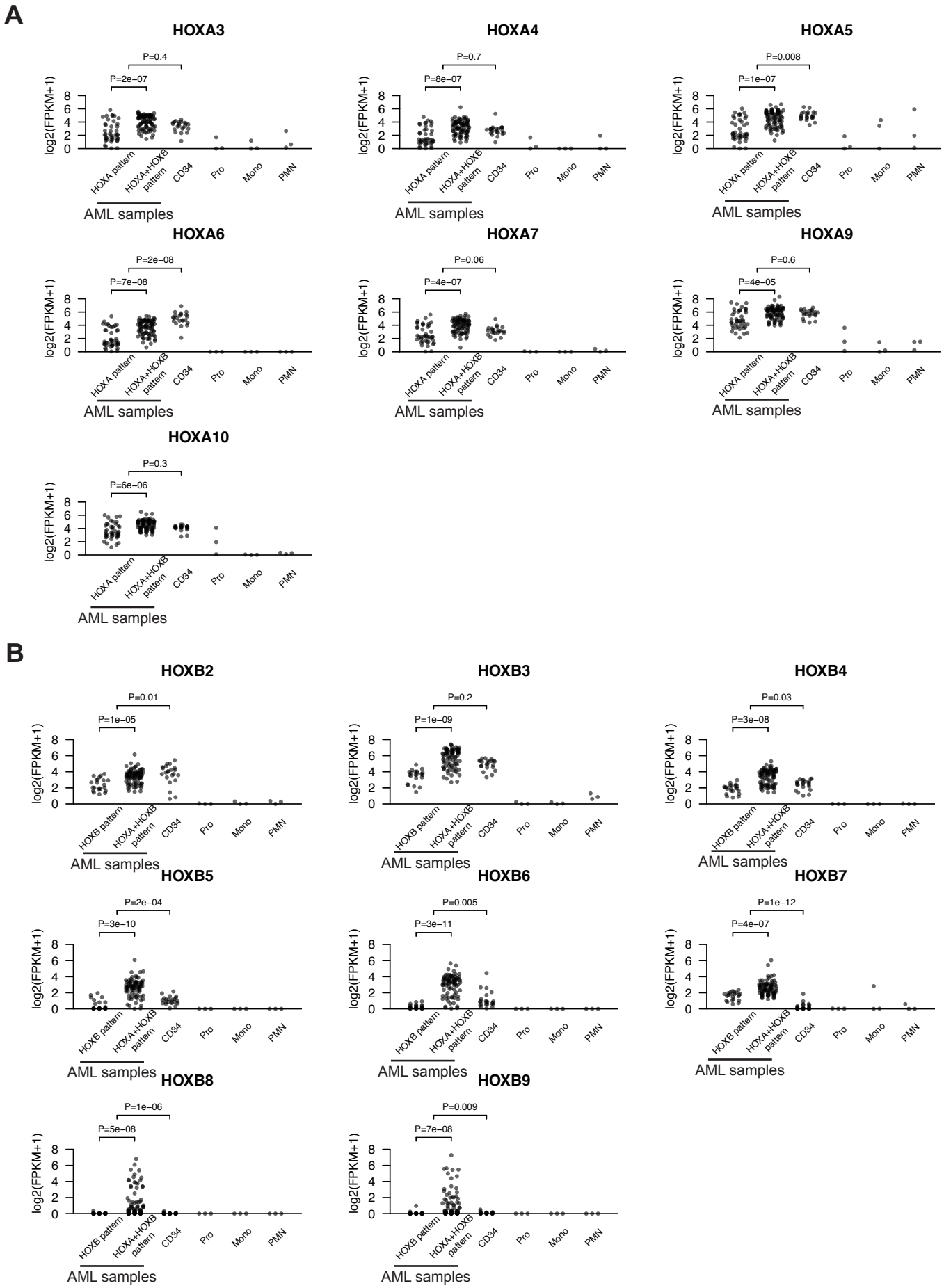

A
HOXA3
P=0.4
P=2e−07
HOXA4
P=0.7
P=8e−07
HOXA5
P=0.008
P=1e−07
HOXA6
P=2e−08
P=7e−08
HOXA7
P=0.06
P=4e−07
HOXA9
P=0.6
P=4e−05
HOXA10
P=0.3
P=6e−06
log2(FPKM+1)
HOXA pattern
HOXA+HOXB pattern
CD34
Pro
Mono
PMN
AML samples
B
HOXB2
P=0.01
P=1e−05
HOXB3
P=0.2
P=1e−09
HOXB4
P=0.03
P=3e−08
HOXB5
P=2e−04
P=3e−10
HOXB6
P=0.005
P=3e−11
HOXB7
P=1e−12
P=4e−07
HOXB8
P=1e−06
P=5e−08
HOXB9
P=0.009
P=7e−08
HOXB pattern
HOXA+HOXB pattern

Figure S4

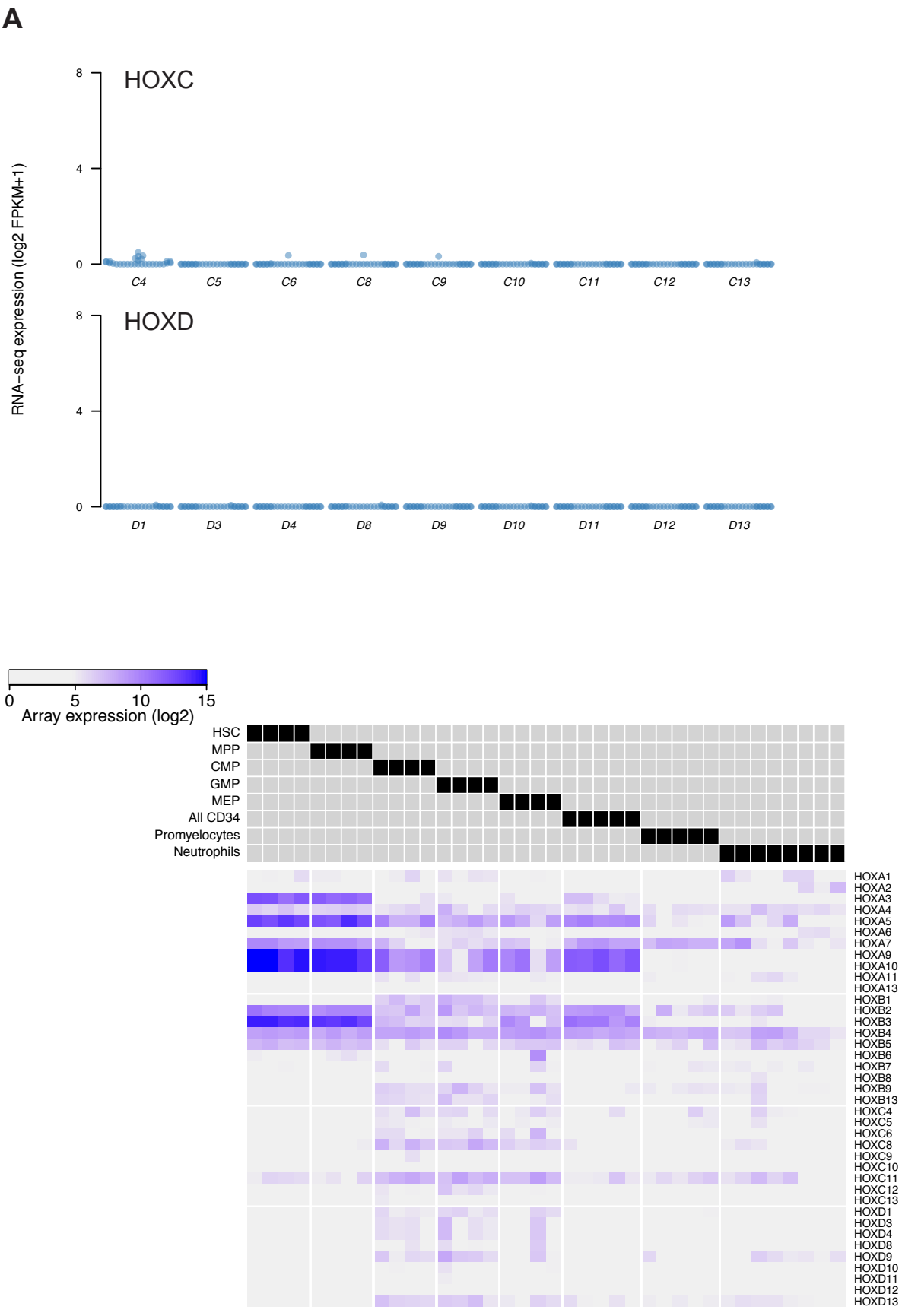

A
HOXC
HOXD
RNA−seq expression (log2 FPKM+1)
8
4
0
C4
C5
C6
C8
C9
C10
C11
C12
C13
D1
D3
D4
D8
D9
D10
D11
D12
D13
B
0
5
10
15
Array expression (log2)
HSC
MPP
CMP
GMP
MEP
All CD34
Promyelocytes
Neutrophils
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
HOXB1
HOXB2
HOXB3
HOXB4
HOXB5
HOXB6
HOXB7
HOXB8
HOXB9
HOXB13
HOXC4
HOXC5
HOXC6
HOXC8
HOXC9
HOXC10
HOXC11
HOXC12
HOXC13
HOXD1
HOXD3
HOXD4
HOXD8
HOXD9
HOXD10
HOXD11
HOXD12
HOXD13

Figure S5

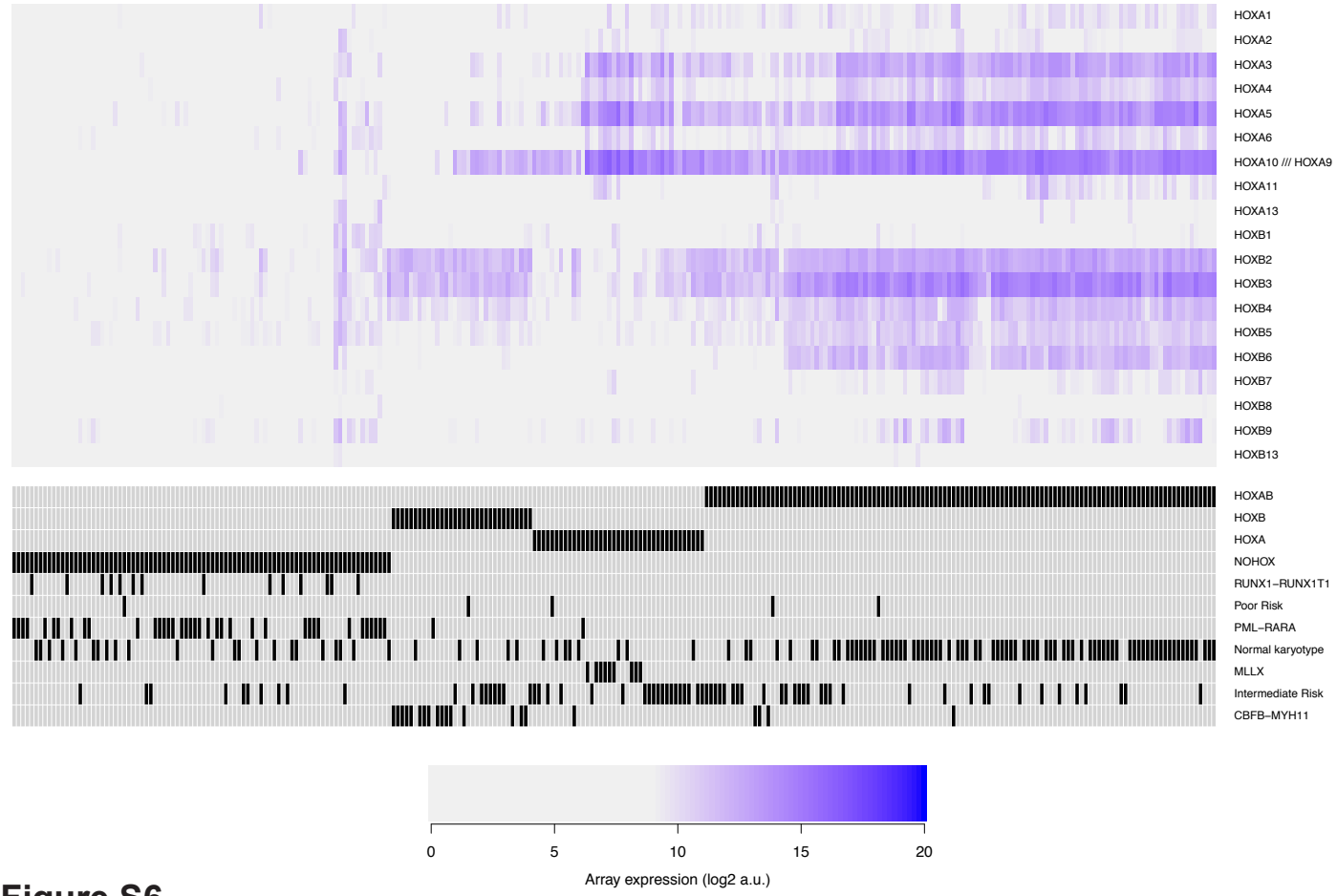


Figure S6

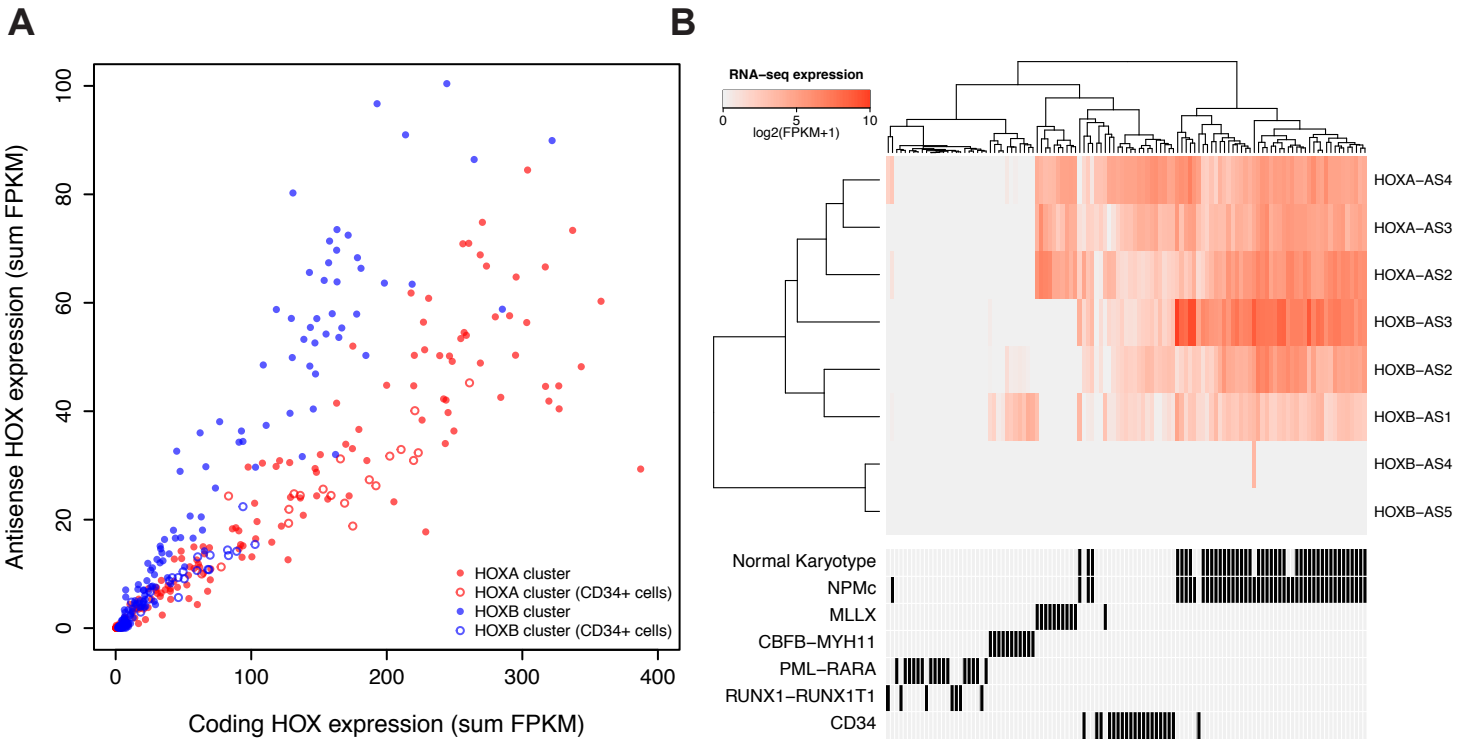
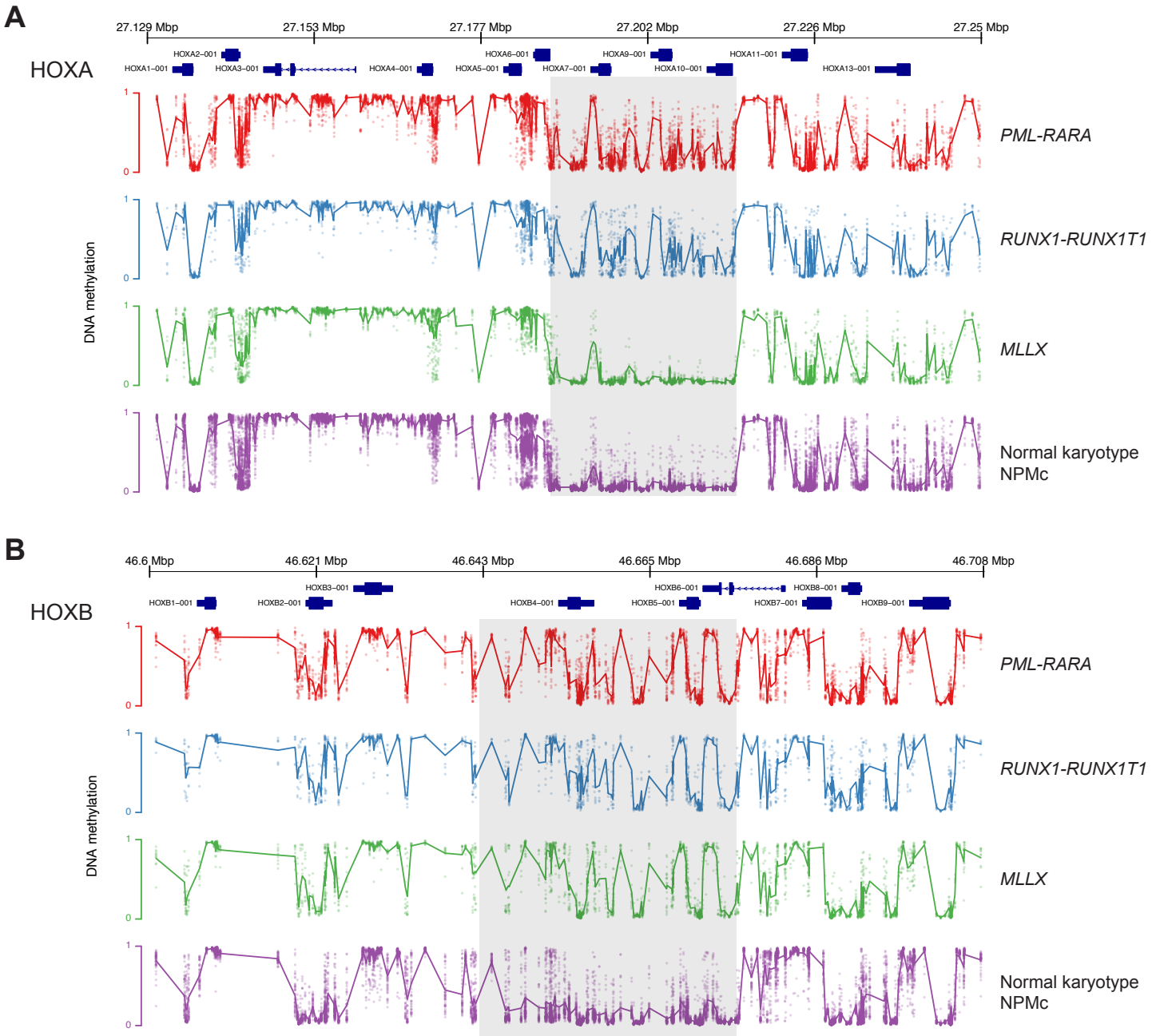


Figure S7


A
HOXA
27.129 Mbp
27.153 Mbp
27.177 Mbp
27.202 Mbp
27.226 Mbp
27.25 Mbp
HOXA1–001
HOXA2–001
HOXA3–001
HOXA4–001
HOXA5–001
HOXA6–001
HOXA7–001
HOXA9–001
HOXA10–001
HOXA11–001
HOXA13–001
DNA methylation
PML-RARA
RUNX1-RUNX1T1
MLLX
Normal karyotype NPMc
B
HOXB
46.6 Mbp
46.621 Mbp
46.643 Mbp
46.665 Mbp
46.686 Mbp
46.708 Mbp
HOXB1–001
HOXB2–001
HOXB3–001
HOXB4–001
HOXB5–001
HOXB6–001
HOXB7–001
HOXB8–001
HOXB9–001
DNA methylation
PML-RARA
RUNX1-RUNX1T1
MLLX
Normal karyotype NPMc