Platform validation: Comparison of sequence-specific transcription factor determinations by ChIP-seq and ChIP-qPCR

From the laboratory of Richard M. Myers
HudsonAlpha Institute for Biotechnology
601 Genome Way
Huntsville, AL 35806

Prepared by: Jason Gertz, Timothy E Reddy, Florencia Pauli, and Richard M Myers
September 9, 2011

As part of our contribution to the ENCODE Project, as well as our extended interests in understanding gene regulation in humans, we have performed large numbers of chromatin immunoprecipitation experiments, followed by high-throughput DNA sequencing (ChIP-seq). This has involved the analysis of more than 60 DNA binding proteins, including general and sequence-specific transcription factors. In an effort to understand the performance of the DNA sequencing, we sought to evaluate ChIP-seq results with an alternative approach, ChIP-qPCR. We quantitatively compared the density of reads (binding site signal) in ChIP-seq binding sites to measurements of binding site enrichment using ChIP-qPCR. We tested the concordance of ChIP-seq and ChIP-qPCR results for 12 transcription factors and found an overall good agreement between ChIP-seq signal and ChIP-qPCR enrichment. These results indicate that high-throughput DNA sequencing maintains an overall robust representation of immunoprecipitated material.

Methods

Chromatin Immunoprecipitation of sequence-specific factors

We performed chromatin immunoprecipitation (ChIP) experiments as previously described (Johnson et al. 2007) with $8x10^7$ cells for the human cell lines GM12878 or

K562. ChIP-seq libraries constructed using the method described in Johnson et al. are denoted as PCR2x, because there are two PCR reactions performed, before and after size selection. ChIP-seq libraries labeled as PCR1x were constructed with one PCR reaction that was performed after size selection. Libraries were sequenced on Genome Analyzers (Illumina).

Target selection and primer design

Binding sites for sequence-specific transcription factors were identified using MACS (Zhang et al. 2008) by comparing a sequenced input library (sonicated chromatin with formaldehyde crosslinks reversed) to a sequenced ChIP library. Binding sites were then ranked by their fold-enrichment, and 44 binding sites were randomly chosen to uniformly cover the fold-enrichment levels. Primer pairs were designed for each binding site using Primer3 (Rozen and Skaletsky 2000) with the following parameters: 50 bp to 100 bp amplicon size, 19 bp to 26 bp primer length, 63°C to 67°C melting temperature range for each primer. In addition to experimental primers, 2 pairs of negative primers were designed against non-conserved and non-repetitive intergenic sequences located far from any known genes using the same Primer3 parameters.

qPCR assay and analysis

We used 0.2% of the total DNA purified from a ChIP in each experimental qPCR reaction. A standard curve was also performed for each primer pair using 0.5 ng, 5 ng, and 50 ng of stock Human Genomic DNA (Roche). Two 96-well plates were set up for each ChIP, with 22 binding site primer pairs and both negative control primer pairs on each plate. For each qPCR reaction, DNA was mixed with 0.5 μM (final concentration) of each primer, 10 μL 2x DyNAmo Flash SYBR Green qPCR mix (Finnzymes F-415) in a total reaction volume of 20 μL. The reactions were incubated at 95°C for 7 minutes, followed by 40 cycles of 94°C for 10 seconds and 60°C for 30 seconds on an iCycler IQ Real-Time PCR Detection System (Biorad). Automatically detected threshold cycles were used to determine relative abundance in each reaction and the experimental wells

were compared to the corresponding primers' standard curve to determine the effective concentration of each binding site. Effective concentrations were then compared to the negative controls to determine the enrichment of the binding site in the ChIP. Spearman rank correlations between fold enrichments from qPCR and fold enrichments from ChIP-seq were calculated in Prism. qPCR fold enrichments of 2-fold or greater were considered positive when calculating false positive rates.

Results

We performed qPCR on ChIPs with antibodies that target 12 sequence-specific transcription factors in the human cell lines GM12878 or K562. Rank correlation between ChIP enrichments from qPCR and ChIP-seq were significant for 11 out of the 12 ChIP experiments, with an average correlation of 0.635 (Table 1). NRSF was the only transcription factor that did not show significant correlation between qPCR and ChIP-seq. All of the data points for individual experiments can be found in the Supplemental Figures.

The qPCR results were also used to experimentally evaluate the false positive rate of ChIP-seq binding sites. The positive predictive value (PPV) for each transcription factor is shown in Table 1. On average, 89.7% of ChIP-seq binding sites exhibited enrichment when assayed by qPCR. For every transcription factor more than 75% of ChIP-seq binding sites were enriched over negative controls. Overall, these results show that the vast majority of ChIP-seq binding sites can also be detected by qPCR of ChIP material and that the quantitative enrichments found in ChIP-seq data are usually recapitulated by qPCR.

References

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**(5830): 1497-1502.

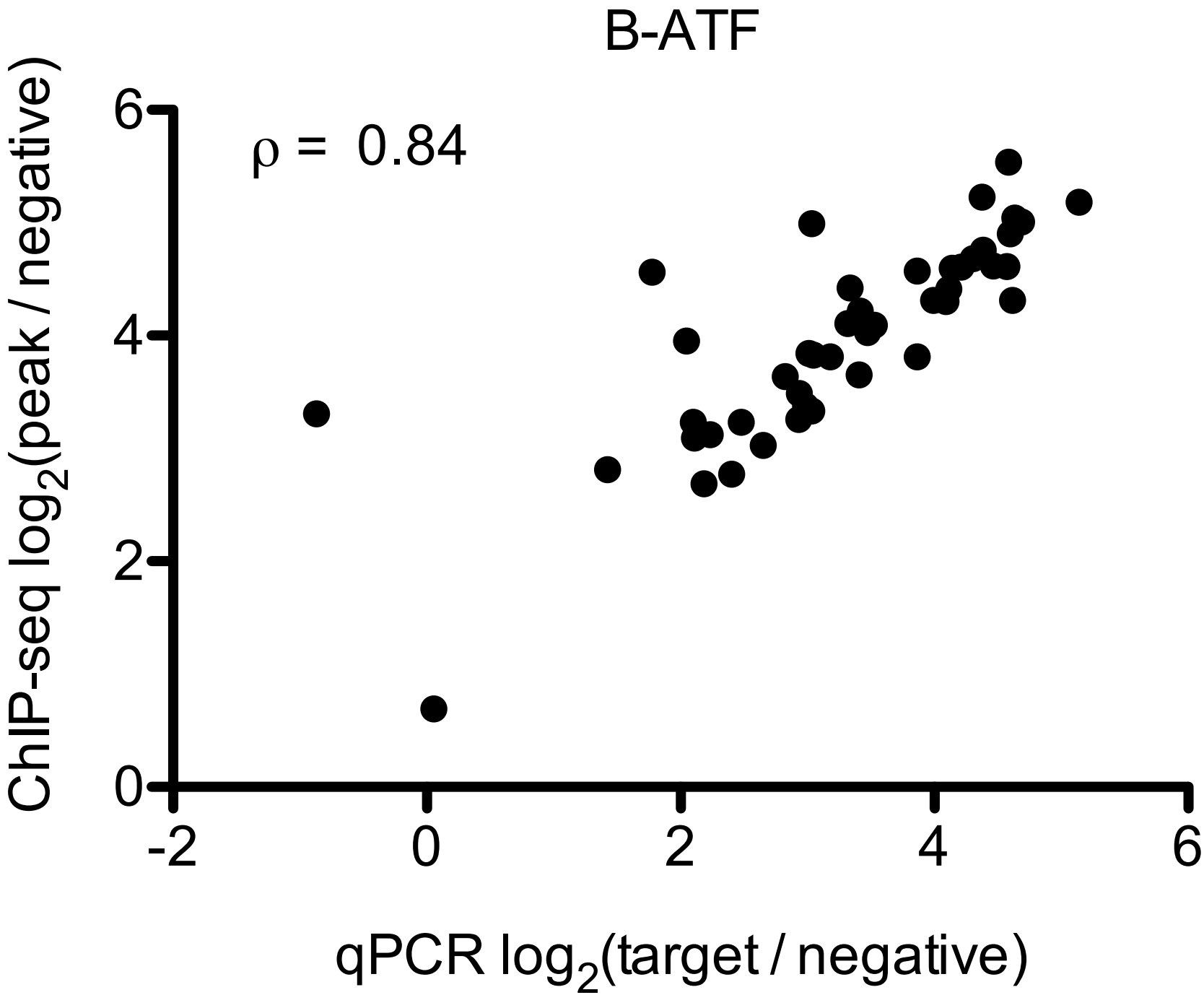
Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-386.

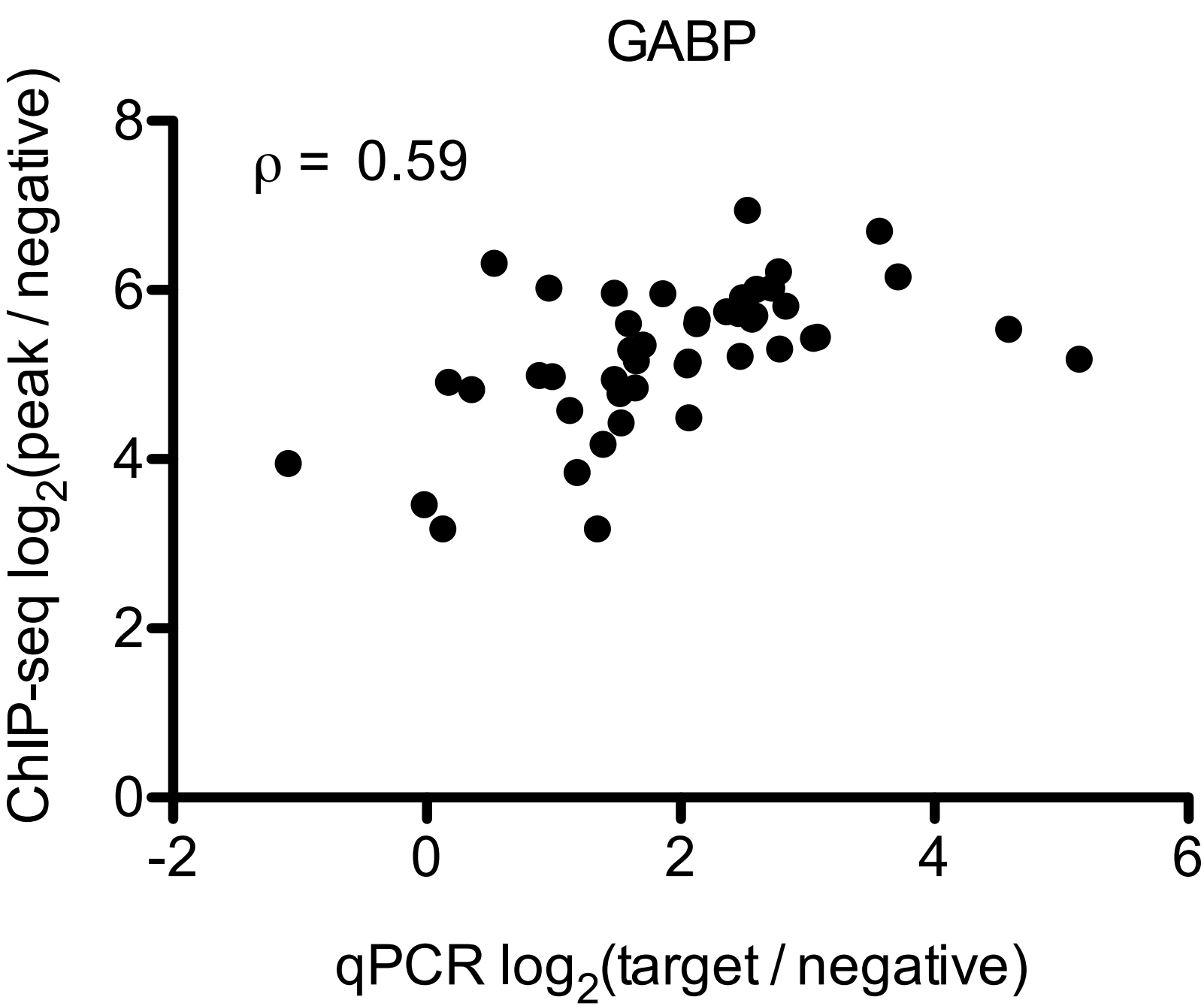
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

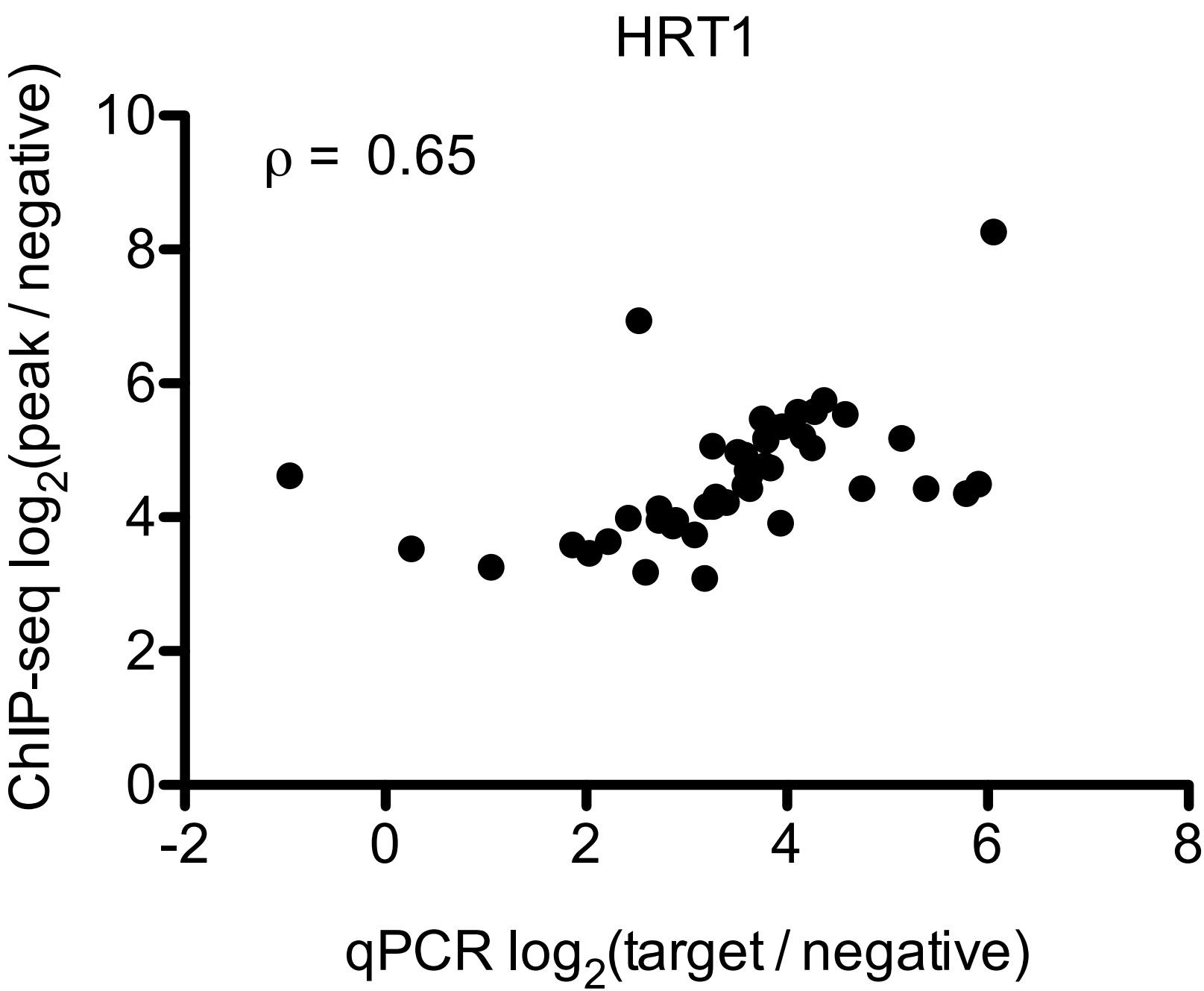
Table 1. Performance of qPCR validation

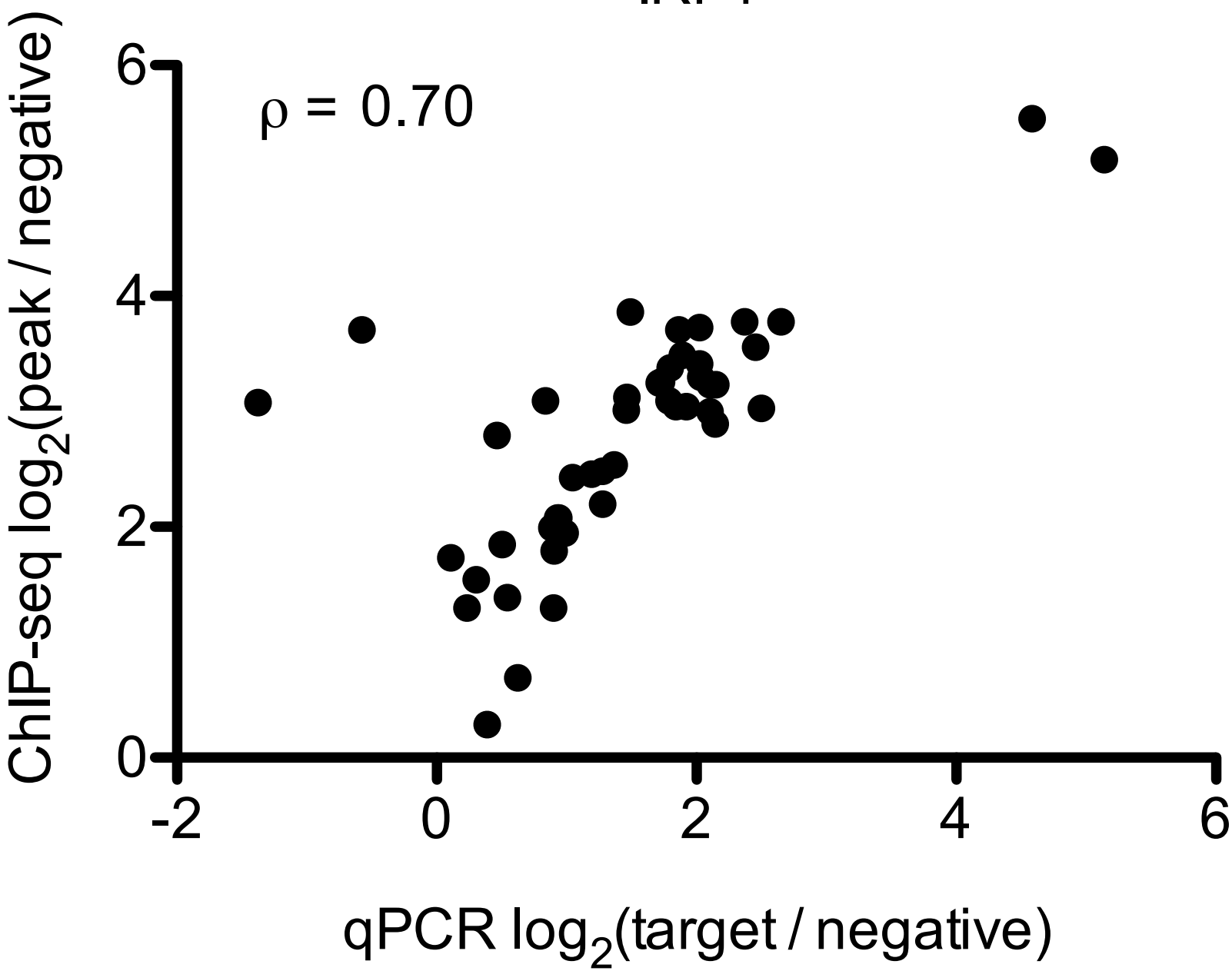
Transcription Factor	DNA type	Cell type	Dataset identifier	Protocol	Correlation between Chip-seq and qPCR	Positive predictive value (PPV)
BATF	ChIP	GM12878	SL839	PCR1x	0.84	0.978
GABP	ChIP	GM12878	SL203	PCR2x	0.59	0.864
HRT1	ChIP	K562	SL845	PCR1x	0.65	0.953
IRF4	ChIP	GM12878	SL838	PCR1x	0.70	0.773
NRSF	ChIP	Gm12878	SL202	PCR2x	-0.13	0.767
OCT2	ChIP	GM12878	SL614	PCR1x	0.58	0.773
P300	ChIP	GM12878	SL551	PCR1x	0.75	0.955
PAX5	ChIP	GM12878	SL675	PCR1x	0.64	0.864
PBX3	ChIP	GM12878	SL647	PCR1x	0.76	0.909
POL2	ChIP	GM12878	SL748	PCR1x	0.76	1
PU.1	ChIP	K562	SL646	PCR1x	0.69	1
SIX5	ChIP	K562	SL842	PCR1x	0.79	0.932

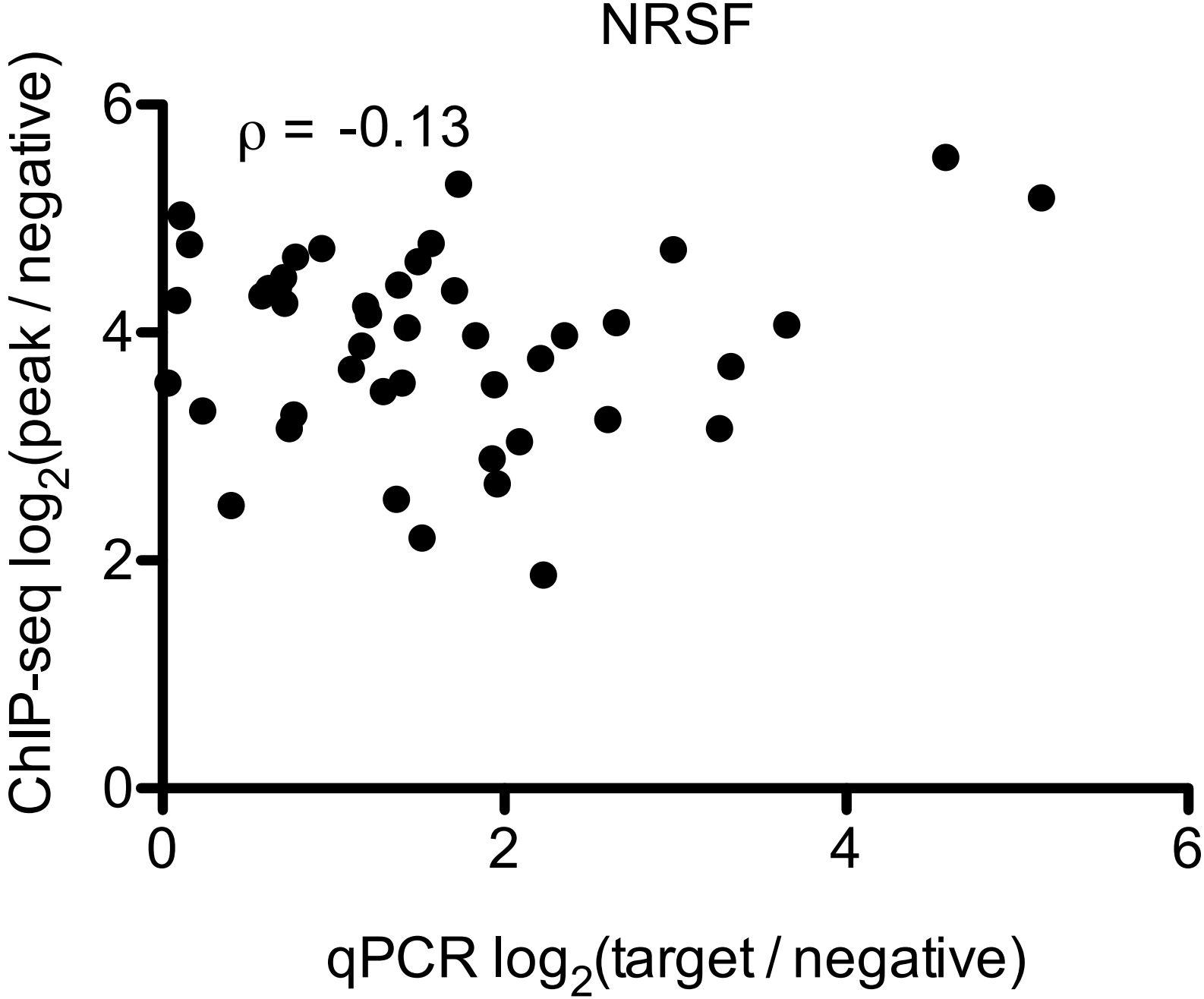
Supplementary Figures. Comparison of ChIP-seq signal and relative qPCR expression. Each graph shows the 44 binding sites assayed by qPCR for a particular factor with log base 2 of the ChIP-seq signal divided by the reverse crosslink control on the y-axis and the log base 2 of the enrichment of the binding site compared to negative control regions measured by qPCR on the x-axis.

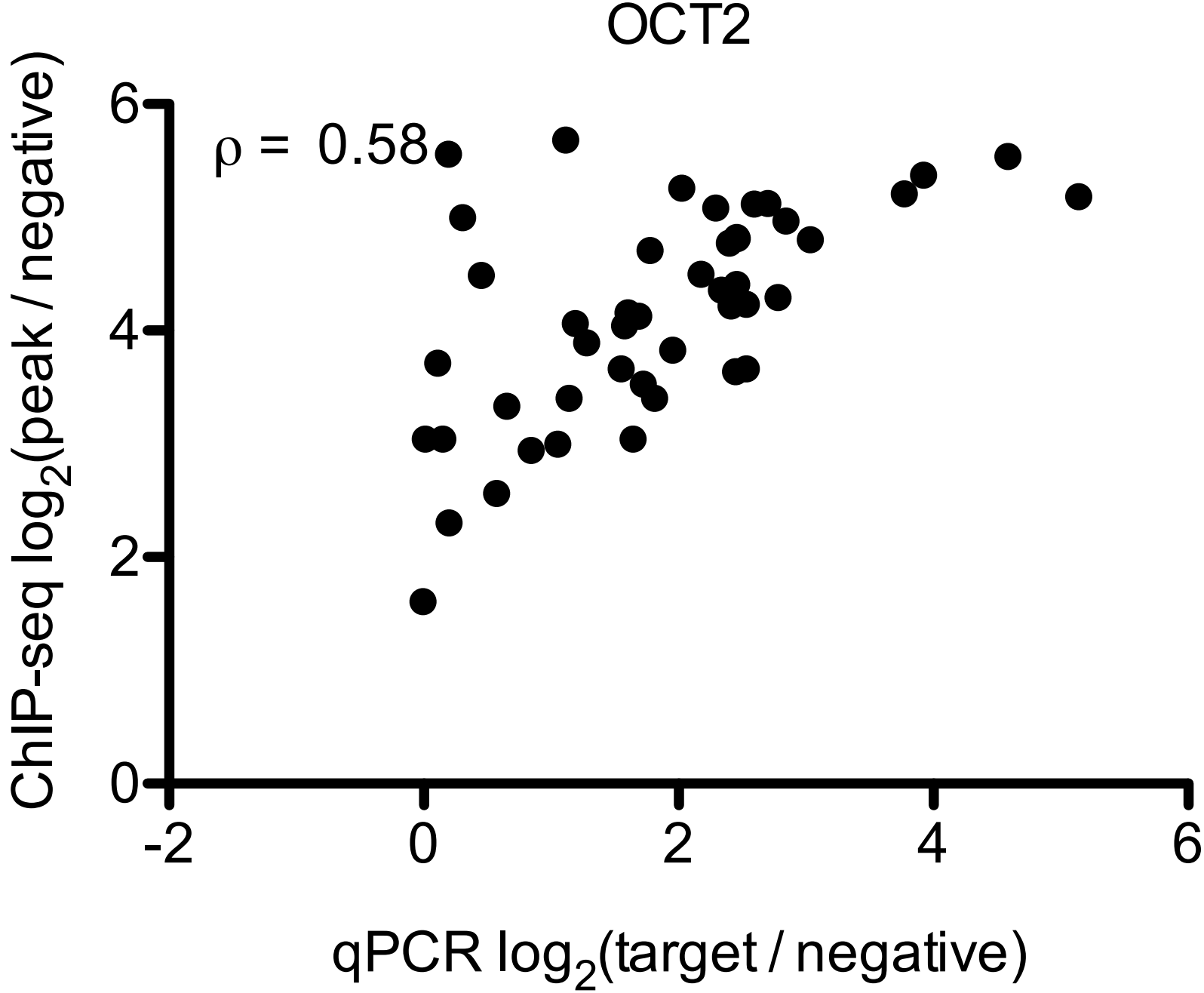

B-ATF
ρ = 0.84
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
6
4
2
0
-2
0
2
4
6

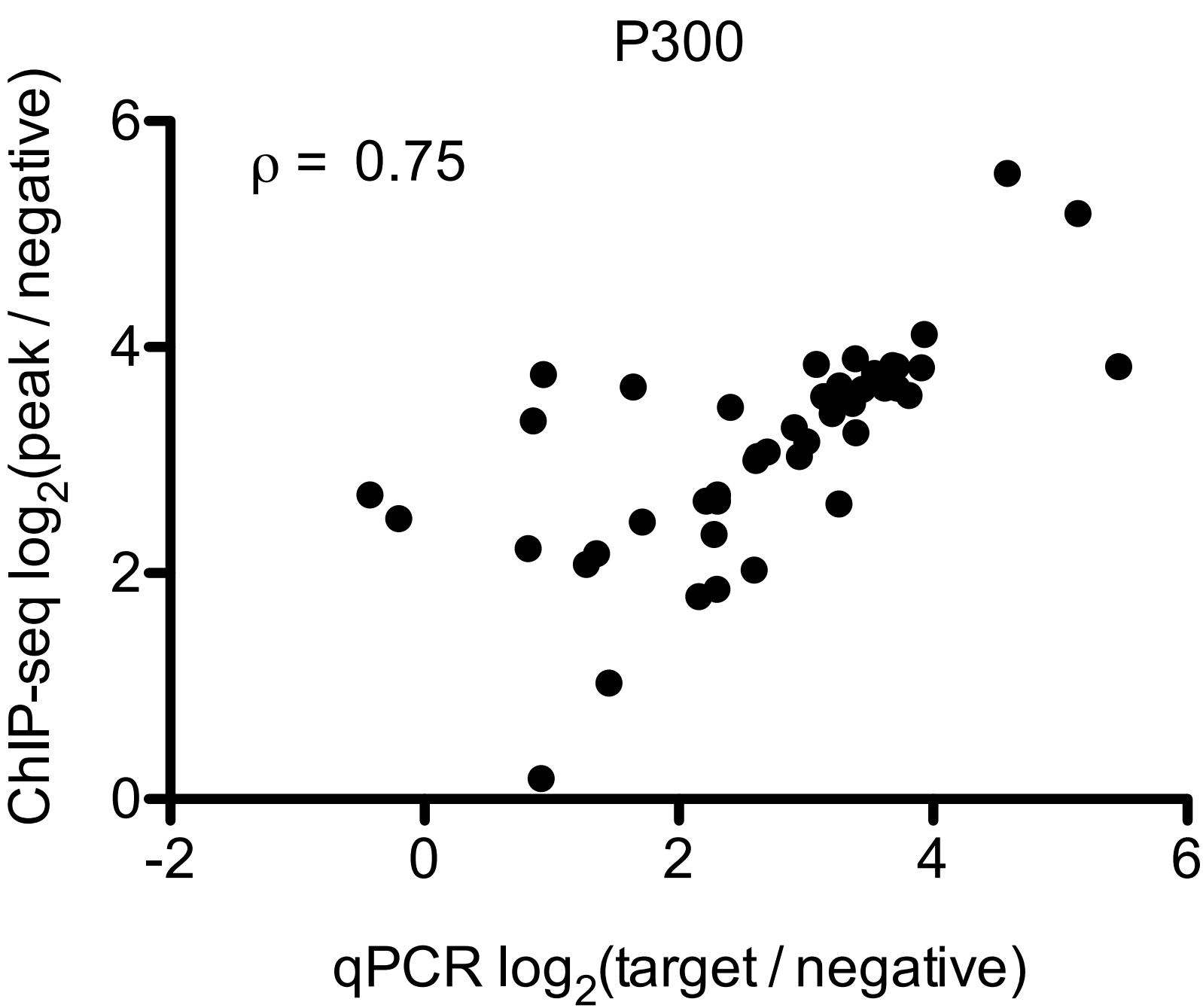

GABP
ρ = 0.59
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
8
6
4
2
0
-2
0
2
4
6

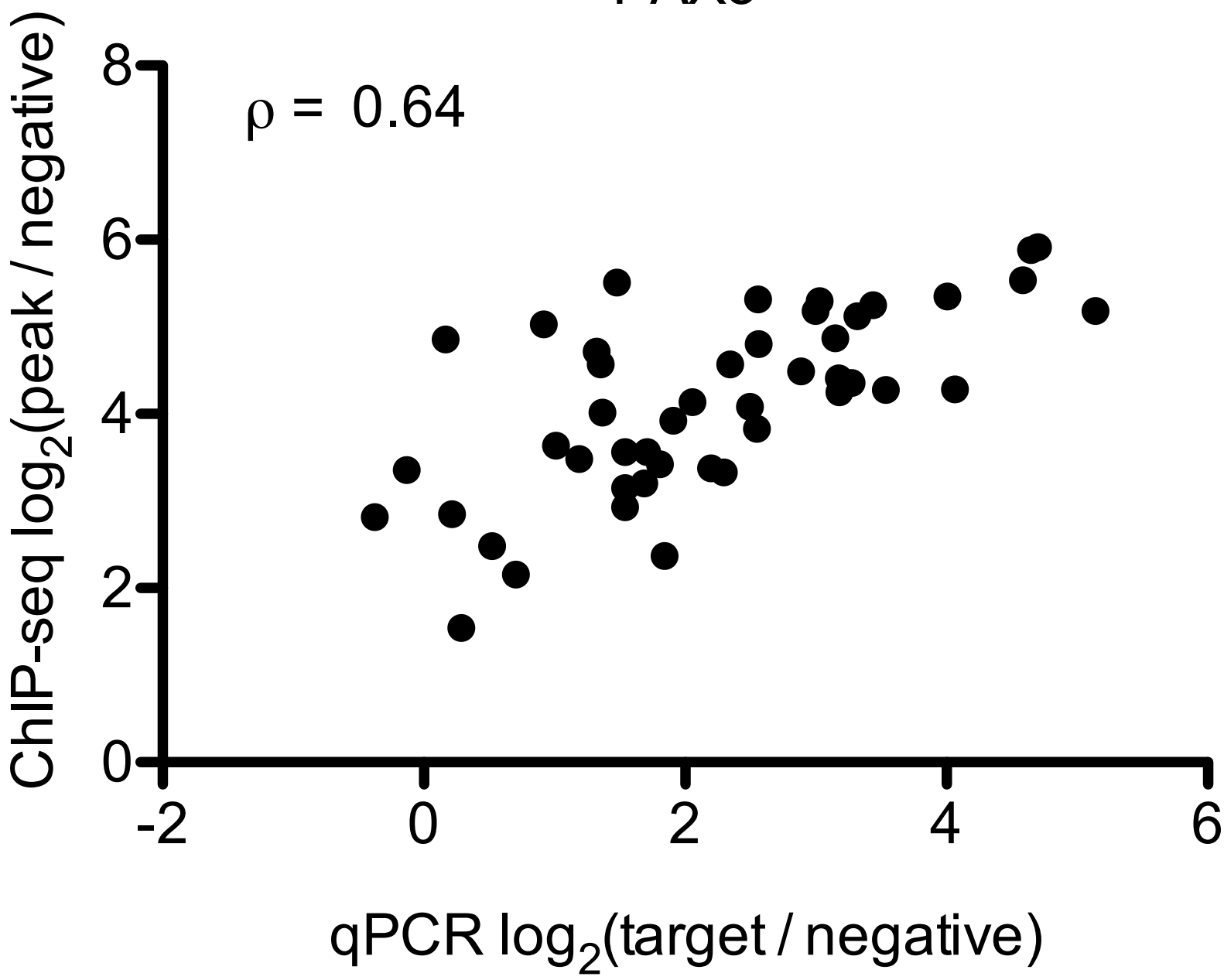

HRT1
ρ = 0.65
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
10
8
6
4
2
0
-2
0
2
4
6
8

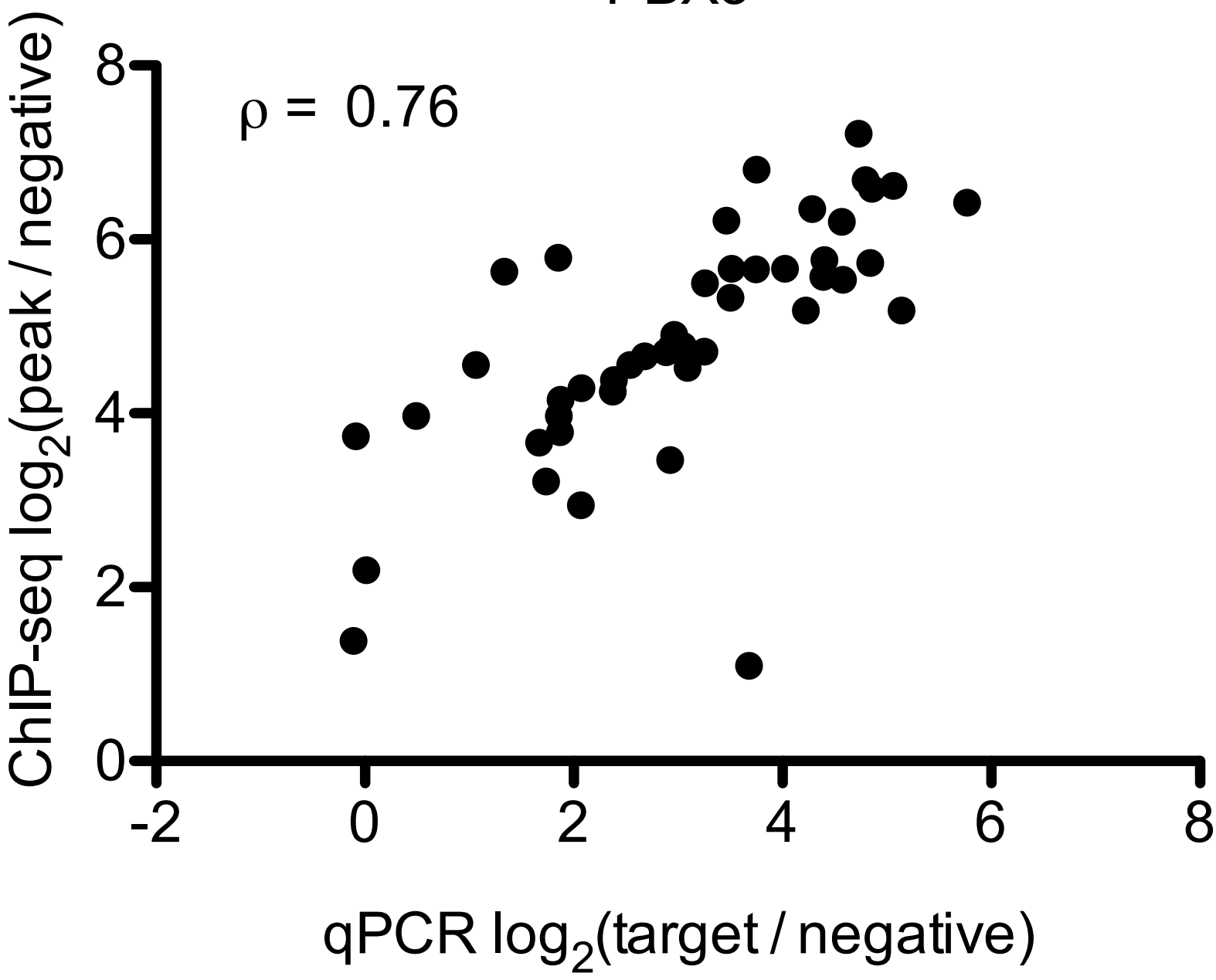

IRF4
ρ = 0.70
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
6
4
2
0
-2
0
2
4
6

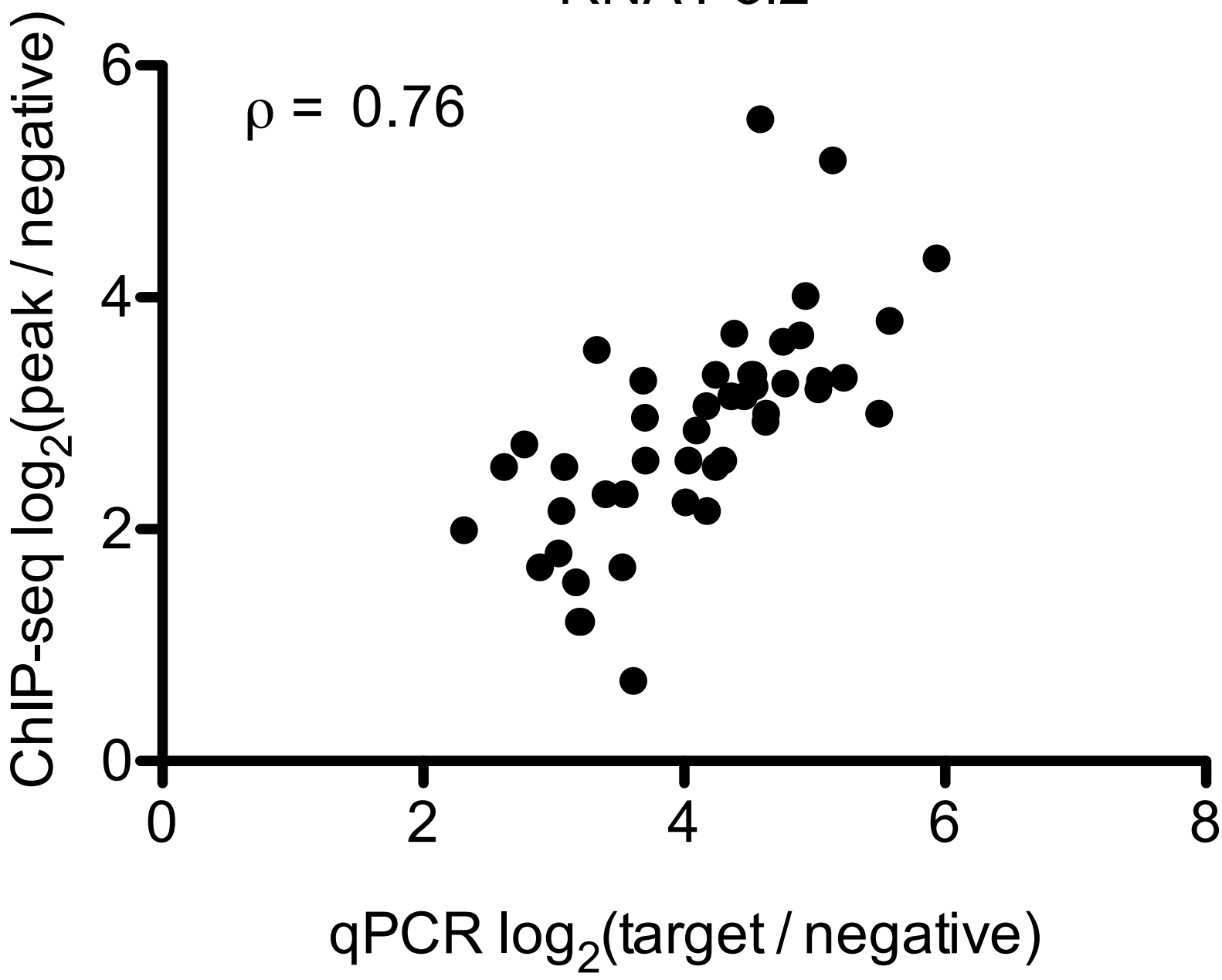

NRSF
ρ = -0.13
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
0
2
4
6
0
2
4
6

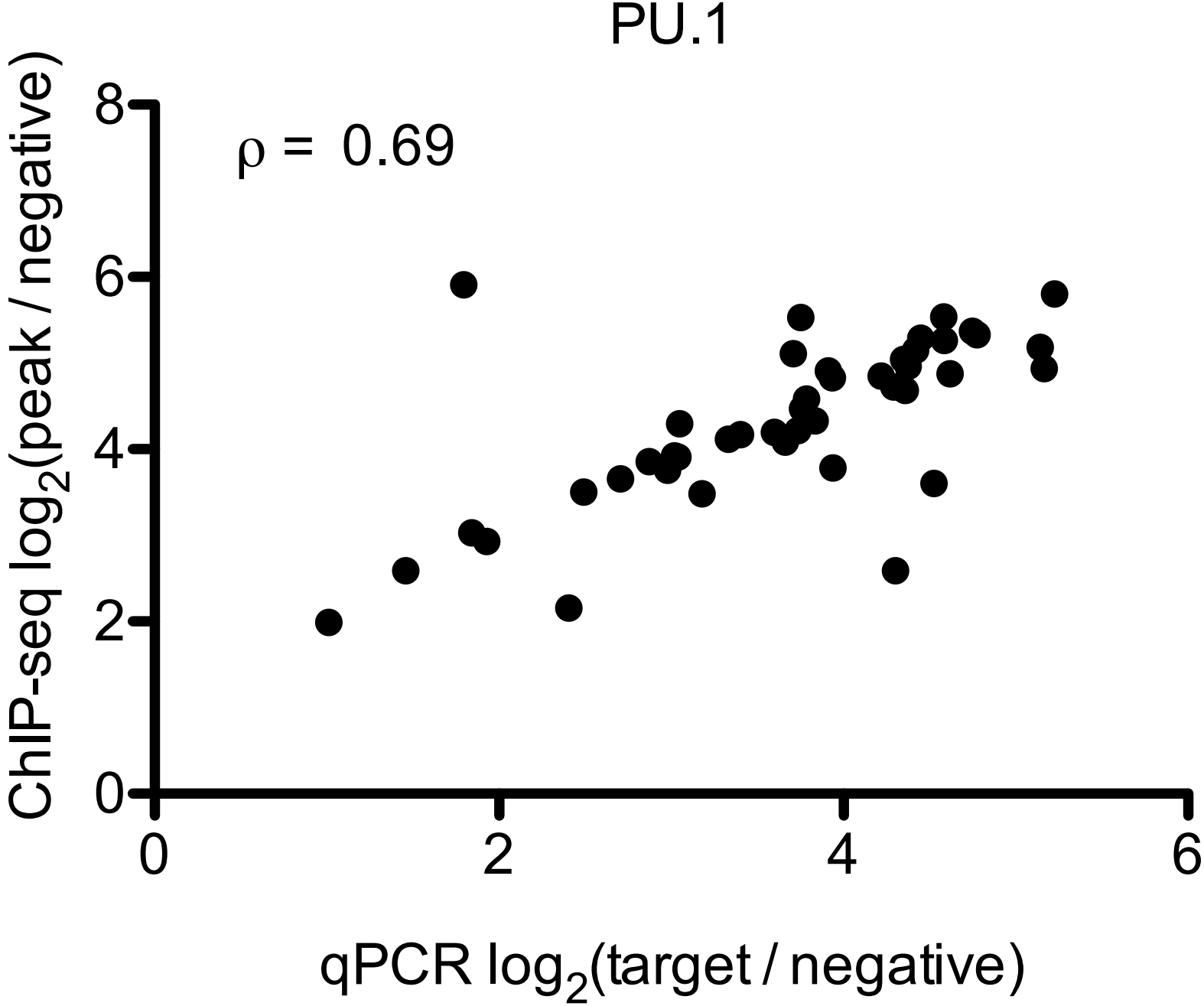

OCT2
ρ = 0.58
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
6
4
2
0
-2
0
2
4
6

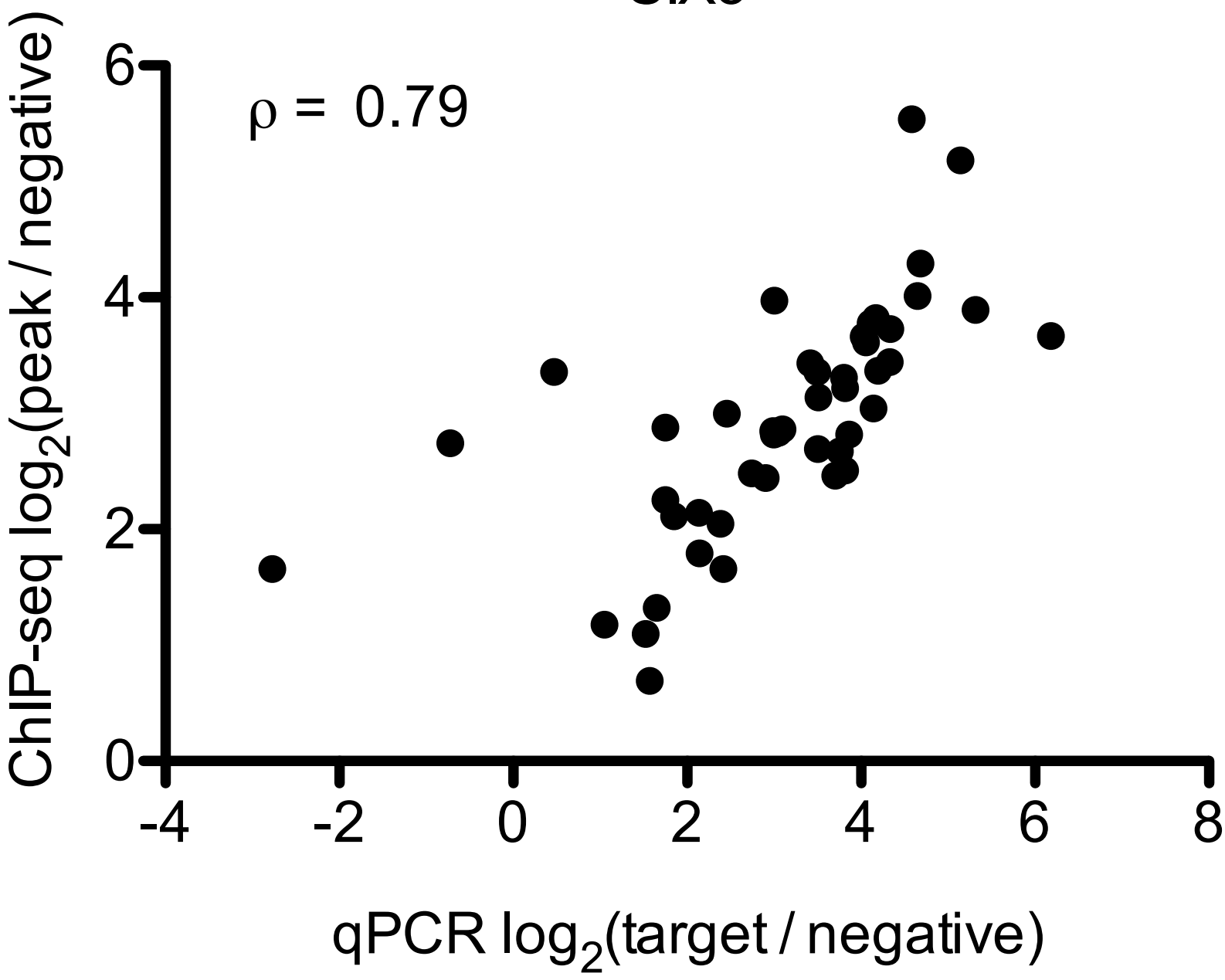

P300
ρ = 0.75
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
6
4
2
0
-2
0
2
4
6


PAX5
ρ = 0.64
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
8
6
4
2
0
-2
0
2
4
6


PBX3
ρ = 0.76
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
8
6
4
2
0
-2
0
2
4
6
8


RNA Pol2
ρ = 0.76
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
0
2
4
6
8


PU.1
ρ = 0.69
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
0
2
4
6
8
0
2
4
6


SIX5
ρ = 0.79
ChIP-seq log2(peak / negative)
qPCR log2(target / negative)
6
4
2
0
-4
-2
0
2
4
6
8