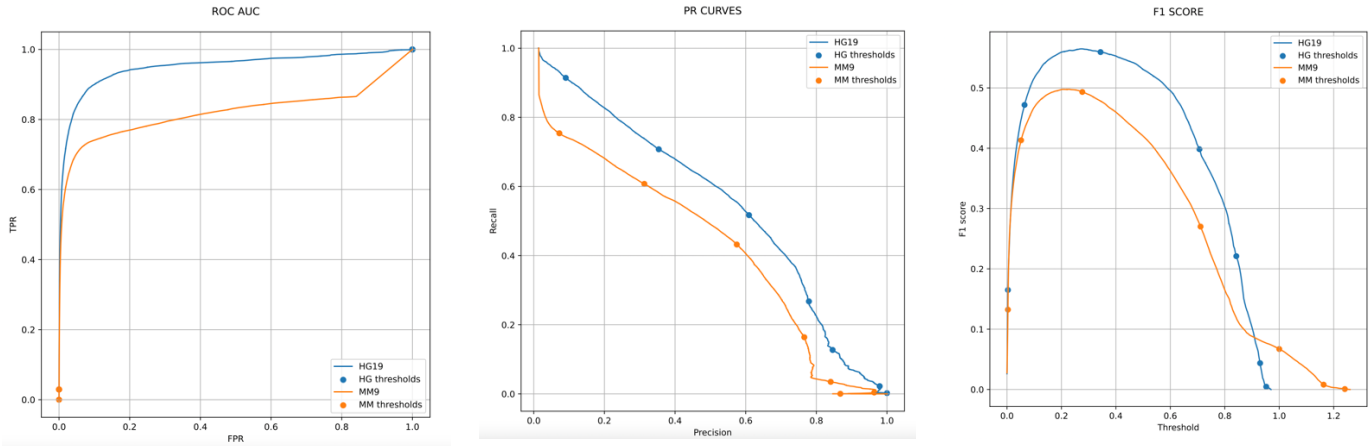
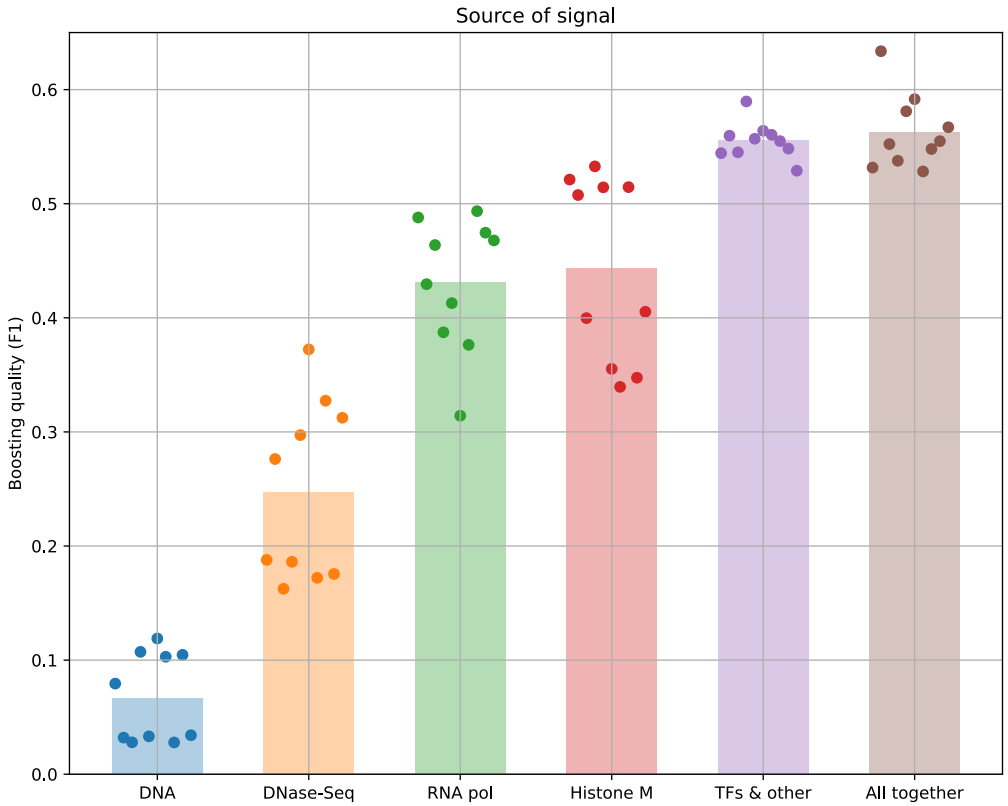
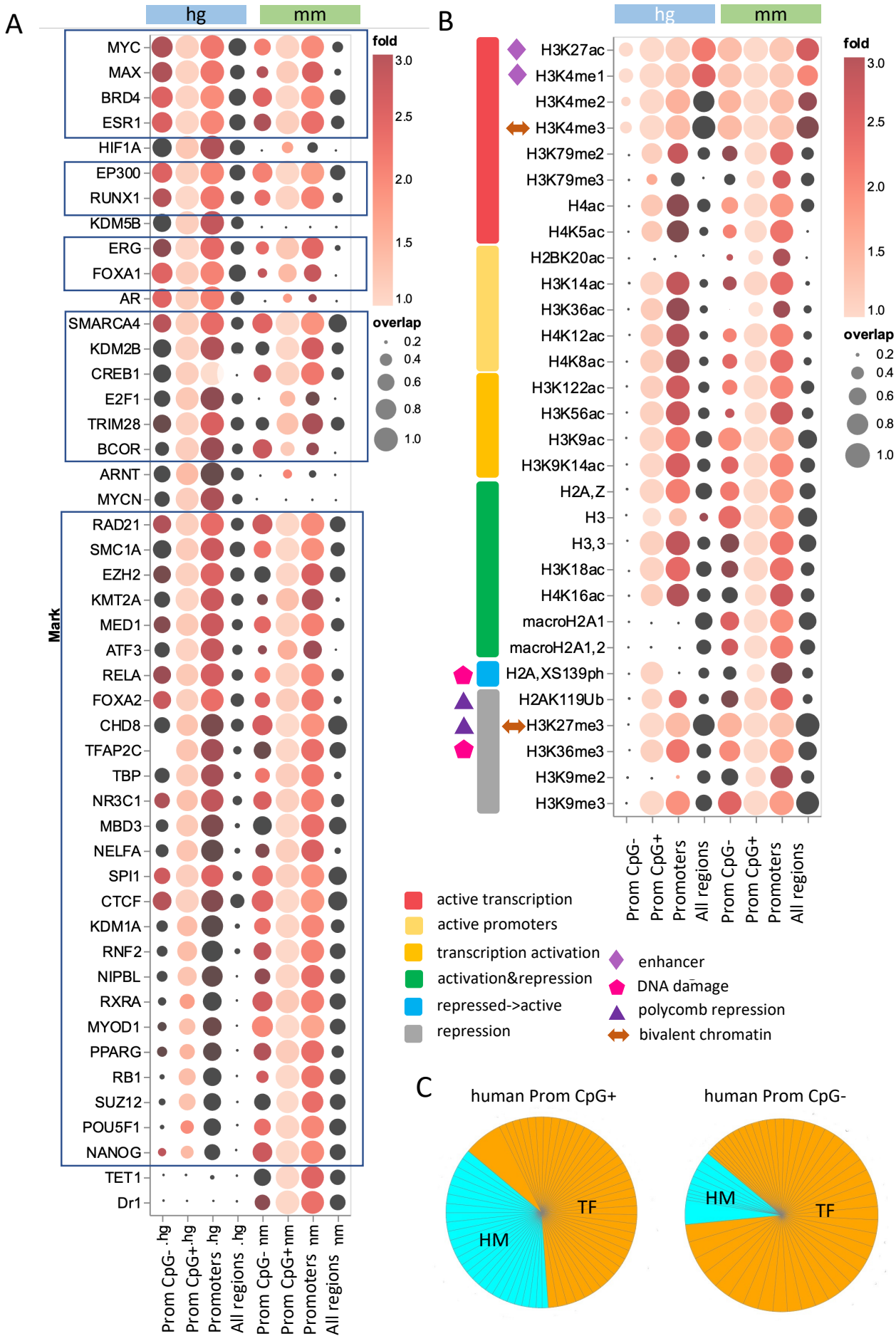


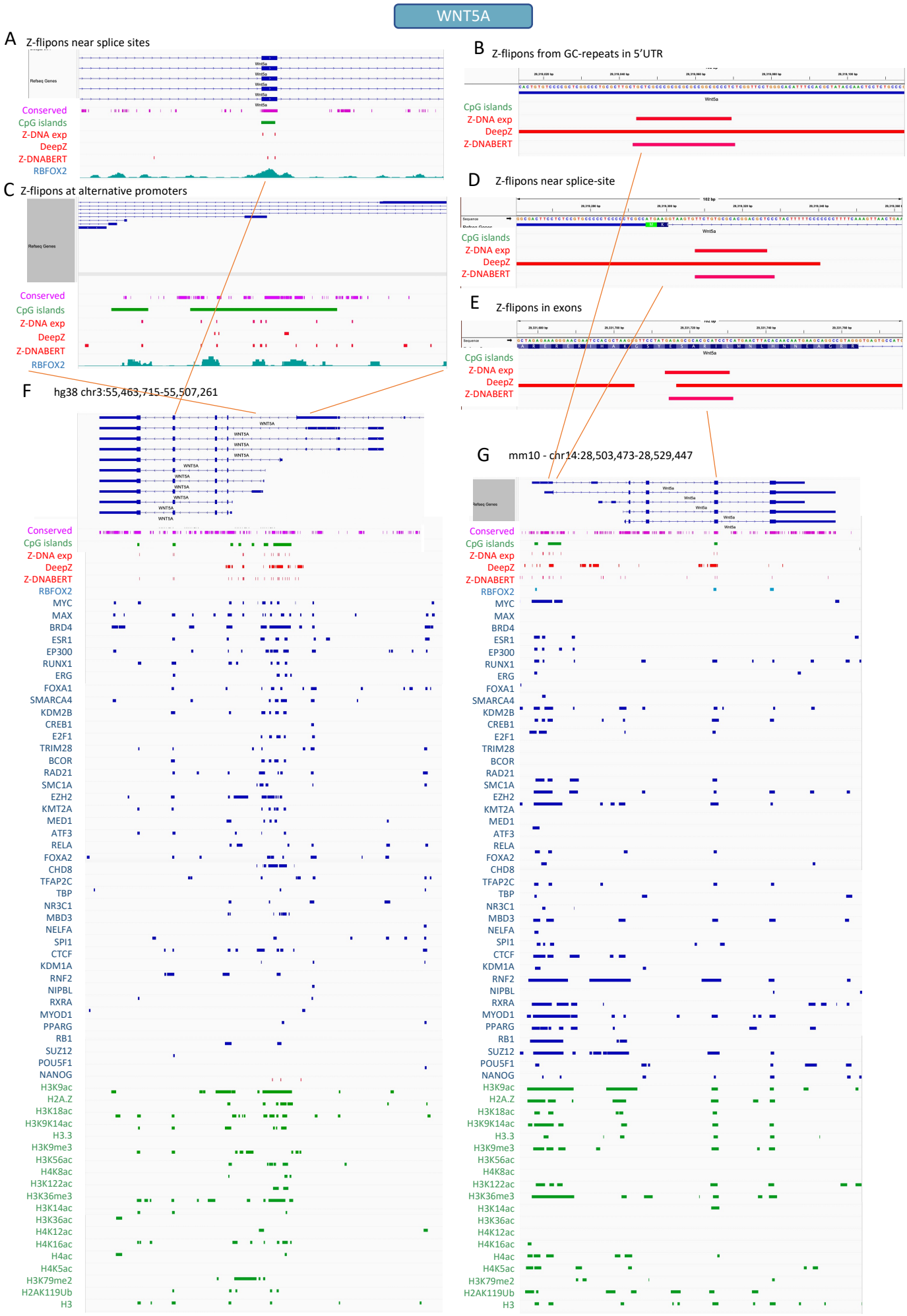
Supplementary Figure 1. ROC-, PR-, and F-curves of DeepZ models for human and mouse genomes. DeepZ models were trained using the same number of omics features.



Supplementary Figure 2. Ablation study for group feature importance performed with Gradient Boosting models on reduced data sets for human and mouse (see Methods for details).


A
hg
mm
fold
3.0
2.5
2.0
1.5
1.0
overlap
0.2
0.4
0.6
0.8
1.0
MYC
MAX
BRD4
ESR1
HIF1A
EP300
RUNX1
KDM5B
ERG
FOXA1
AR
SMARCA4
KDM2B
CREB1
E2F1
TRIM28
BCOR
ARNT
MYCN
RAD21
SMC1A
EZH2
KMT2A
MED1
ATF3
RELA
FOXA2
CHD8
TFAP2C
TBP
NR3C1
MBD3
NELFA
SPI1
CTCF
KDM1A
RNF2
NIPBL
RXRA
MYOD1
PPARG
RB1
SUZ12
POU5F1
NANOG
TET1
Dr1
Mark
Prom CpG- .hg
Prom CpG+.hg
Promoters .hg
All regions .hg
Prom CpG- mm
Prom CpG+mm
Promoters mm
All regions mm
B
hg
mm
fold
3.0
2.5
2.0
1.5
1.0
overlap
0.2
0.4
0.6
0.8
1.0
H3K27ac
H3K4me1
H3K4me2
H3K4me3
H3K79me2
H3K79me3
H4ac
H4K5ac
H2BK20ac
H3K14ac
H3K36ac
H4K12ac
H4K8ac
H3K122ac
H3K56ac
H3K9ac
H3K9K14ac
H2A,Z
H3
H3,3
H3K18ac
H4K16ac
macroH2A1
macroH2A1,2
H2A,XS139ph
H2AK119Ub
H3K27me3
H3K36me3
H3K9me2
H3K9me3
Prom CpG-
Prom CpG+
Promoters
All regions
Prom CpG-
Prom CpG+
Promoters
All regions
active transcription
active promoters
transcription activation
activation&repression
repressed->active
repression
enhancer
DNA damage
polycomb repression
bivalent chromatin
C
human Prom CpG+
TF
HM
human Prom CpG-
HM
TF

Supplementary Figure 3. Patterns of transcription factors and histone marks enriched around all Deep-Z predicted Z-flipons in human and mouse genome. **A.** Enrichment of transcription factors around Z-flipons for the entire genome, promoters, CpG-promoters and non-CpG promoters. **B.** Enrichment of histone marks around Z-flipons. **C.** Share of enriched histone marks and transcription factors around Z-flipons for CpG- and non-CpG promoters.


WNT5A
A
Z-flipons near splice sites
Refseq Genes
Conserved
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
B
Z-flipons from GC-repeats in 5'UTR
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
C
Z-flipons at alternative promoters
Refseq Genes
Conserved
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
D
Z-flipons near splice-site
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
E
Z-flipons in exons
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
F
hg38 chr3:55,463,715-55,507,261
WNT5A
Conserved
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
MYC
MAX
BRD4
ESR1
EP300
RUNX1
ERG
FOXA1
SMARCA4
KDM2B
CREB1
E2F1
TRIM28
BCOR
RAD21
SMC1A
EZH2
KMT2A
MED1
ATF3
RELA
FOXA2
CHD8
TFAP2C
TBP
NR3C1
MBD3
NELFA
SPI1
CTCF
KDM1A
RNF2
NIPBL
RXRA
MYOD1
PPARG
RB1
SUZ12
POU5F1
NANOG
H3K9ac
H2A.Z
H3K18ac
H3K9K14ac
H3.3
H3K9me3
H3K56ac
H4K8ac
H3K122ac
H3K36me3
H3K14ac
H3K36ac
H4K12ac
H4K16ac
H4ac
H4K5ac
H3K79me2
H2AK119Ub
H3
G
mm10 - chr14:28,503,473-28,529,447
Refseq Genes
Conserved
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
MYC
MAX
BRD4
ESR1
EP300
RUNX1
ERG
FOXA1
SMARCA4
KDM2B
CREB1
E2F1
TRIM28
BCOR
RAD21
SMC1A
EZH2
KMT2A
MED1
ATF3
RELA
FOXA2
CHD8
TFAP2C
TBP
NR3C1
MBD3
NELFA
SPI1
CTCF
KDM1A
RNF2
NIPBL
RXRA
MYOD1
PPARG
RB1
SUZ12
POU5F1
NANOG
H3K9ac
H2A.Z
H3K18ac
H3K9K14ac
H3.3
H3K9me3
H3K56ac
H4K8ac
H3K122ac
H3K36me3
H3K14ac
H3K36ac
H4K12ac
H4K16ac
H4ac
H4K5ac
H3K79me2
H2AK119Ub
H3

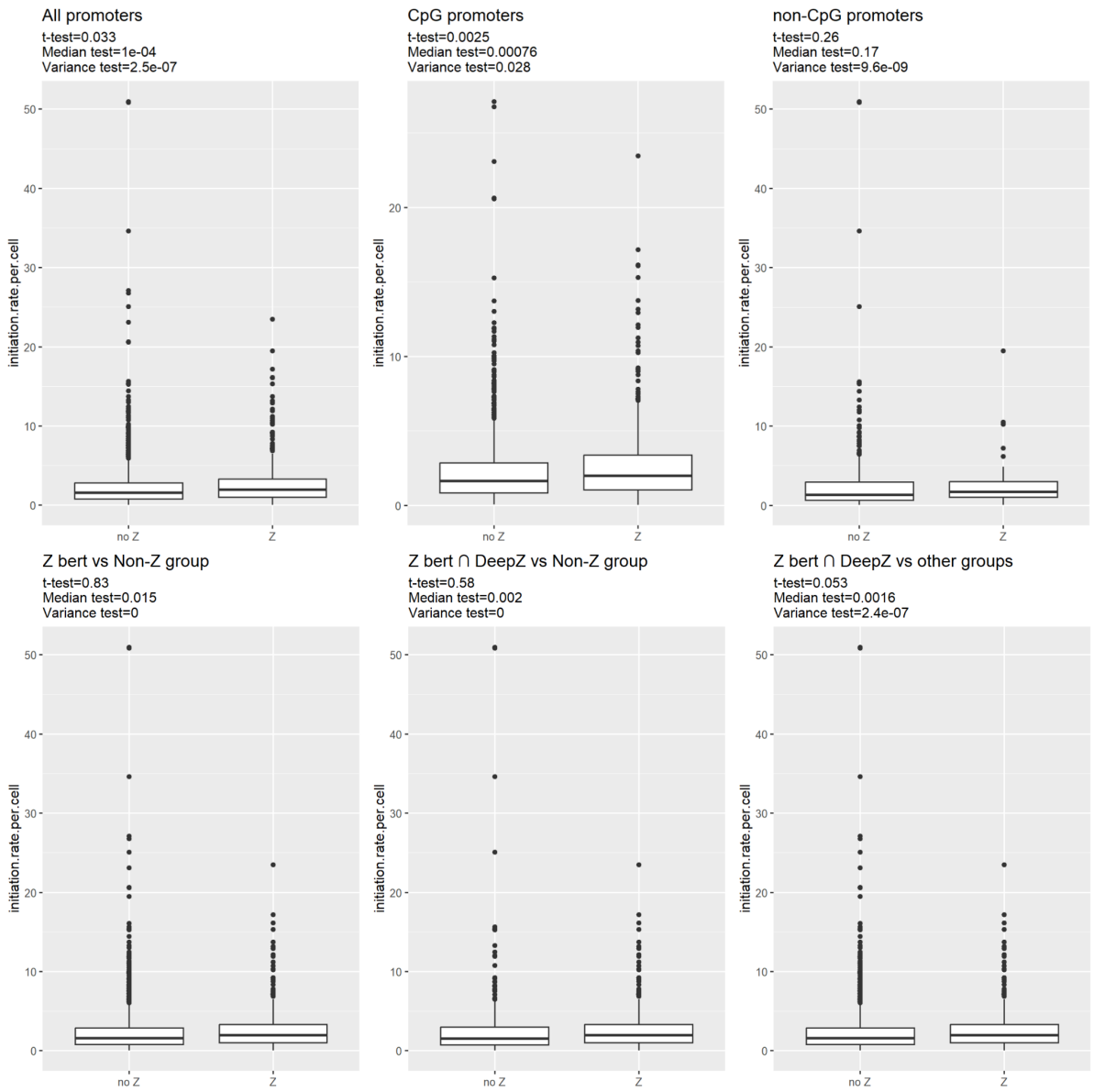
Supplementary Figure 4. Z-flipons in WNT5A gene from Presynapse assembly pathway that was enriched in common cluster of conserved human and mouse Z-flipons (see Figure 3 in the main text). A-E. Regions of Z-flipons detected by three methods – DeepZ, KEx, Z-DNABERT. **A.** Conserved Z-flipons near splice sites of the second exon in WNT5A in human. **B.** Z-flipons composed from GC-repeats in 5'UTR region in mouse. **C**. Z-flipons at alternative promoters of WNT5A in human. **D**. Z-flipons near splice-site after the first exon of Wnt5a in mouse. **E**. Z-flipon detected by three methods – DeepZ, KEx, Z-DNABERT in the exon of Wnt5a in mouse. **F-G**. Region of WNT5A gene with signals from omics features enriched in Z-flipons both in human (**F**) and mouse (**G**) genomes. Omics features are aggregated signals from all tissues as they were used in DeepZ model.


TMEM51-AS1
A
GT-repeats
Refseq
CpG Isl
Exp Z-DNA
DeepZ
Z-DNABERT
B
GT-repeats
Refseq
CpG Isl
Exp Z-DNA
DeepZ
Z-DNABERT
C
Refseq
CpG Isl
Exp Z-DNA
DeepZ
Z-DNABERT
RBFOX2
bidirectional transcription
D
Refseq
CpG Isl
Exp Z-DNA
DeepZ
Z-DNABERT
Rbfox2
bidirectional transcription
E
hg19 chr1:15,473,840-15,486,748
all tissues
RefSeq
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
MYC
MAX
BRD4
ESR1
EP300
RUNX1
ERG
FOXA1
SMARCA4
KDM2B
CREB1
E2F1
TRIM28
BCOR
RAD21
SMC1A
EZH2
KMT2A
MED1
ATF3
RELA
FOXA2
CHD8
TFAP2C
TBP
NR3C1
MBD3
NELFA
SPI1
CTCF
KDM1A
RNF2
NIPBL
RXRA
MYOD1
PPARG
RB1
SUZ12
POU5F1
NANOG
H3K9ac
H2A.Z
H3K18ac
H3K9K14ac
H3.3
H3K9me3
H3K56ac
H4K8ac
H3K122ac
H3K36me3
H3K14ac
H3K36ac
H4K12ac
H4K16ac
H4ac
H4K5ac
H3K79me2
H2AK119Ub
H3
F
mm9 chr4:141,635,890-141,644,323
all tissues
RefSeq
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
Rbfox2
Myc
Max
Brd4
Esr1
Ep300
Runx1
Rrg
Foxa1
Smarca4
Kdm2b
Creb1
E2f1
Trim28
Bcor
Rad21
Smc1a
Ezh2
Kmt2a
Med1
Atf3
Rela
Foxa2
Chd8
Tfap2c
Tbp
Nr3c1
Mbd3
Nelfa
Spi1
Ctcf
Kdm1a
Rnf2
Nipbl
Rxra
Myod1
Pparg
Rb1
Suz12
Pou5f1
Nanog
H3K9ac
H2A.Z
H3K18ac
H3K9K14ac
H3.3
H3K9me3
H3K56ac
H4K8ac
H3K122ac
H3K36me3
H3K14ac
H3K36ac
H4K12ac
H4K16ac
H4ac
H4K5ac
H3K79me2
H2AK119Ub
H3
G
RefSeq
Z-DNA exp
DeepZ
Z-DNABERT
human blood
CTCF
ERG
MAX
RAD21
SMARCA4
SPI1
MYC
ETS1
EP300
SUZ12
H3ac
H3K4me2
H3K4me1
H3K27me3
H3K27ac
H3K4me3
H3K9me2
H3K36me3
H3K9me3
H
RefSeq
Z-DNA exp
DeepZ
Z-DNABERT
mouse blood
Ctcf
Erg
Max
Rad21
Smarca4
Spi1
Myc
Ets1
Ep300
Suz12
H3ac
H3K4me2
H3K4me1
H3K27me3
H3K27ac
H3K4me3
H3K9me2
H3K36me3
H3K9me3

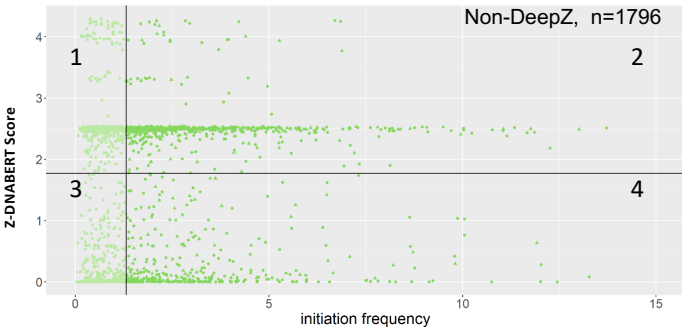
Supplementary Figure 5. Z-flipons at bidirectional and alternative promoters of TMEM51-AS1 in human and mouse genomes. **A-D.** Regions of Z-flipons detected by three methods – DeepZ, KEx, Z-DNABERT in human and mouse genomes. **E-F**. Omics features are aggregated signals from all tissues as they were used in DeepZ model. **G-H**. Selected omics features for human (**G**) and mouse (**H**) genome for blood tissue type.


A
hg19 chr17:41,269,932-41,286,752
all tissues
BRCA1-NBR2
B
mm9 chr11:101,406,150-101,422,355
all tissues
RefSeq
CpG islands
Z-DNA exp
DeepZ
Z-DNABERT
RBFOX2
MYC
MAX
BRD4
ESR1
EP300
RUNX1
ERG
FOXA1
SMARCA4
KDM2B
CREB1
E2F1
TRIM28
BCOR
RAD21
SMC1A
EZH2
KMT2A
MED1
ATF3
RELA
FOXA2
CHD8
TFAP2C
TBP
NR3C1
MBD3
NELFA
SPI1
CTCF
KDM1A
RNF2
NIPBL
RXRA
MYOD1
PPARG
RB1
SUZ12
POU5F1
NANOG
H3K9ac
H2A.Z
H3K18ac
H3K9K14ac
H3.3
H3K9me3
H3K56ac
H4K8ac
H3K122ac
H3K36me3
H3K14ac
H3K36ac
H4K12ac
H4K16ac
H4ac
H4K5ac
H3K79me2
H2AK119Ub
H3
C
liver
Z-DNA exp
DeepZ
Z-DNABERT
DNASE
CTCF
ARID3A
ATF3
CEBPA
CREB1
EP300
GABPA
GRP20
HDAC2
HET1
HNF4A
HSF1
JUND
MAX
MXI1
MYC
RFX5
SIN3A
SP1
SREBF1
SRF
STAG1
TBP
YY1
H3K4me1
H3K4me2
H3K4me3
H3K9ac
H3K27ac
H3K27me3
H3K79me2
H3K36me3
D
liver
Z-DNA exp
DeepZ
Z-DNABERT
DNASE
IKZF1
CLOCK
FOXA2
CREB1
BCL6
CEBPA
CEBPB
THRB
RARA
RARB
EP300
RXRA
NCOR2
NR1H2
NR3C1
ARNTL
HDAC3
PPARA
H2AZ
H3K4me3
H3K9ac
H3K27ac
H3K36me3
H3K4me2
H3K4me1

Supplementary Figure 6. Z-flipons at bidirectional and alternative promoters of BRCA1-NBR2 in human and mouse genomes. **A-B.** Region around bidirectional promoter of BRCA1-NBR2 and signals from omics features enriched in Z-flipons both in human (**A**) and mouse (**B**) genomes. Omics features are aggregated signals from all tissues as they were used in DeepZ model. **C-D**. Selected omics features for human (**C**) and mouse (**D**) genome for liver tissue type.



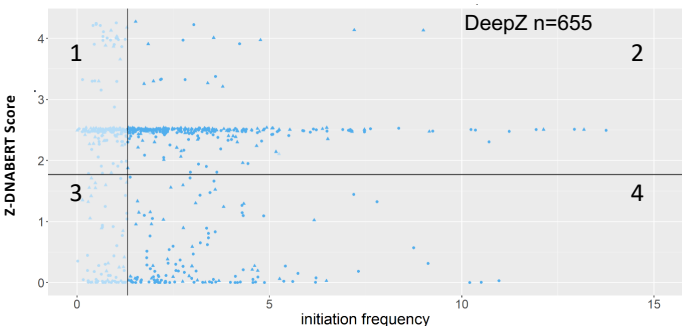
Supplementary Figure 7. Transcription initiation rate for different types of promoters with or without Z-flipons. The groups are exactly those as in the main Figure 8. The comparison of groups is made based on t-test, median-test and variance-test.



Reinitiation Rate

Z-DNABERT	low	high
high	25%	36%
low	19%	20%

Graph Region (percentages)		
	1	2
	3	4



Reinitiation Rate

Z-DNABERT	low	high
high	24%	46%
low	10%	20%

Z-score Test

	Non-DeepZ	DeepZ
proportion (ratio region 4:2)	0.56	0.64

Z= 5.13, one-sided p <0.00001

Supplementary Figure 8. Z-DNABERT prediction scores plotted against initiation rates for promoters with Z-flipons (blue) and without Z-flipons (green). The graph is divided into four quadrants: quadrant 2 corresponds to promoters with both high Z-scores and reinitiation rates; quadrant 1 corresponds to promoters with low initiation rate; quadrant 4 corresponds to promoters with to low Z-Scores; quadrant 3 corresponds to promoters with both low Z-scores and reinitiation rates.