Supplementary Fig. S1

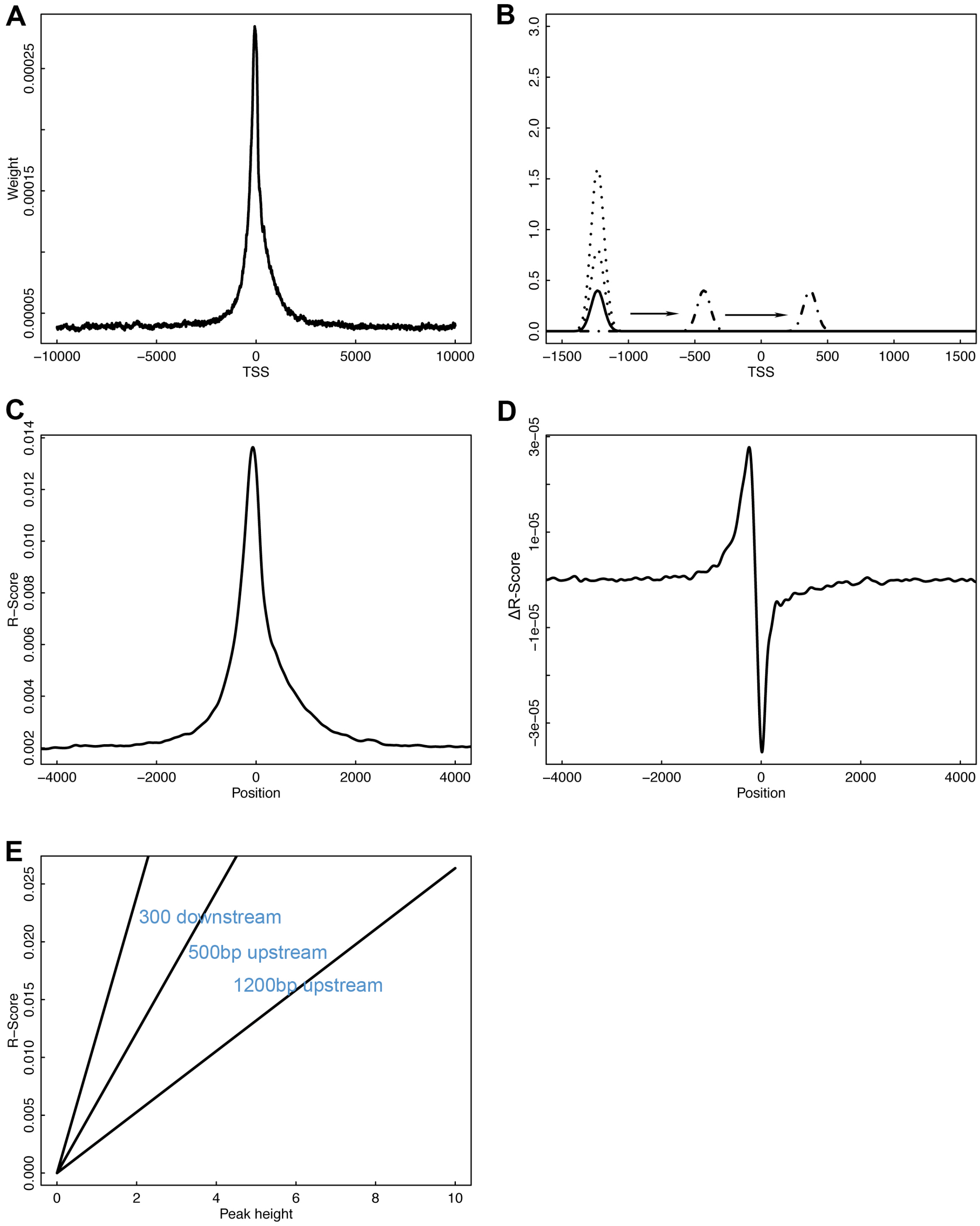

Fig. S1: Effect of peak position and height on regulatory score. (A) The binding profile of STAT4 around TSSs of mouse genes. (B) A simulated binding peak nearby the TSS of a gene. (C) Relationship between regulatory score and peak position. (D) Relationship between the change of regulatory score and peak position. Note the intensive change occurs nearby the TSS. (E) Relationship between regulatory score and peak height. Note the regulatory score increase linearly with the increase of peak height in a position dependent manner.

Supplementary Fig. S2

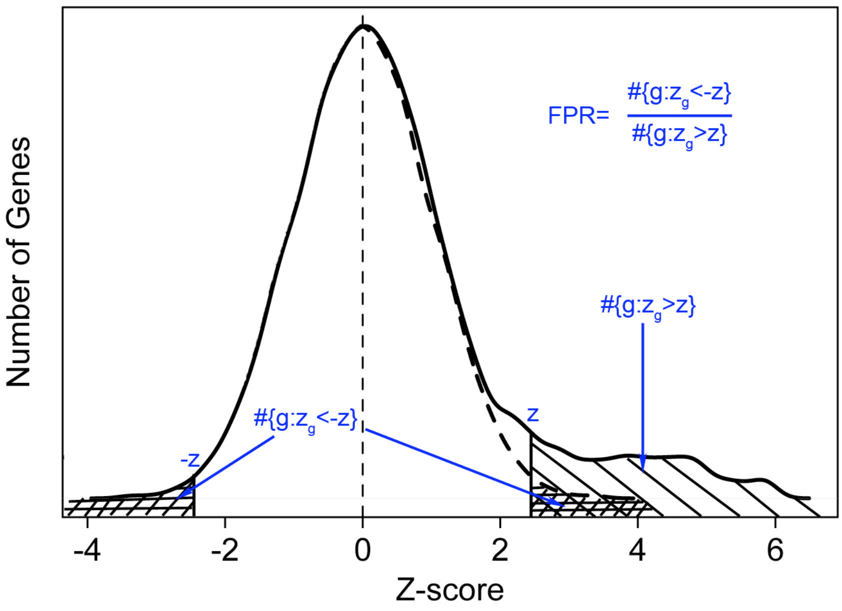

Fig. S2: Schematic diagram for false positive rate (FPR) calculation. Note the heavy tail on the right side correspond to the scores for target genes.

Supplementary Fig. S3

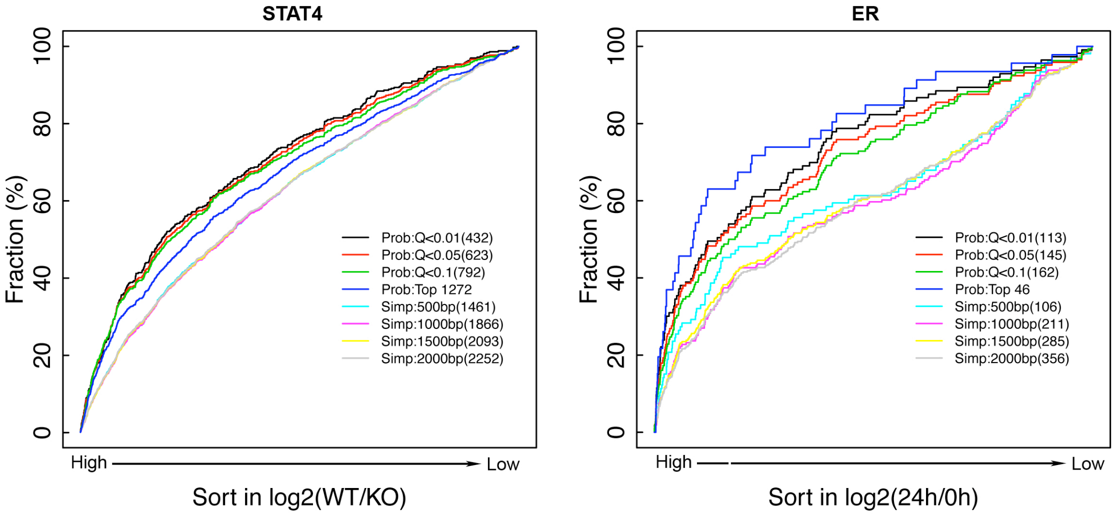

Fig. S3: A large version of Fig. 2C (A) and Fig. 3B (B).

Supplementary Fig. S4

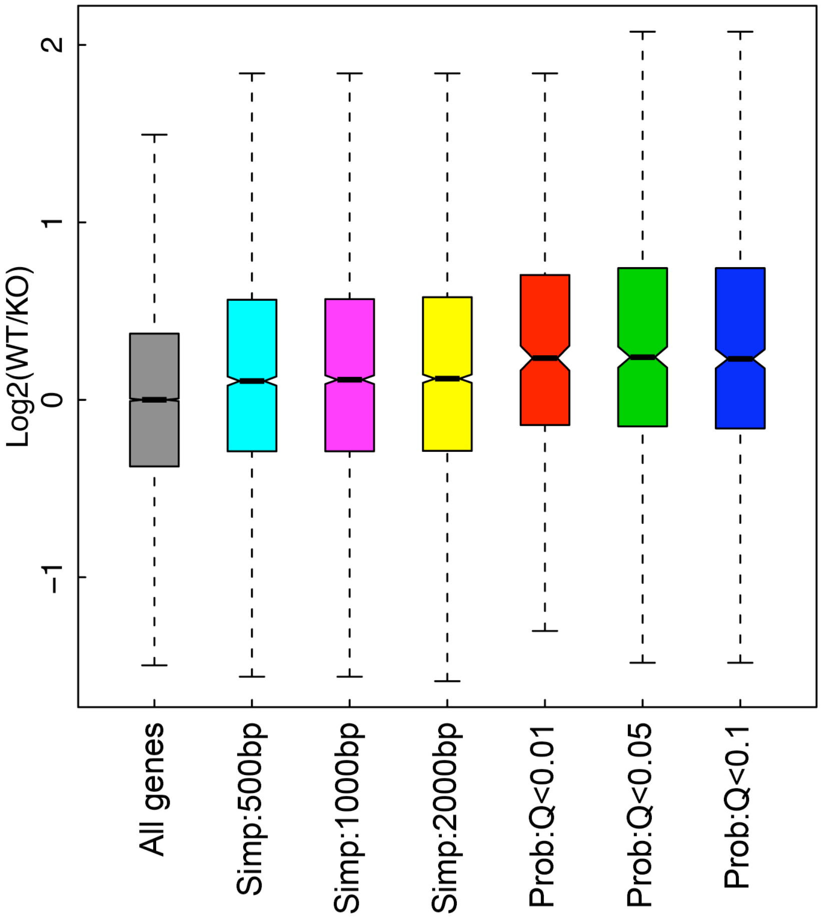

Fig. S4: Expression changes of STAT6 target genes in wild type versus STAT6 knockout Th2 cells. Note the target genes identified by the probabilistic model are more differentially expressed than those by the simple method.

Supplementary Fig. S5

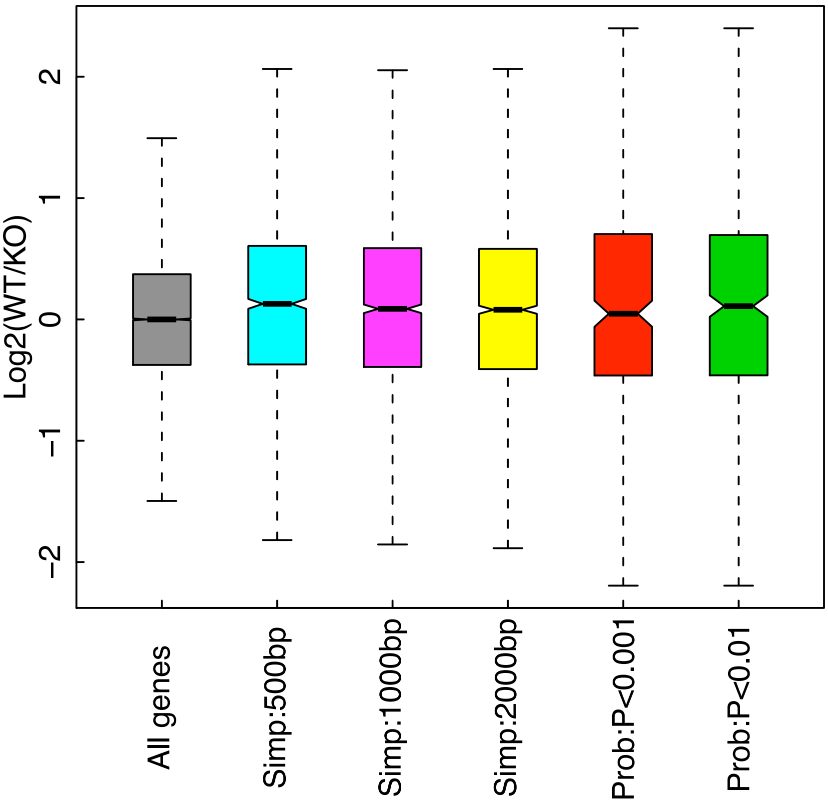

Fig. S5: Expression changes of STAT4 target genes in Th1 cells with WT versus KO of STAT4. Note the median expression changes of target genes identified by applying the probabilistic model on binding peak data are slightly higher than those by the simple method (not significant).

Supplementary Table S6

	Cut-off	**ERα**	**mutERα**
Simple method	[-500bp, 500] bp	349	10
	[-1000, 1000] bp	651	20
	[-1500, 1500] bp	883	31
	[-2000, 2000] bp	1,091	39
Probabilistic model	Q<0.001	244	30
	Q<0.01	312	41
	Q<0.05	406	73
	Q<0.1	492	90

Table S6: The number of human ER target genes identified by the simple method and the probabilistic model in MDA-MB-231 cell lines with wild type (ERα) and mutant ERα (mutERα).

Supplementary Table S7

Number of Mapped reads (M)	Simple Method (Peakseq)		#Target genes by Prob. Model (P<0.01)
	#Peaks (Q-value<0.001)	#Target genes	
5	7716	1458	372
6	9151	1793	380
7	10638	2160	360
8	12193	2609	375
9	13646	2948	381
10	15147	3361	379
11	16691	3727	376
12	18113	4118	372
13	19847	4546	369
14	21340	4859	375
15	22904	5208	376
16	24451	5537	365
17	26135	5757	372
18	27820	6233	363
19	29466	6390	370
20	31199	6740	363
21	32990	6867	389
22	34861	7281	375
23	36620	7468	373
24	38441	7672	357
25	40318	7932	374

Table S7: The number of peak and target genes in 21 simulated datasets for TCF4. Datasets with read depth from 5M to 25M were generated by randomly sampling from ~18M experimental derived mapped reads for TCF4. Input DNA was used as control for peak calling by PeakSeq.

Supplementary Fig. S8

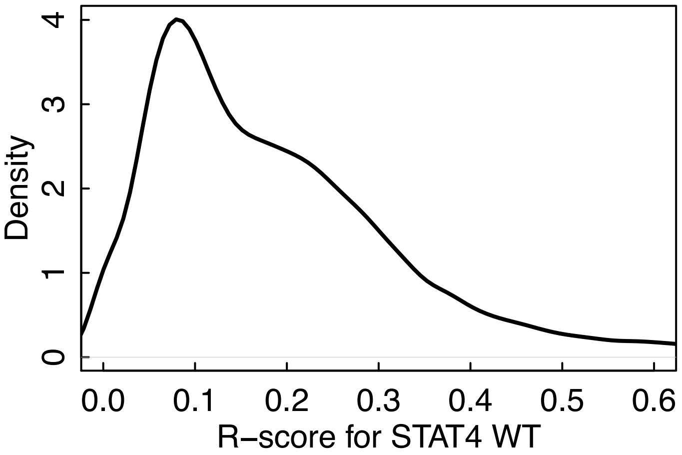

Fig. S8: The distribution of regulatory scores of all genes for STAT4 in wild type mouse Th1 cells. Note it is a mixture of non-target (left side) and target (right side) genes.