Supplementary Figure Legends

Figure S1. Schematic of the modified cytosine PBM methodology. Single-stranded Agilent HK arrays containing 40,000 features were made double-stranded with a nucleotide mix that contains either cytosine, 5mC, or 5hmC using a common primer. An epitope-tagged (GST) TF is allowed to bind to the sequences on the microarray slide. Binding is then detected using a fluor-conjugated antibody to the followed by measurement of bound protein (fluorescence intensities) for each of the array features.

Figure S2. Comparison of CREB1-GST binding to DNA 8-mers between replicates. A. Z-scores for CREB1 binding to 8-mers containing cytosine from replicate1 (Y-axis) and cytosine from replicate 2 (X-axis). Three spots are highlighted: 1.the canonical CRE 8-mer TGAC|GTCA (red spot), 2.the CRE|C/EBP chimeric 8-mer TGAC|GCAA (green spot), and the 8-mer TGAT|GCAA (blue spot). **B.** Z-scores for CREB1binding to 8-mers containing 5mC from replicate1 (Y-axis) and 5mC from replicate 2 (X-axis). **C.** Z-scores for CREB1binding to 8-mers containing 5hmC from replicate 1 (Y-axis) and 5hmC from replicate 2 (X-axis).

Figure S3. Comparison of ATF1-GST binding to DNA 8-mers between replicates. A. Z-scores for ATF1 binding to 8-mers containing cytosine from replicate 1 (Y-axis) and cytosine from replicate 2 (X-axis). **B.** Z-scores for ATF1 binding to 8-mers containing 5mC from replicate1 (Y-axis) and 5mC from replicate 2 (X-axis). **C.** Z-scores for ATF1 binding to 8-mers containing 5hmC from replicate 1 (Y-axis) and 5hmC from replicate 2 (X-axis).

Figure S4. Comparison of JUND-GST binding to DNA 8-mers between replicates. Z-scores for JUND binding to 8-mers containing cytosine from replicate 1 (Y-axis) and cytosine from replicate 2 (X-axis).

Figure S5. Heat-map of enrichment of six motifs in DHS derived from 125 human cells. Hierarchical clustering of enrichment scores for six motifs (CRE, C/EBP, CRE|C/EBP, TGAT|GCAA, and two negative control sequences (-)) in DHS derived from 125 samples in human. Samples were classified according to their karyotype (cancer, normal, brain, ESC or B- or T-cells) as indicated by the color legend.

Figure S6. Enrichment of GO biological process terms of promoters containing TGAT|GCAA. Promoters are defined as -10kbp to +10kbp from the transcription start site.

Figure S1

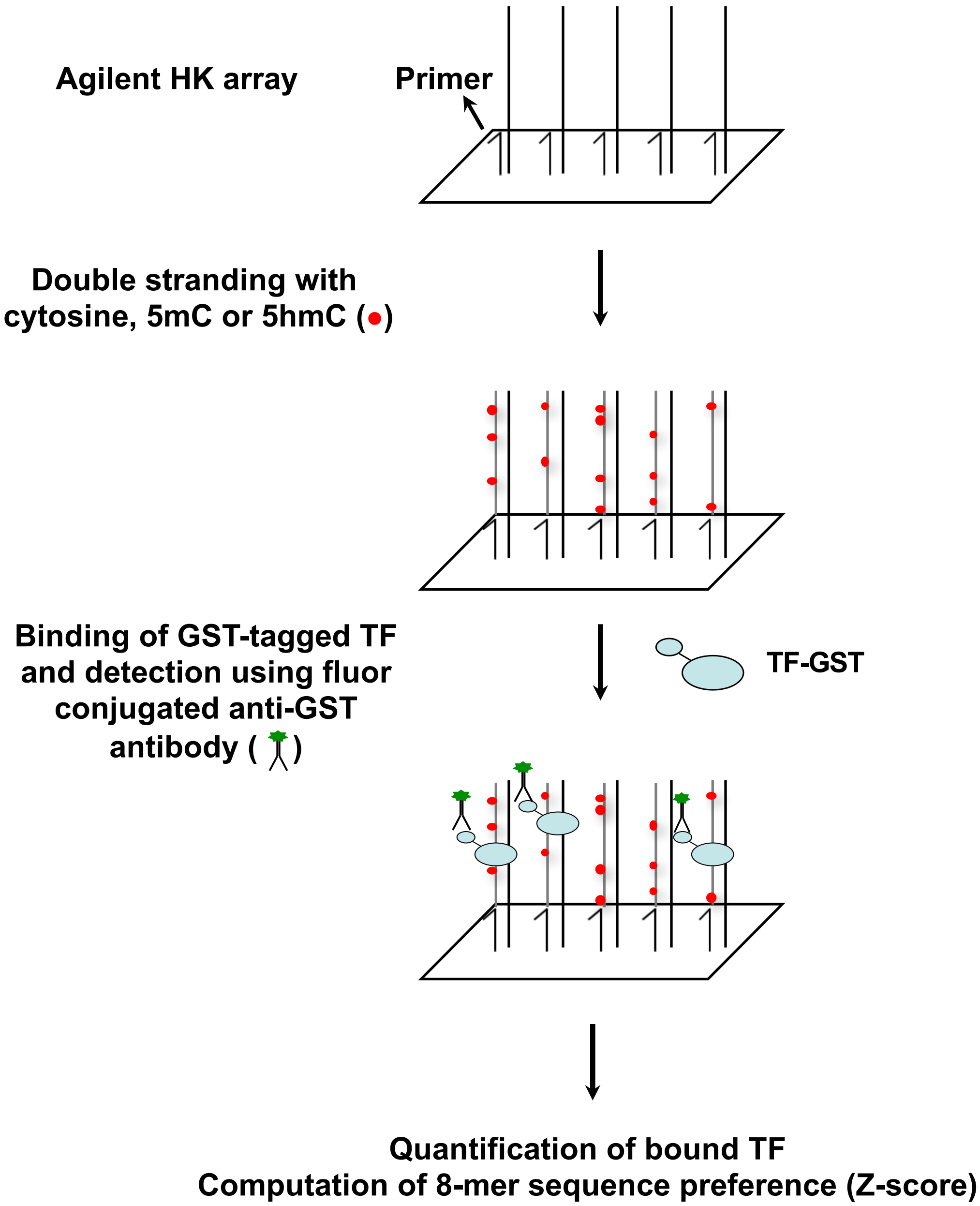

Agilent HK array
Primer
Double stranding with cytosine, 5mC or 5hmC (●)
Binding of GST-tagged TF and detection using fluor conjugated anti-GST antibody ( )
TF-GST
Quantification of bound TF
Computation of 8-mer sequence preference (Z-score)

Figure S2

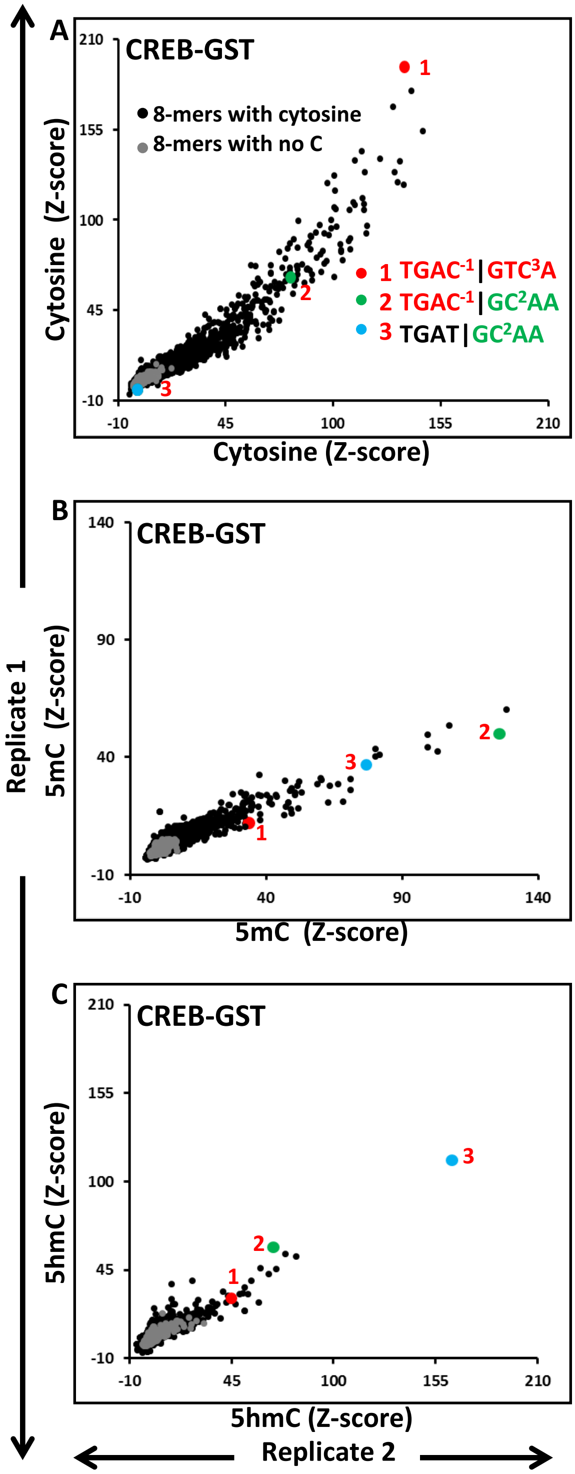

A
CREB-GST
8-mers with cytosine
8-mers with no C
1 TGAC^{-1}|GTC^{3}A
2 TGAC^{-1}|GC^{2}AA
3 TGAT|GC^{2}AA
Cytosine (Z-score)
Cytosine (Z-score)
-10
45
100
155
210
B
CREB-GST
5mC (Z-score)
5mC (Z-score)
-10
40
90
140
C
CREB-GST
5hmC (Z-score)
5hmC (Z-score)
-10
45
100
155
210
1
2
3
Replicate 1
Replicate 2

Figure S3

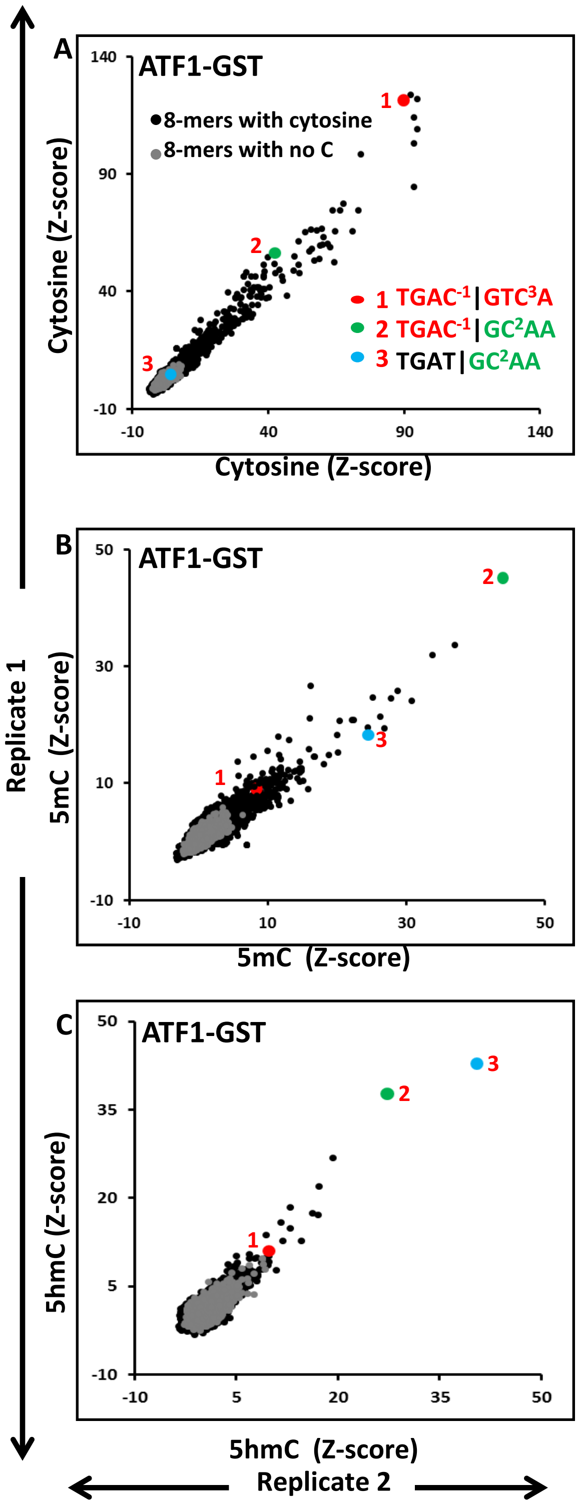

A
ATF1-GST
8-mers with cytosine
8-mers with no C
1 TGAC^{-1}|GTC^{3}A
2 TGAC^{-1}|GC^{2}AA
3 TGAT|GC^{2}AA
1
2
3
140
90
40
-10
-10
40
90
140
Cytosine (Z-score)
Cytosine (Z-score)
B
ATF1-GST
50
30
10
-10
-10
10
30
50
5mC (Z-score)
5mC (Z-score)
C
ATF1-GST
50
35
20
5
-10
-10
5
20
35
50
5hmC (Z-score)
5hmC (Z-score)
Replicate 1
Replicate 2

Figure S4

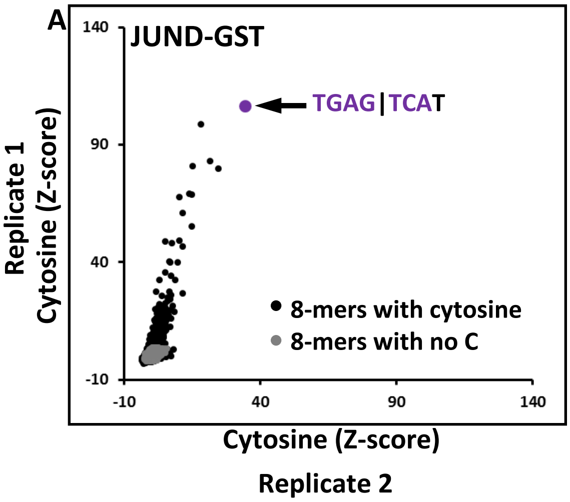

A
JUND-GST
TGAG|TCAT
8-mers with cytosine
8-mers with no C
140
90
40
-10
-10
40
90
140
Replicate 1
Cytosine (Z-score)
Cytosine (Z-score)
Replicate 2

Figure S5

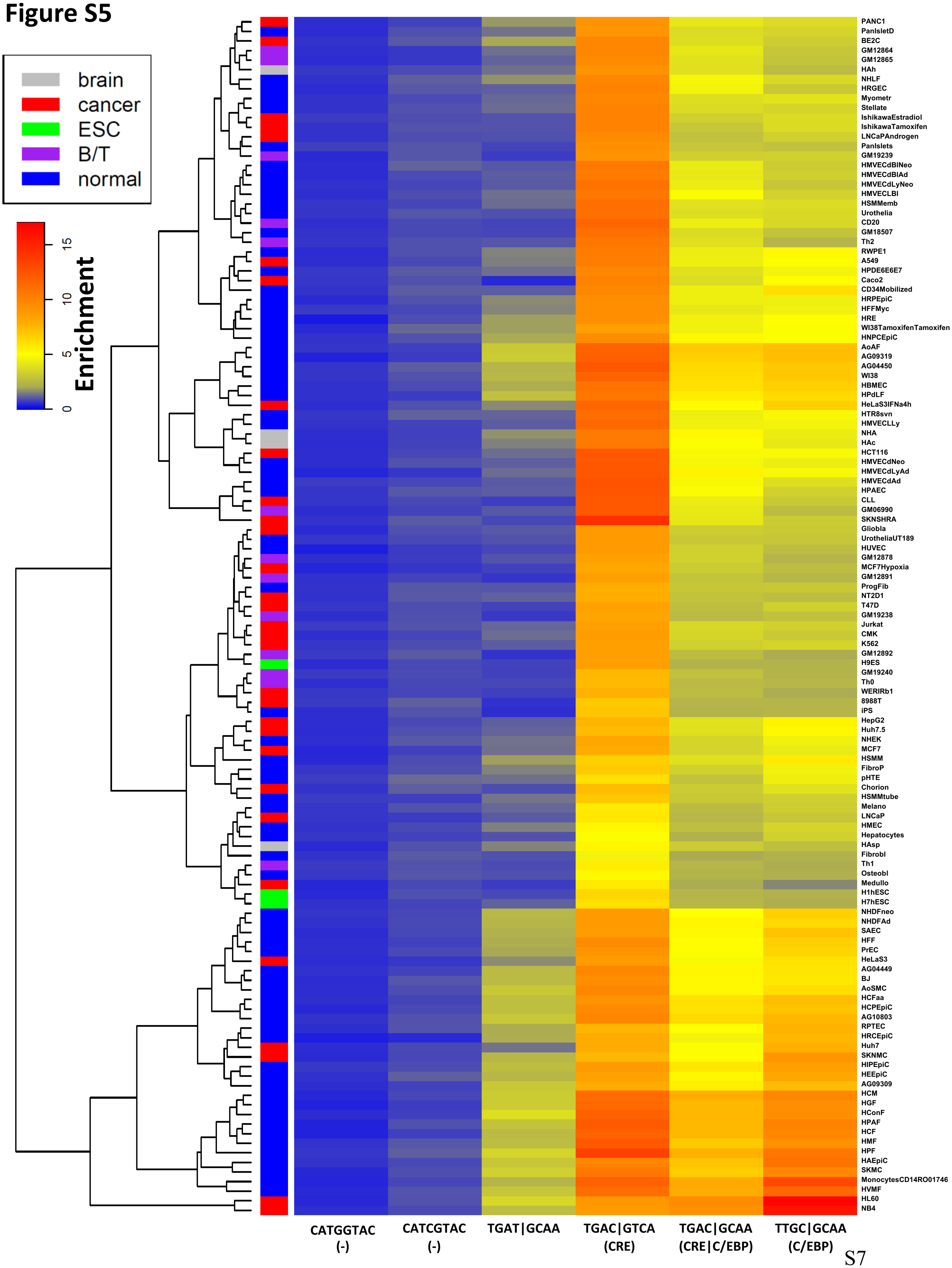
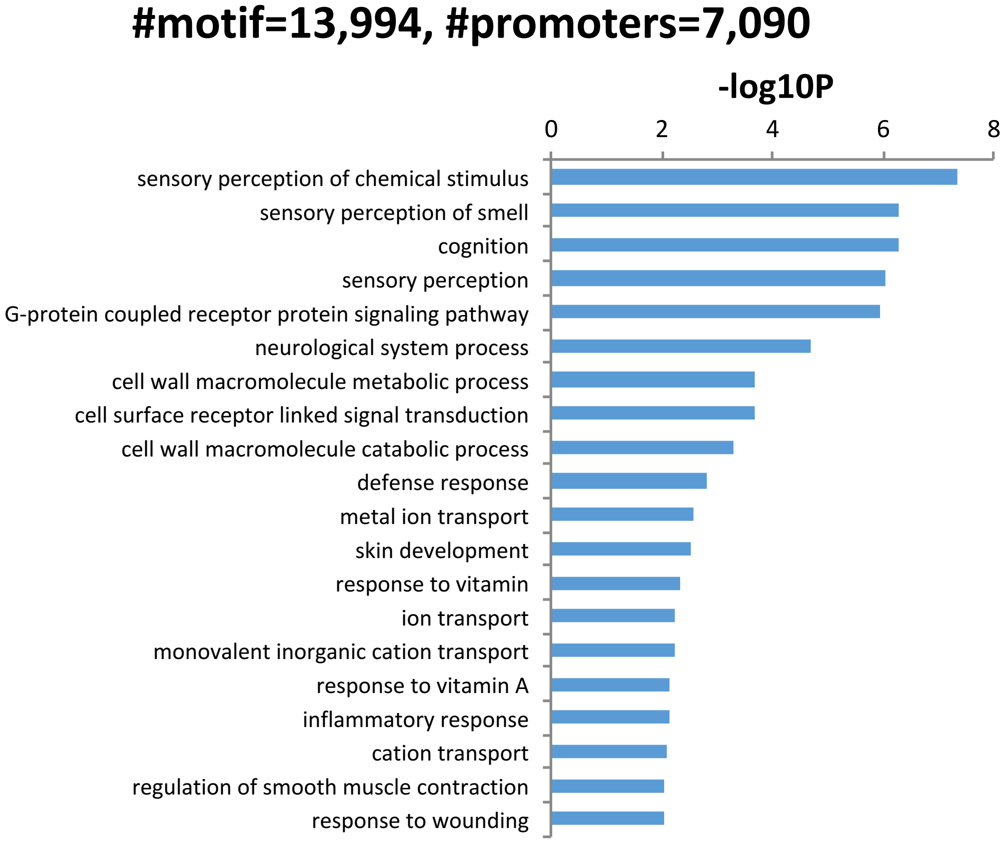

brain
cancer
ESC
B/T
normal
0
5
10
15
Enrichment
PANC1
PanIsletD
BE2C
GM12864
GM12865
HAh
NHLF
HRGEC
Myometr
Stellate
IshikawaEstradiol
IshikawaTamoxifen
LNCaPAndrogen
PanIslets
GM19239
HMVECdBlNeo
HMVECdBlAd
HMVECdLyNeo
HMVECLBl
HSMMemb
Urothelia
CD20
GM18507
Th2
RWPE1
A549
HPDE6E6E7
Caco2
CD34Mobilized
HRPEpiC
HFFMyc
HRE
WI38TamoxifenTamoxifen
HNPCEpiC
AoAF
AG09319
AG04450
WI38
HBMEC
HPdLF
HeLaS3IFNa4h
HTR8svn
HMVECLLy
NHA
HAc
HCT116
HMVECdNeo
HMVECdLyAd
HMVECdAd
HPAEC
CLL
GM06990
SKNSHRA
Gliobla
UrotheliaUT189
HUVEC
GM12878
MCF7Hypoxia
GM12891
ProgFib
NT2D1
T47D
GM19238
Jurkat
CMK
K562
GM12892
H9ES
GM19240
Th0
WERIRb1
8988T
iPS
HepG2
Huh7.5
NHEK
MCF7
HSMM
FibroP
pHTE
Chorion
HSMMtube
Melano
LNCaP
HMEC
Hepatocytes
HAsp
Fibrobl
Th1
Osteobl
Medullo
H1hESC
H7hESC
NHDFneo
NHDFAd
SAEC
HFF
PrEC
HeLaS3
AG04449
BJ
AoSMC
HCFaa
HCPEpiC
AG10803
RPTEC
HRCEpiC
Huh7
SKNMC
HIPEpiC
HEEpiC
AG09309
HCM
HGF
HConF
HPAF
HCF
HMF
HPF
HAEpiC
SKMC
MonocytesCD14RO01746
HVMF
HL60
NB4
CATGGTAC (-)
CATCGTAC (-)
TGAT|GCAA
TGAC|GTCA (CRE)
TGAC|GCAA (CRE|C/EBP)
TTGC|GCAA (C/EBP)

Figure S6


#motif=13,994, #promoters=7,090
-log10P
0
2
4
6
8
sensory perception of chemical stimulus
sensory perception of smell
cognition
sensory perception
G-protein coupled receptor protein signaling pathway
neurological system process
cell wall macromolecule metabolic process
cell surface receptor linked signal transduction
cell wall macromolecule catabolic process
defense response
metal ion transport
skin development
response to vitamin
ion transport
monovalent inorganic cation transport
response to vitamin A
inflammatory response
cation transport
regulation of smooth muscle contraction
response to wounding