

Supplemental information for

Genome-wide identification of protein binding sites in mammalian cells

Fenglin Liu¹, Tianyu Ma¹, Yuxiang Zhang¹

Supplemental Items:

Supplemental Fig 1. Reaction for conjugating beads and proteins with Sulfo-SMCC

Supplemental Fig 2. Chromosomal distribution of PBSs

Supplemental Fig 3. Confirmation of PBSs from different groups by PCR

Supplemental Fig 4. GO molecular function terms of proteins in different salt fractions

Supplemental Fig 5. Percentages of PBSs associated with different gene regions

Supplemental Fig 6. Verification of predictive CTCF-TFBSs by ChIP-exo

Supplemental Fig 7. The families and numbers of TFBSs on 23 chromosomes

Supplemental Fig 8. The GC content and length of PBSs affect their binding with TFs

Supplemental Fig 9. Distribution of PBSs in the promoters of the top 50 highly expressed genes and 50 non-expressed genes

Supplemental Table 1. PBS-peaks of the six NaCl groups (see separate excel file)

Supplemental Table 2. Sequences of the six chosen PBSs

Supplemental Table 3. Protein mass spectra of the five groups (see separate excel file)

Supplemental Table 4. Gene expression levels (FPKM) of Hela cells (see separate excel file)

Supplemental Table 5. Predictions for the five groups of TFBSs (see separate excel file)

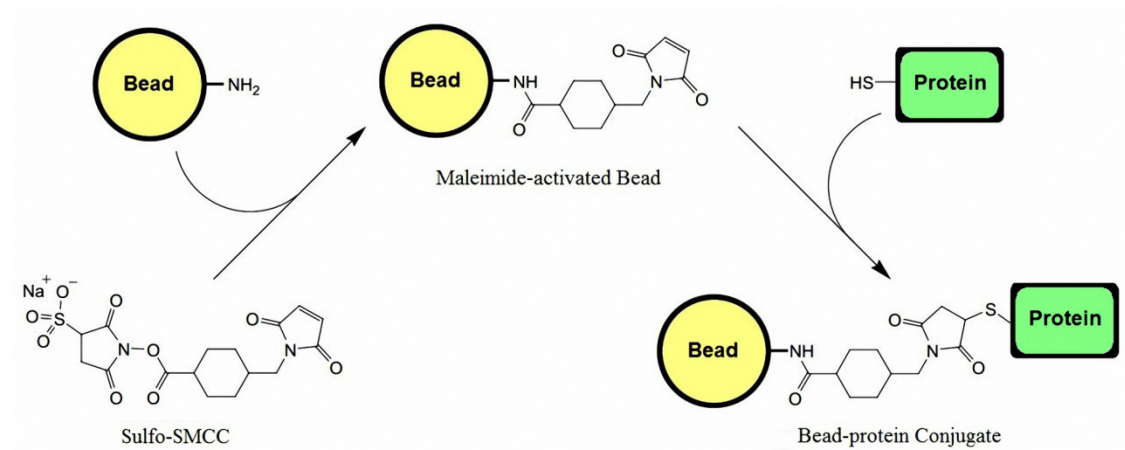
Supplemental Table 6. Chromosomal locations of the chosen NFYB-TFBSs

Supplemental Table 7. Chromosomal locations of the chosen SP1-TFBSs

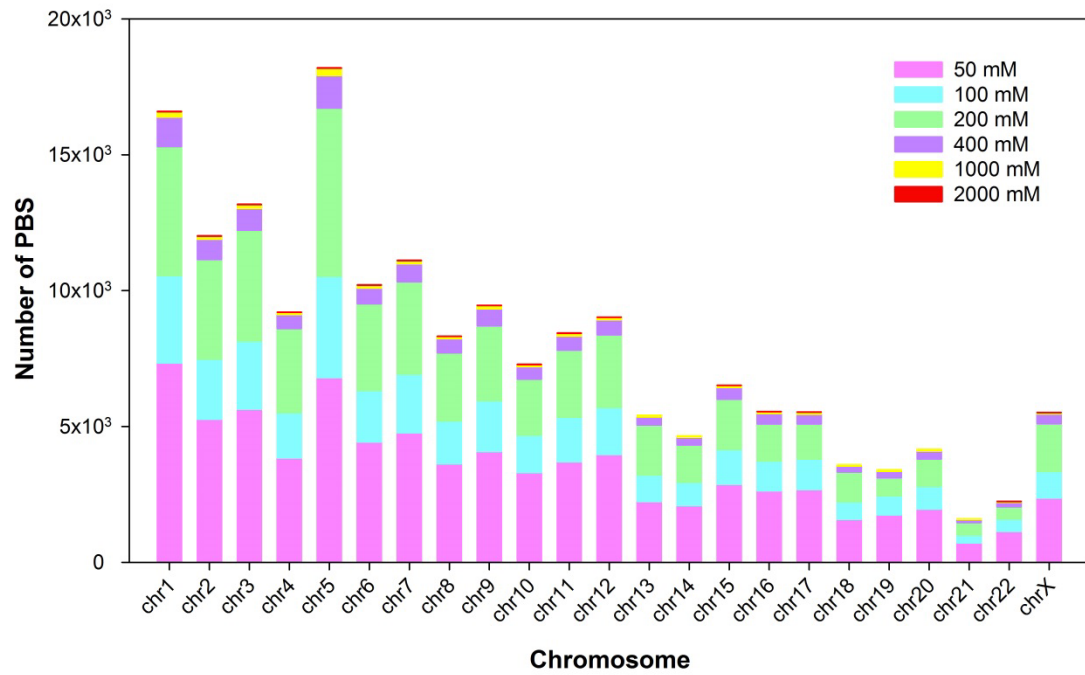
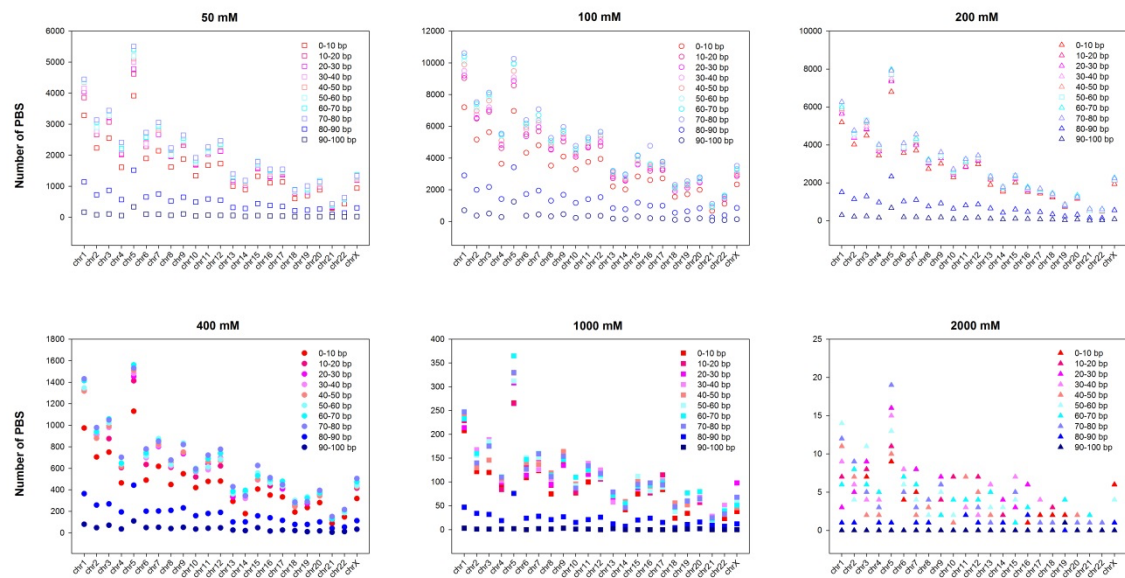
Supplemental Table 8. Chromosomal locations of the chosen EGR1-ETS1-TFBSs

Supplemental Table 9. Sequences of the chosen SP1-TFBSs and PCR primers used

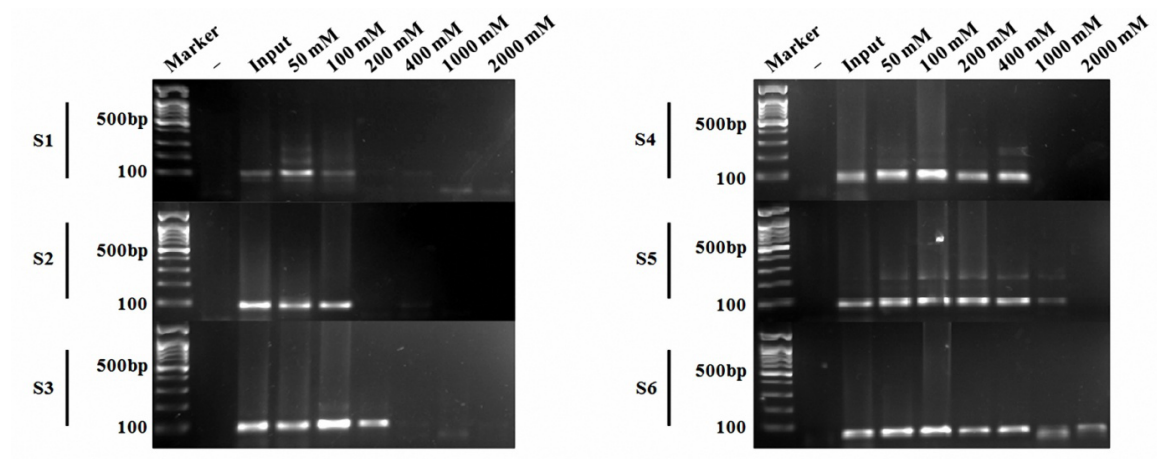
Supplemental Table 10. FPKM data for genes related to promoter regions in the six groups (see separate excel file)



Supplemental Figure 1. Reaction for conjugating beads and proteins with Sulfo-SMCC. The cross-linker is first reacted with the amine groups on the beads to produce maleimide groups, and then these are reacted with sulfhydryl groups on the chromatin proteins.

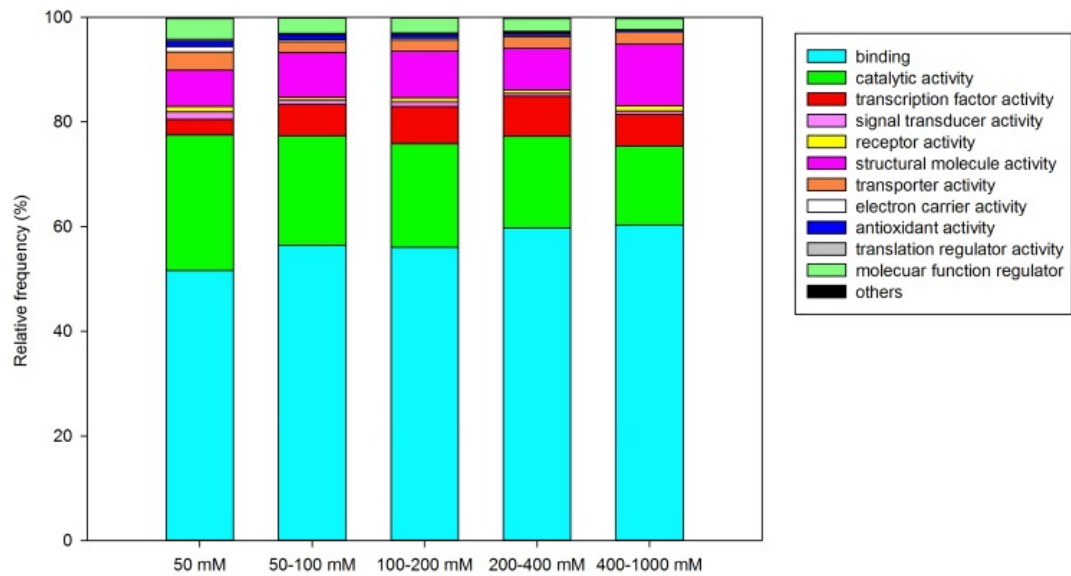
A**B**

Supplemental Figure 2. Chromosomal distribution of PBSs. (A) Overall numbers of PBS peaks on different chromosomes. **(B)** Numbers of PBS peaks of different lengths on different chromosomes.

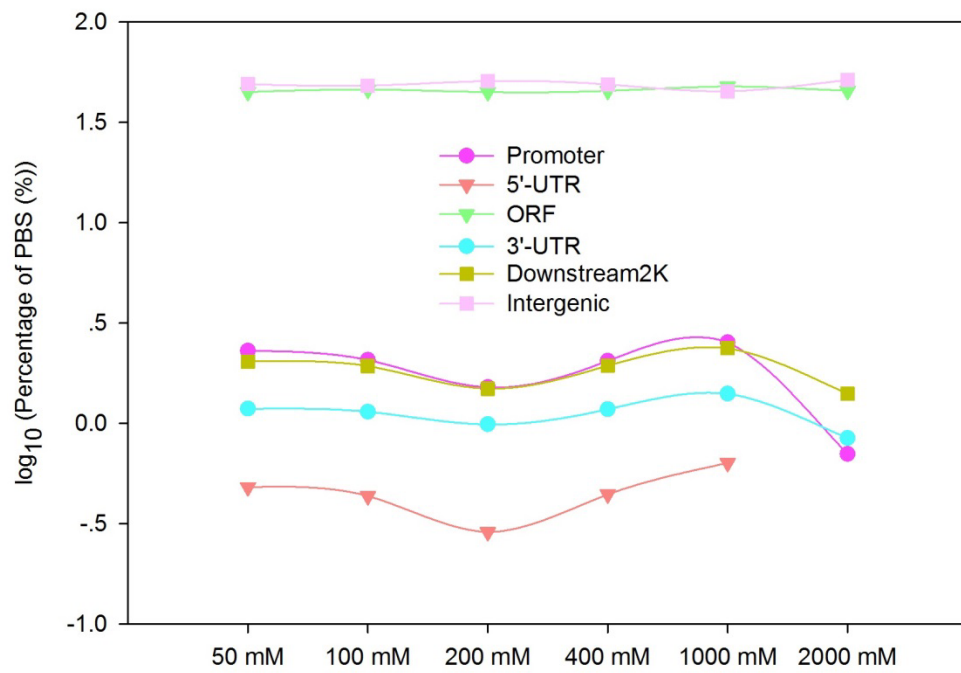


Supplemental Figure 3. Confirmation of PBSs from different groups by PCR.

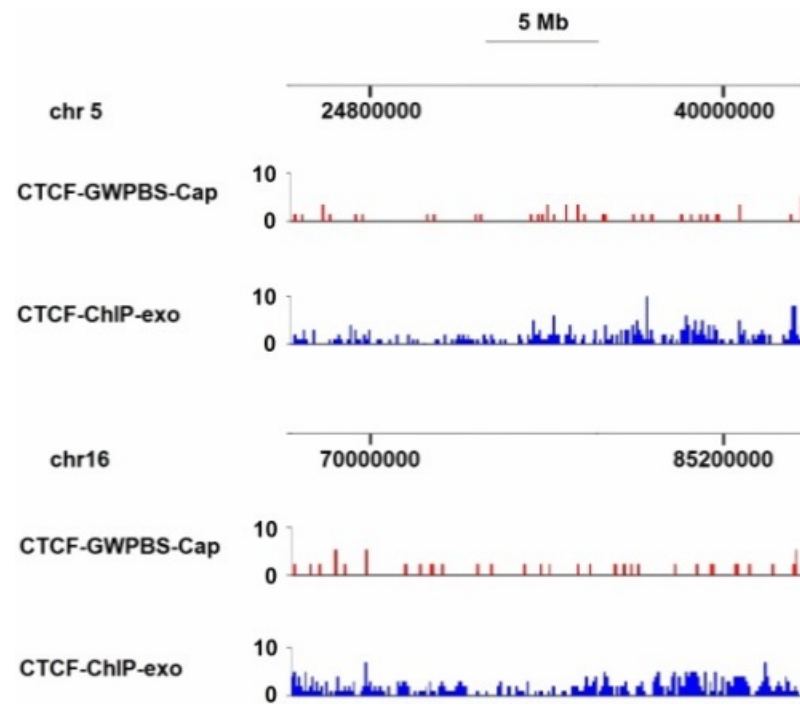
PCR primers were designed for PBSs from the different NaCl washing groups (S1-S6). The GoldView-stained agarose gels show the PCR products. The first two lanes ("- " and "Input") represent negative and positive controls, respectively.



Supplemental Figure 4. GO molecular function terms of proteins in different salt fractions. “50 mM” means that the nuclear proteins were extracted with 50 mM salt. “50-100 mM” means that the nuclei were washed with 50 mM salt first and then 100 mM salt was used to extract the nuclear proteins. “100-200 mM” means that the nuclei were washed with 100 mM salt first and then 200 mM salt was used to extract the nuclear proteins. “200-400 mM” means that the nuclei were washed with 200 mM salt first and then 400 mM salt was used to extract the nuclear proteins. “400-1000 mM” means that the nuclei were washed with 400 mM salt first and then 1000 mM salt was used to extract the nuclear proteins.

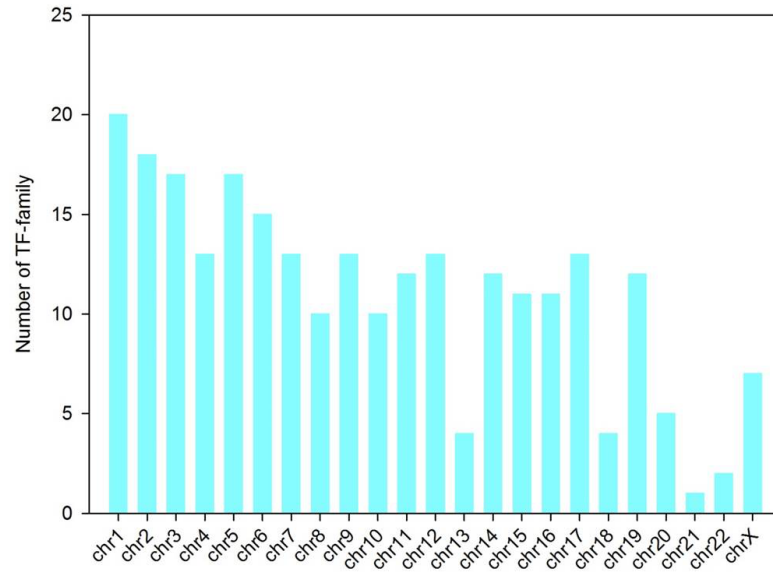
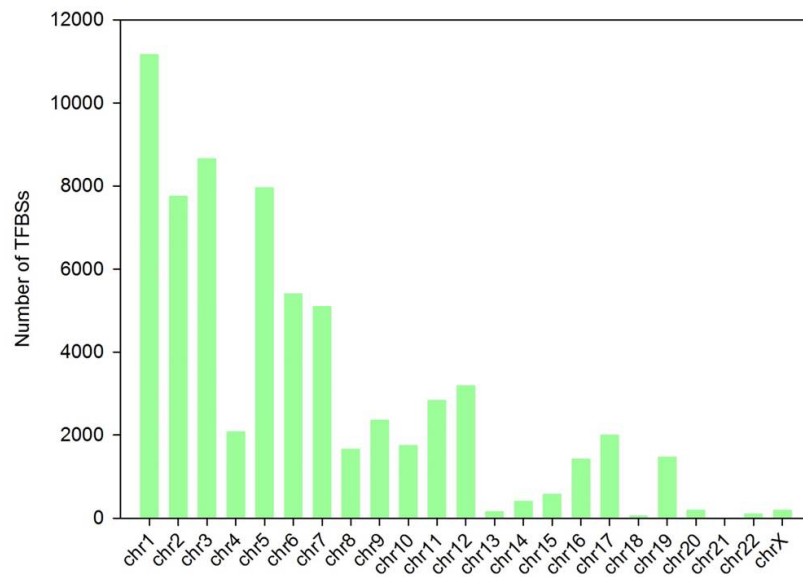


Supplemental Figure 5. Percentages of PBSs associated with different gene regions.



Supplemental Figure 6. Verification of predictive CTCF-TFBSs by ChIP-exo.

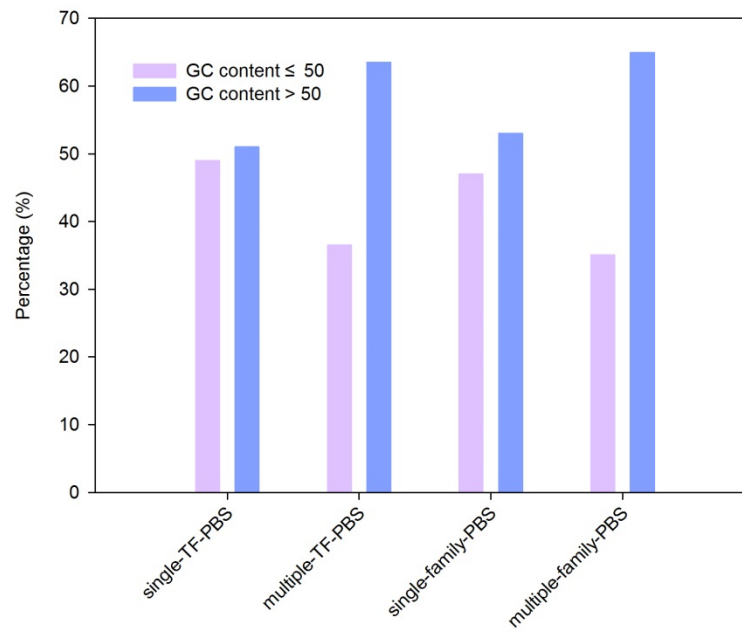
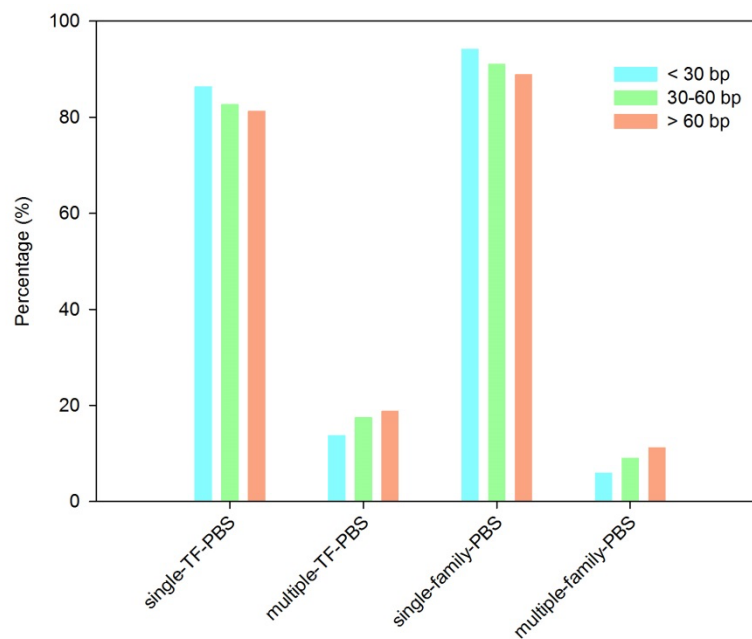
The CTCF-TFBSs predicted by GWPBS-Cap are contained within the CTCF locations identified by ChIP-exo.

A**B**

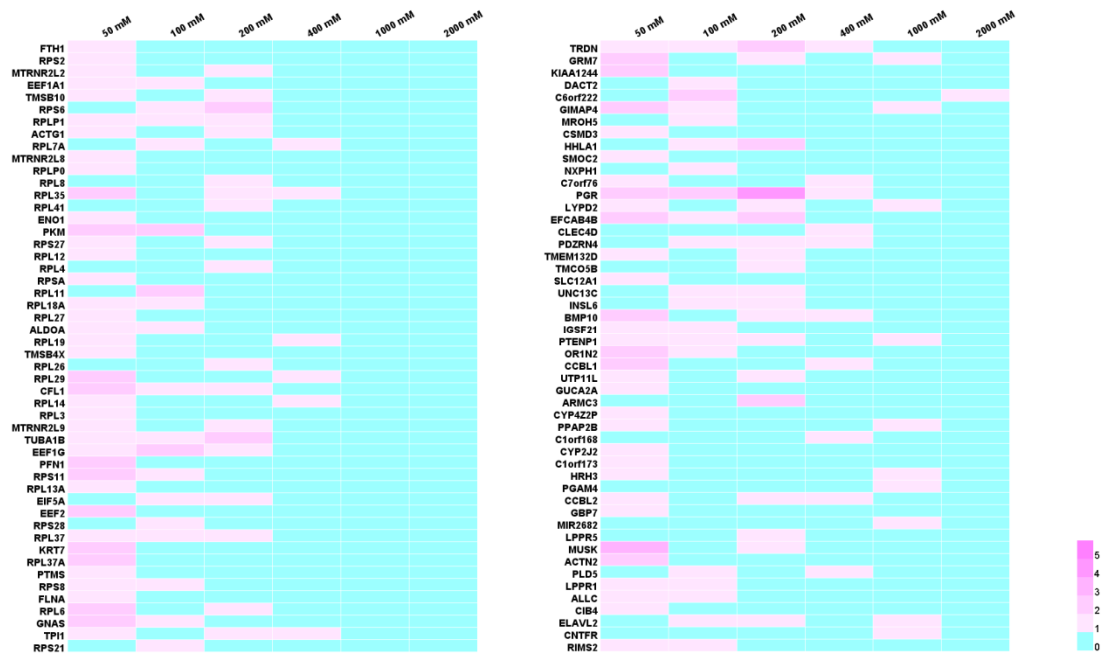
Supplemental Figure 7. The families and numbers of TFBSs on 23 chromosomes.

(A) The distribution of 21 TF families of TFBSs on different chromosomes. **(B)**

Numbers of TFBSs contained in PBSs on different chromosomes.

A**B**

Supplemental Figure 8. The GC content and length of PBSs affect their binding with TFs. (A) PBSs with high GC contents (>50%) tend to bind multiple TFs or TF families rather than single TFs or single TF families. **(B)** The percentages of TF TFBSs for single TFs or single TF families are negatively correlated with the lengths of TFBSs.



Supplemental Figure 9. Distribution of PBSs in the promoters of the top 50 highly expressed genes and 50 non-expressed genes. Left panel shows the distribution of PBSs from the different groups in the promoters of the top 50 highly expressed genes. Right panel shows the distribution of PBSs from the different groups in the promoters of 50 randomly selected non-expressed genes. The PBSs from the lower NaCl PBS groups, but not from the 1000 mM or 2000 mM PBS groups, tend to be enriched in the promoters of highly expressed genes. This is in contrast to the distribution of PBSs in weakly expressed genes, whose promoters tend to contain PBSs from the 1000 mM and 2000 mM NaCl PBS groups.

Supplemental Table 1. PBS-peaks of the six NaCl groups (see separate excel file).

Supplemental Table 2. Sequences of the six chosen PBSs.

Name	Group	Chrom	Start (bp)	End (bp)	Length (bp)
S1	50 mM	chr1	39619838	39619930	94
S2	100 mM	chr11	45970425	45970514	91
S3	200 mM	chr11	27351478	27351569	92
S4	400 mM	chr1	178603106	178603199	94
S5	1000 mM	chr4	74163765	74163848	84
S6	2000 mM	chr11	90420859	90420927	69

Supplemental Table 3. Protein mass spectra of the five groups (see separate excel file).

Supplemental Table 4. Gene expression levels (FPKM) of Hela cells (see separate excel file).

Supplemental Table 5. Predictions for the five groups of TFBSs (see separate excel file).

Supplemental Table 6. Chromosomal locations of the chosen NFYB-TFBSs.

Name	Chrom	Start (bp)	End (bp)
NFYB-S1	chr1	28907625	28907703
NFYB-S2	chr1	114300332	114300387
NFYB-S3	chr1	119682734	119682796
NFYB-S4	chr1	145213228	145213285
NFYB-S5	chr1	205199203	205199267
NFYB-S6	chr1	207226000	207226059
NFYB-S7	chr1	173793051	173793081
NFYB-S8	chr1	200374722	200374783
NFYB-S9	chr1	68697869	68697901
NFYB-S10	chr1	233086593	233086664

Supplemental Table 7. Chromosomal locations of the chosen SP1-TFBSs.

Name	Chrom	Start (bp)	End (bp)
SP1-S1	chr1	29505102	29505171
SP1-S2	chr1	40023614	40023692
SP1-S3	chr1	150274247	150274323
SP1-S4	chr1	230350095	230350165
SP1-S5	chr2	46499055	46499131
SP1-S6	chr2	54132403	54132483
SP1-S7	chr2	70393477	70393557
SP1-S8	chr2	152929383	152929461
SP1-S9	chr3	48025730	48025793
SP1-S10	chr5	414065	414163

Supplemental Table 8. Chromosomal locations of the chosen EGR1-ETS1-TFBSs.

Name	Chrom	Start (bp)	End (bp)
EGR1-ETS1-S1	chr4	177436969	177437046
EGR1-ETS1-S2	chr5	37031312	37031392
EGR1-ETS1-S3	chr5	37281265	37281317
EGR1-ETS1-S4	chr5	77326620	77326664
EGR1-ETS1-S5	chr5	94841229	94841288
EGR1-ETS1-S6	chr5	94831966	94832033
EGR1-ETS1-S7	chr5	112152685	112152728
EGR1-ETS1-S8	chr5	133878845	133878893
EGR1-ETS1-S9	chr5	180090166	180090237
EGR1-ETS1-S10	chr5	180105239	180105295

Supplemental Table 9. Sequences of the chosen SP1-TFBSs and PCR primers used.

	SP1-S_{100 mM}	SP1-S_{1000 mM}
Group	100 mM	1000 mM
Chrom	chr1	chr1
Start (bp)	211949661	24382952
End (bp)	211949745	24383015
Length (bp)	85	64
Forward- Primer	TTGAGGTCAGGAGTTCGAGA	AGGTTCAACGGATTCTCCCA
Reverse- Primer	CACCATGCCTGGCTAATTTT	ATGGTGGCAGGTGCCTGTAAT

Supplemental Table 10. FPKM data for genes related to promoter regions in the six groups (see separate excel file).