Saturation analysis of ChIP-seq data for reproducible identification of binding peaks

- Supplemental Material -

Peter Hansen[1,2], Jochen Hecht[1,2,3], Daniel M. Ibrahim[1,3],
Alexander Krannich[4], Matthias Truss[5] Peter N. Robinson[1,2,3,6]

[1]*Institute for Medical and Human Genetics, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany.*
[2]*Berlin Brandenburg Center for Regenerative Therapies (BCRT), Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany.*
[3]*Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany.*
[4]*Department of Biostatistics, Clinical Research Unit, Berlin Institute of Health, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany.*
[5]*Labor für Pädiatrische Molekularbiologie, Charité-Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany.*
[6]*Institute for Bioinformatics, Department of Mathematics and Computer Science, Freie Universität Berlin, Takustrasse 9, 14195 Berlin, Germany.*

Contents

Supplemental Methods . . . 3
Figure S1 Fragment length estimation . . . 5
Figure S2 Coverage profiles for RNAPII in HeLa-S3 cells . . . 6
Figure S3 Empirical distribution of 5' ends . . . 7
Figure S4 Effect of size of control dataset on P-values . . . 8
Figure S5 CWOP analyis . . . 9
Figure S6 CWOP analyis for TAF1 in hESC . . . 10
Figure S7 CWOP analysis of dataset K562-NRSF-REP1 . . . 11
Figure S8 Distribution of TFDSs for RNAPII . . . 12
Figure S9 Distance distributions of TFDSs for RNAPII . . . 13
Figure S10 Distribution of TFDSs for H3K4me3 . . . 14
Figure S11 Distance distributions of TFDSs for H3K4me3 . . . 15
Table S1 ChIP-seq datasets . . . 16
Table S2 Runtime - Fragment length estimation . . . 17
Table S3 Peak overlaps . . . 18
Table S4 Pearson correlation coefficients . . . 19
Table S5 CWOP Analysis . . . 20
Table S6 Peak overlaps for peaks with IDR$\leq$ 0.01 . . . 21

Table S7 Motif content analysis 22
Table S8 TFDSs for RNAPII 23
Table S9 TFDSs for H3K4me3 23
Table S10 TFDS - Reproducibility 24
Table S11 TFDS - Simulation study 24
Table S12 Runtime - Peak calling 24

Supplemental Methods

IDR analysis.

The IDR tool (Li et al. 2011) takes as input peak lists in ENCODE narrowPeak format for two replicates of ChIP-seq experiments. The peak lists can be sorted either by signal value, P-Value, or Q-value. For both replicates the peak lists are truncated after the top n peaks (e.g n=100,000). Intuitively, we expect there to be a large degree of consistency in the ranks of peaks that represent true signals, but as we move down the list from strong signals to weak signals, the degree of rank consistency drops, often showing a well delineated transition from signal to noise. This transition is visualized in the correspondence profile Ψ and the first derivative of the correspondence profile, Ψ'. The assumption is that once the signal becomes noisy, the degree of reproducibility will drop substantially, reflected by a change in the slope of the Ψ' curve from around zero to negative or positive values. The correspondence curve is constructed by sequentially determining the proportion of overlapping signals in the upper ranks of peaks for the two pseudoreplicates, i.e., the proportion of overlapping signals in the top 1% of ranks, in the 2% of ranks, and so on. If the proportion of overlapping peaks was always 100%, then the correspondence profile would be simply a diagonal line with slope 1. If the overlap proportion becomes smaller, then the correspondence profile Ψ will move away from the diagonal, with a corresponding change in the slope of Ψ'. At some point the correspondence profile returns to the diagonal, because the fraction of the top 100% ranks must be 100%. This implies that the slope of Ψ becomes greater than one. The later this happens, the more consistent the two ranked lists are (Fig. 3E). A second and independent component of the IDR procedure is based on copula mixture model and assumes the data consists of a reproducible group and an irreproducible group (Li et al. 2011). The IDR procedure assigns each overlapping signal a posterior probability of being part of the irreproducible group, which is reported as the Irreproducible Discovery Rate (IDR), which is similar to the false discovery rate (FDR). The IDR is interpreted as the expected probability that selected signals come from the irreproducible group (Fig. 3F).

CWOP analysis.

We defined consistently weak overlapping peaks (CWOPs) as overlapping peaks with IDR$\leq$0.01 for which both pseudoreplicates have signal scores smaller than the mean score of all overlapping peaks with IDR$>$0.01 (i.e., "irreproducible peaks") (Fig. S5). We then calculated the proportion of CWOPs amongst all peaks with IDR$\leq$0.01. For an optimal dataset, this proportion would be zero. Q had more than 10% CWOPs in only one of the 38 datasets, and identified no CWOPs in 31 of the 38 datasets. In contrast, MACS2 and PeakSeq identified more than 10% CWOPs in 13 and 14 experiments, and for PeakSeq, the IDR procedure fails completely in three cases by classifying only consistently weak overlapping peaks as reproducible (Table S5). Figures S6 and S7 show two further examples. We then investigated the cases with more than 10% CWOPs and found many ties at the lower peak ranks. The IDR procedure was developed only for ranking systems that produce scores without ties [1]. However, depending on the peak calling method and data quality, all peak callers can produce ties, especially at the lower ranks. This is one reason why some peak callers and significance measures are considered to be compatible with the IDR procedure and others are not. For example, the signal enrichment values generated by SPP are considered to be well compatible [2], which could be largely confirmed by our analysis.

Peak calling parameters

For Q the fragment length estimation via hamming distance was skipped, by using `-fla 1`. We set no significance cutoff. In this case Q reports by default all positions that are covered by at least

one qfrag. For MACS2 we set `-gsize` to 2.7×10^9, `-shiftsize` to $\ell/2$, the significance threshold `-pvalue` to 10^{-1} and skipped the model building step via the option `-nomodel`. For PeakSeq we set `Enrichment_fragment_length` to ℓ, `Minimum_interpeak_distance` to $2\ell/3$, `target_FDR` to 0.99 and `max_Qvalue` to 0.99, and otherwise used the default settings. To perform fragment-length estimation with SPP, we used the function `remove.local.tag.anomalies` to remove singular positions with extremely high tag count relative to the neighborhood. For peak calling we used the SPP function `find.binding.positions` and set `whs` to whs, `fdr` to 0.99 and used the `WTD` method. For the motif content analysis we kept with the recommendation of the developer of SPP and used the `MTC` method and additionally applied the function `select.informative.tags` to select mapped reads with acceptable alignment quality.

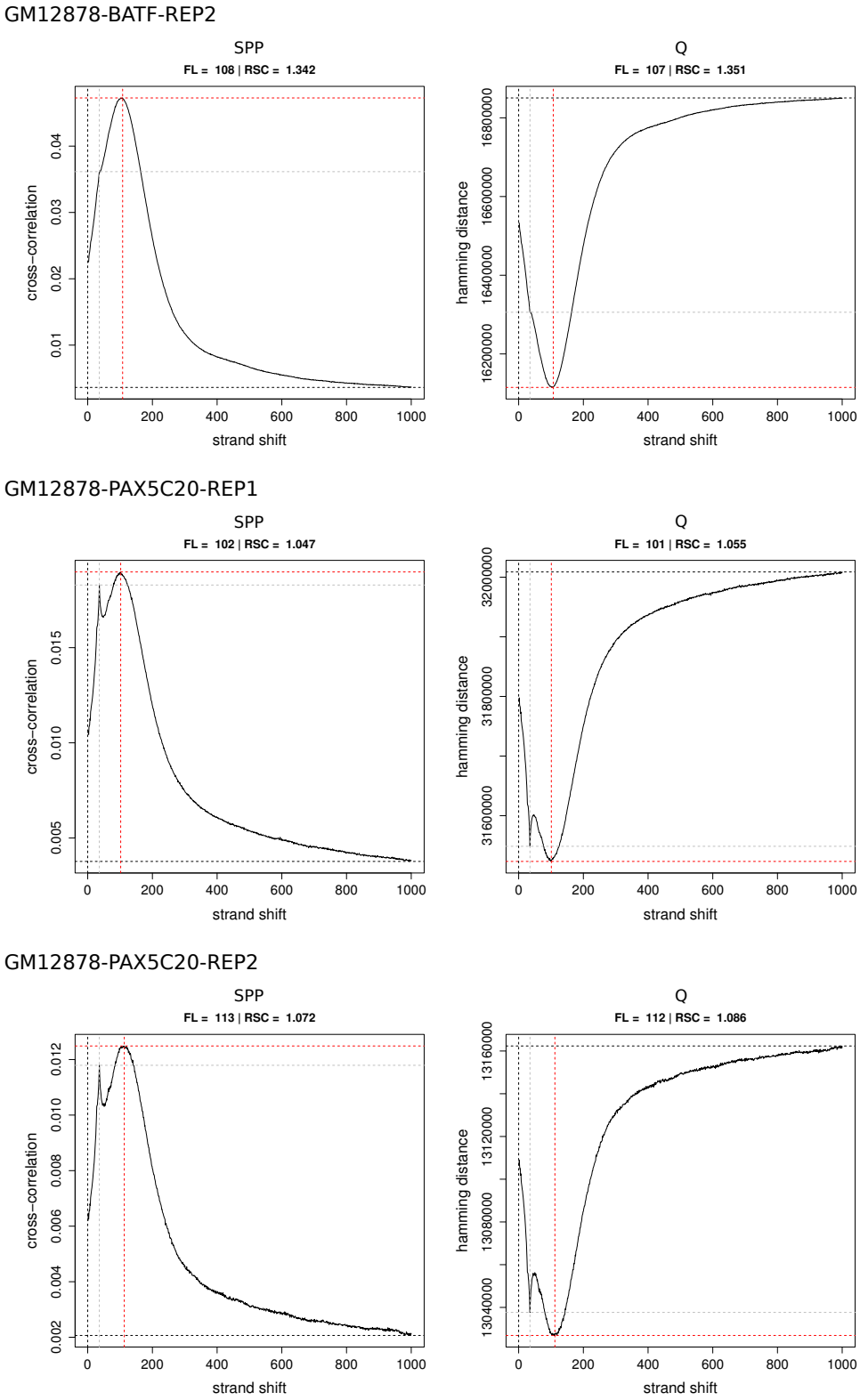

Figure S1. Fragment length estimation. We compared the fragment length estimation of Q and SPP on 38 datasets analysed in this study. In all cases, the estimated fragment length of Q is equal to that of SPP minus one. Three representative examples are shown.

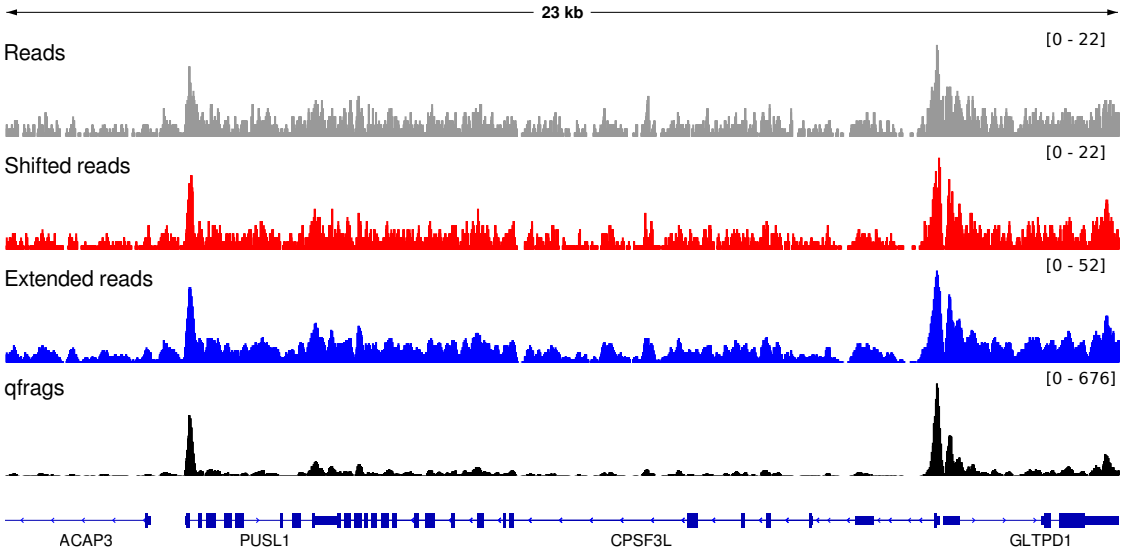

Figure S2. Coverage profiles for RNAPII in HeLa-S3 cells. ChIP-seq peak calling methods transform the raw reads in different ways before statistical hypothesis testing of peak enrichment. The four tracks show histograms of coverage for raw reads (grey), shifted reads (red), extended reads coverage (blue) and qfrags (black). Minimum and maximum coverages within the region chr1:1,240,182-1,263,869 are shown in square brackets and each track is scaled to its maximum value.

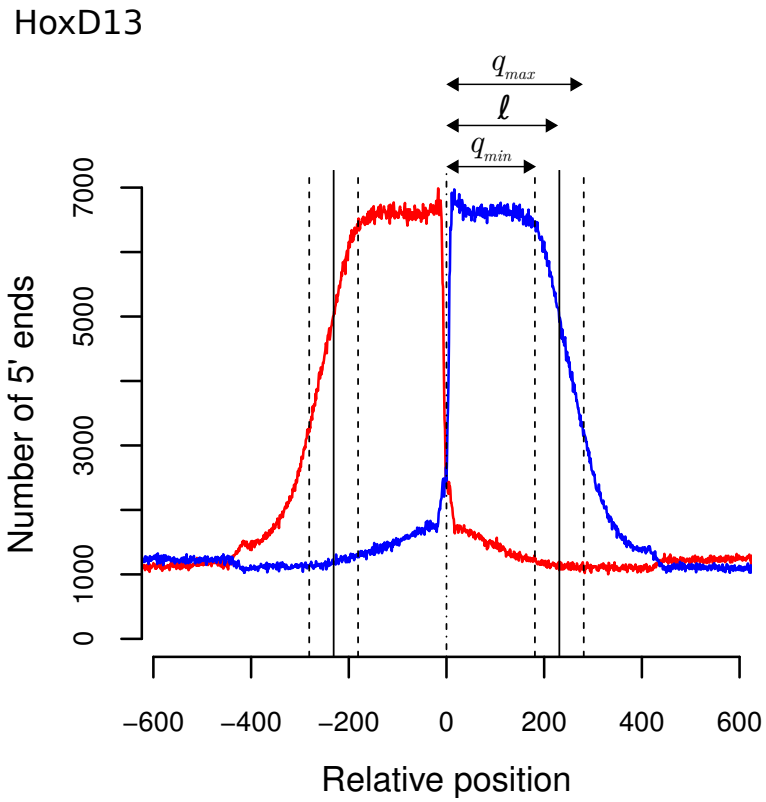

Figure S3. Empirical distribution of 5' ends. We determined the empirical distribution of 5' ends around peaks for a HoxD13 dataset [3]. Initially, we called peaks using SPP and conducted a de novo motif analysis using DREME. To enrich for true binding sites, we filtered for those having an occurence of the primary de novo motif (TTTWATKR) in a distance of at most 20 nucleotides. To eliminate imprecision of predicted binding positions, we centered them to middle position of nearest occurence of the primary motif. For the resulting $29,593$ filtered and centered predicted binding positions the numbers of 5' end positions of mapped reads for the positions $1,000$ nucleotides upstream and downstream were determined and plotted separately for the forward strand (red) and reverse strand (blue). Position 0 corresponds to the predicted centered binding sites. The average fragment length of 231, determined using SPP, is depicted as solid lines upstream and downstream. The dashed lines correspond to $q_{min} = 231 - 50$ and $q_{max} = 231 + 50$.

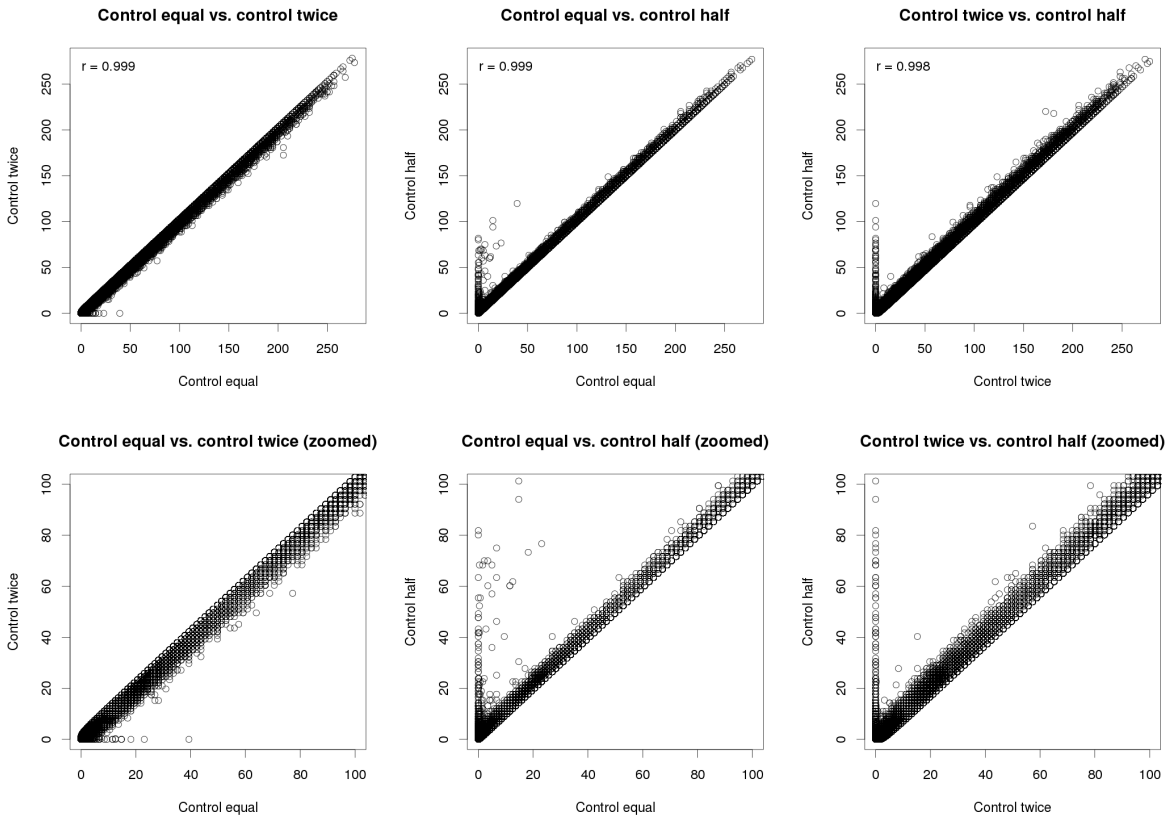

Figure S4. Effect of downsampling the control dataset on P-value. After removal of duplicates for the datasets corresponding to Heals3-Pol2-Rep2, a total of $23,494,468$ mapped reads were left for the treatment sample and $29,454,439$ for the control. The data for treatment was then randomly downsampled to $11,747,234$ mapped reads. Three control datasets of different sizes were generated from the original control data by downsampling to $23,494,468$ (*twice* the size of treatment), $11,747,234$ (*equal* to treatment) and $5,873,617$ (*half* of treatment). Q was applied to the downsampled treatment data with the three different controls and the P-values were ploted against each other. Since the signal detection is performed only on the treatment data, in all three cases the same set of $746,755$ peaks is detected. In the upper row the full range of P-Values is shown (*equal vs. twice* left, *equal vs. half* middle, *twice vs. half*). In the lower row the same plots are shown but zoomed into the range of 0 to 100. Overall the P-values are quite robust against the variation of the size of the control dataset. For smaller controls a relatively small number of peaks accumulate along the axis of the smaller control dataset. Manual inspection showed that those peaks almost exclusively correspond to mapping artefacts, due to repeats mainly in centromeric regions. Altogether, the more control data there is the fewer of those artefacts are detected as significant peaks.

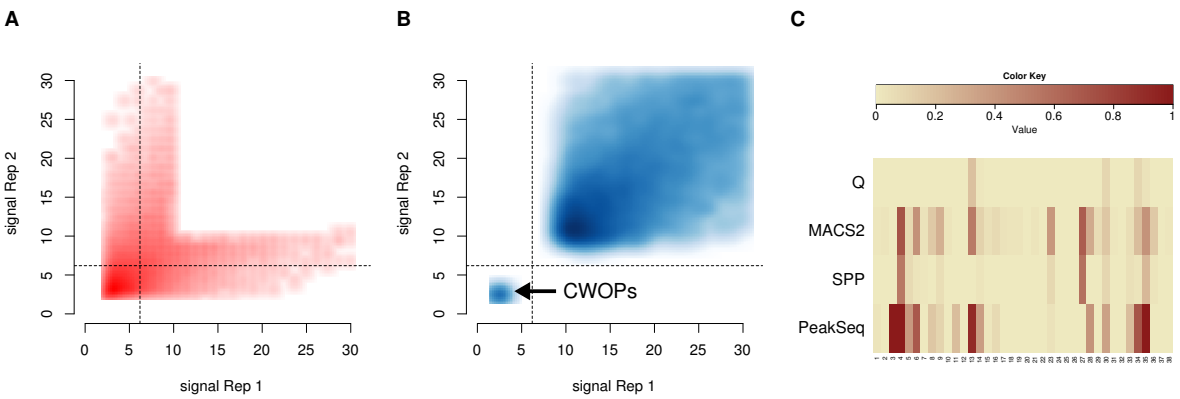

Figure S5. CWOP analyis. Scatter plot of the logarithms of the P-values derived from both pseudoreplicates shown in Figure 3A (Q), with red points in panel **(A)** representing data with an IDR>0.01, and blue points in panel **(B)** representing peaks with IDR≤0.01. The dotted lines show the mean value for all peaks with IDR>0.01. **(C)** Summary of the analyses for all 38 datasets. Q showed the best overall performance in the CWOP analysis (Table S5).

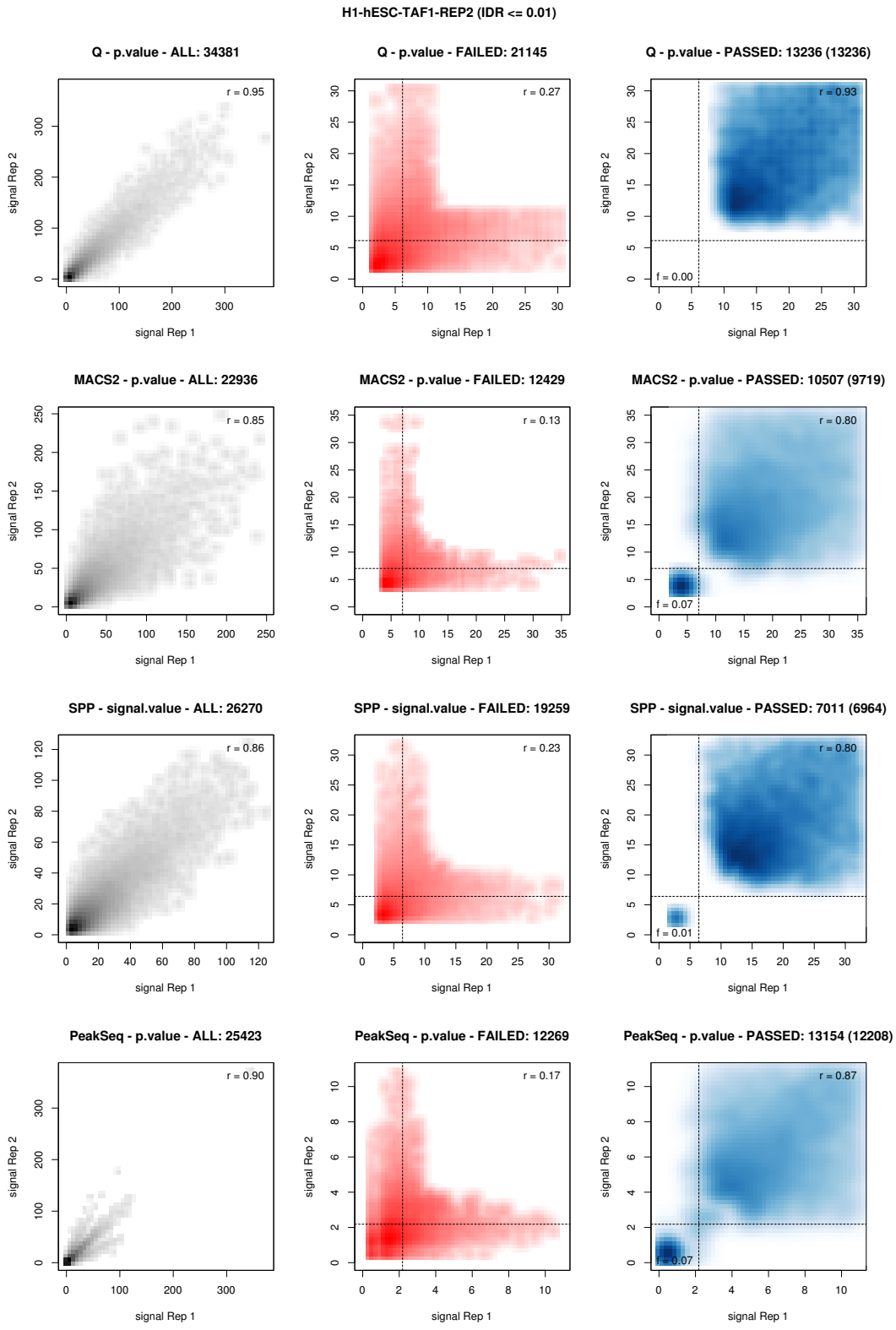

Figure S6. CWOP analyis for TAF1 in hESC. Detailed results for IDR CWOP analysis of dataset H1-hESC-TAF1-REP2 (16). The left column is a scatterplot of $-\log p$ for all P values of the pseudoreplicates, the middle column (in red) shows the P-values for peaks with IDR>0.01, and the right column (in blue) for peaks with IDR$\leq$ 0.01. The four rows show the results for Q, MACS2, SPP, and PeakSeq.

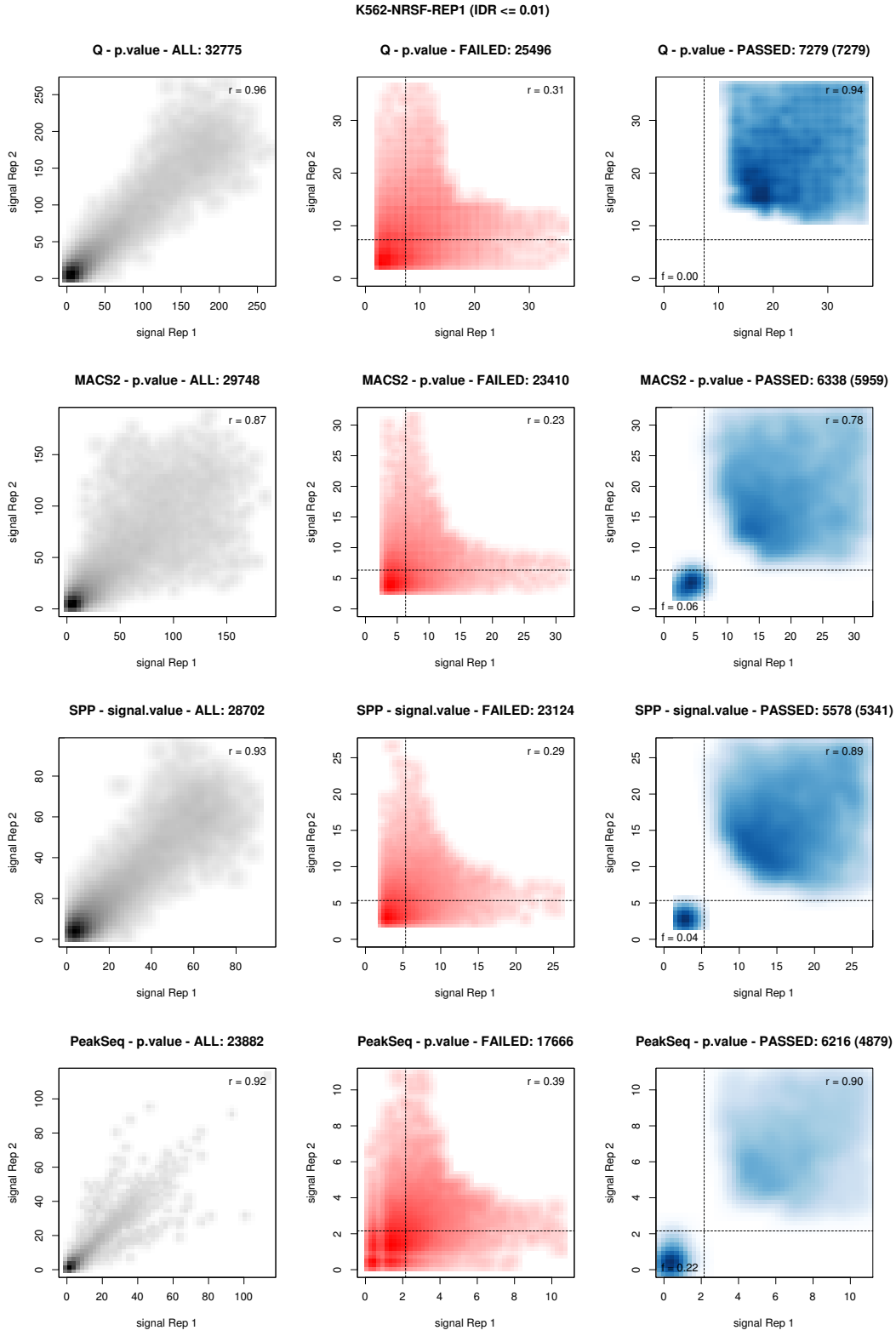

Figure S7. CWOP analysis of dataset K562-NRSF-REP1. See the legend to Figure S6 for explanations.

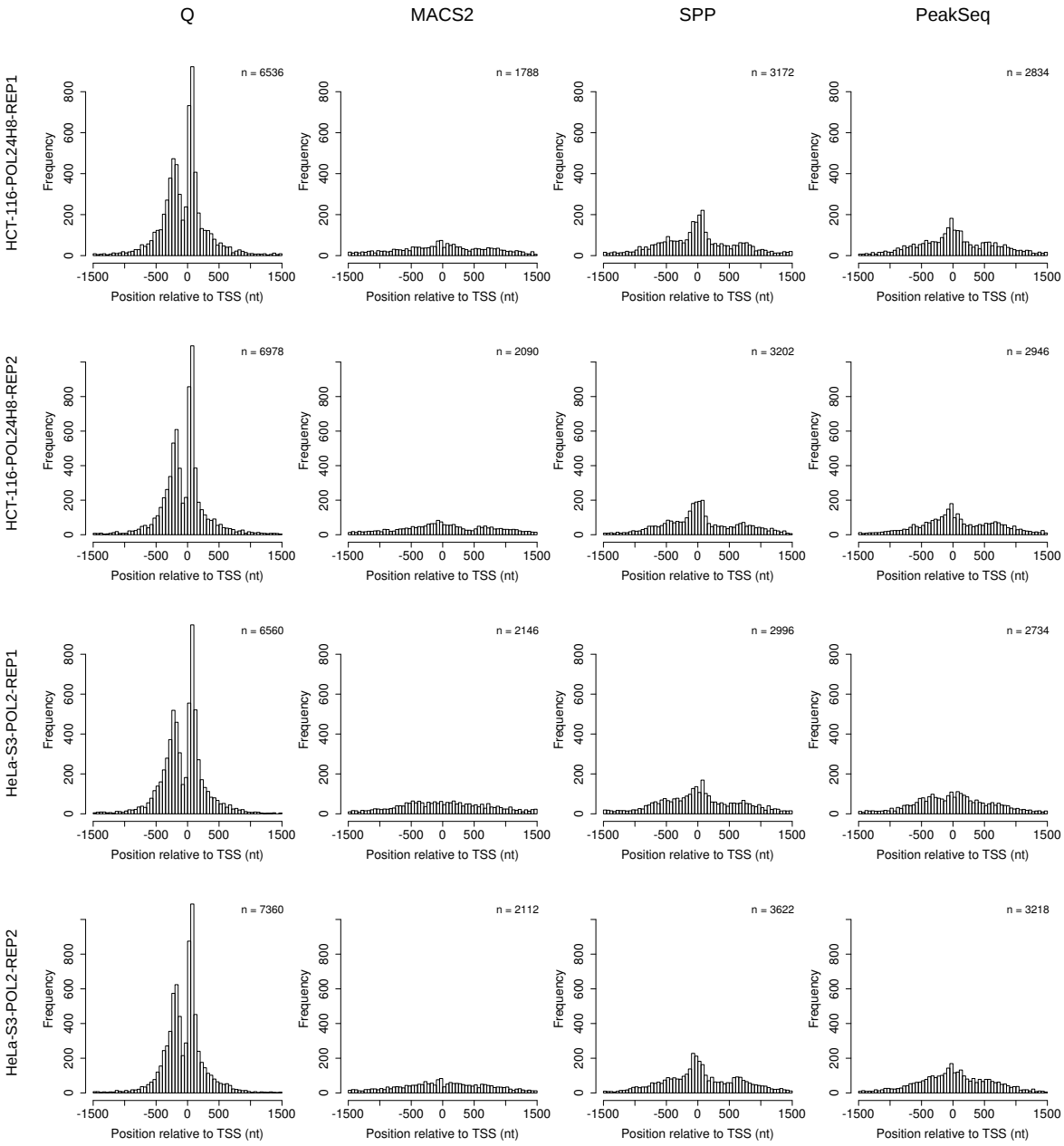

Figure S8. Distribution of TSS flanking double summits for RNAPII. Each row corresponds to one analyzed RNAPII dataset for the cell lines HCT-116 and HeLa-S3 and each column to one peak calling method. Each TSS flanking double summit consists of two summits directly up- and downstream of the transcription start site. The summits were integrated over all non overlapping promoters (TSS±1500) containing a TSS flanking double summit. The subplot at the upper left is identical with Figure 6A in the main text.

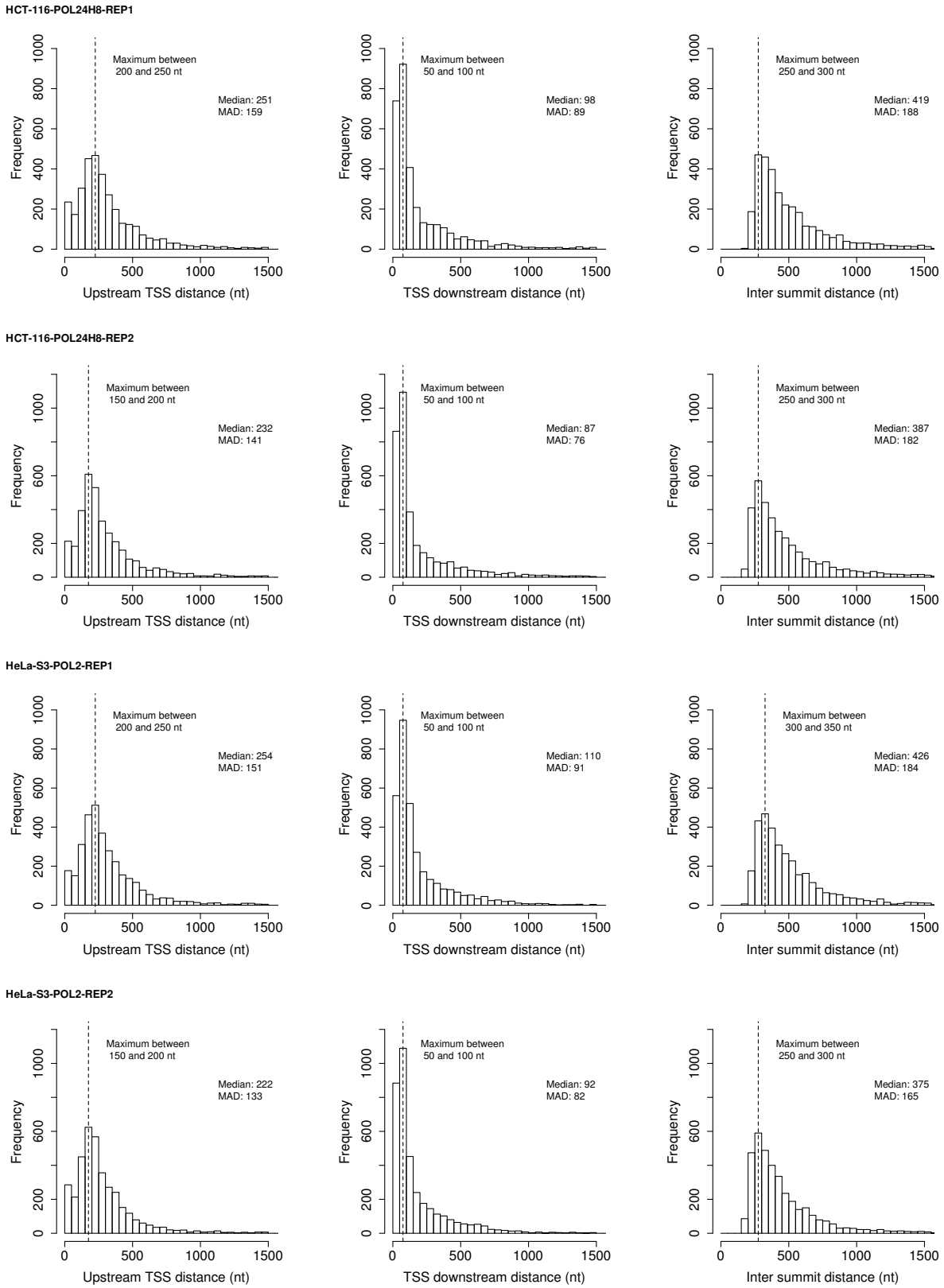

Figure S9. Distance distributions of TSS flanking double summits for RNAPII. Each row corresponds to one analyzed RNAPII dataset for the cell lines HCT-116 and HeLa-S3. The plots are based on the same data as the first column (Q) in Figure S8. The distributions for three different distances are shown. Column 1: The distance between the upstream summit and the TSS; Column 2: the distance between the TSS and the downstream summit; Column 3: the distance between the up- and downstream summits.

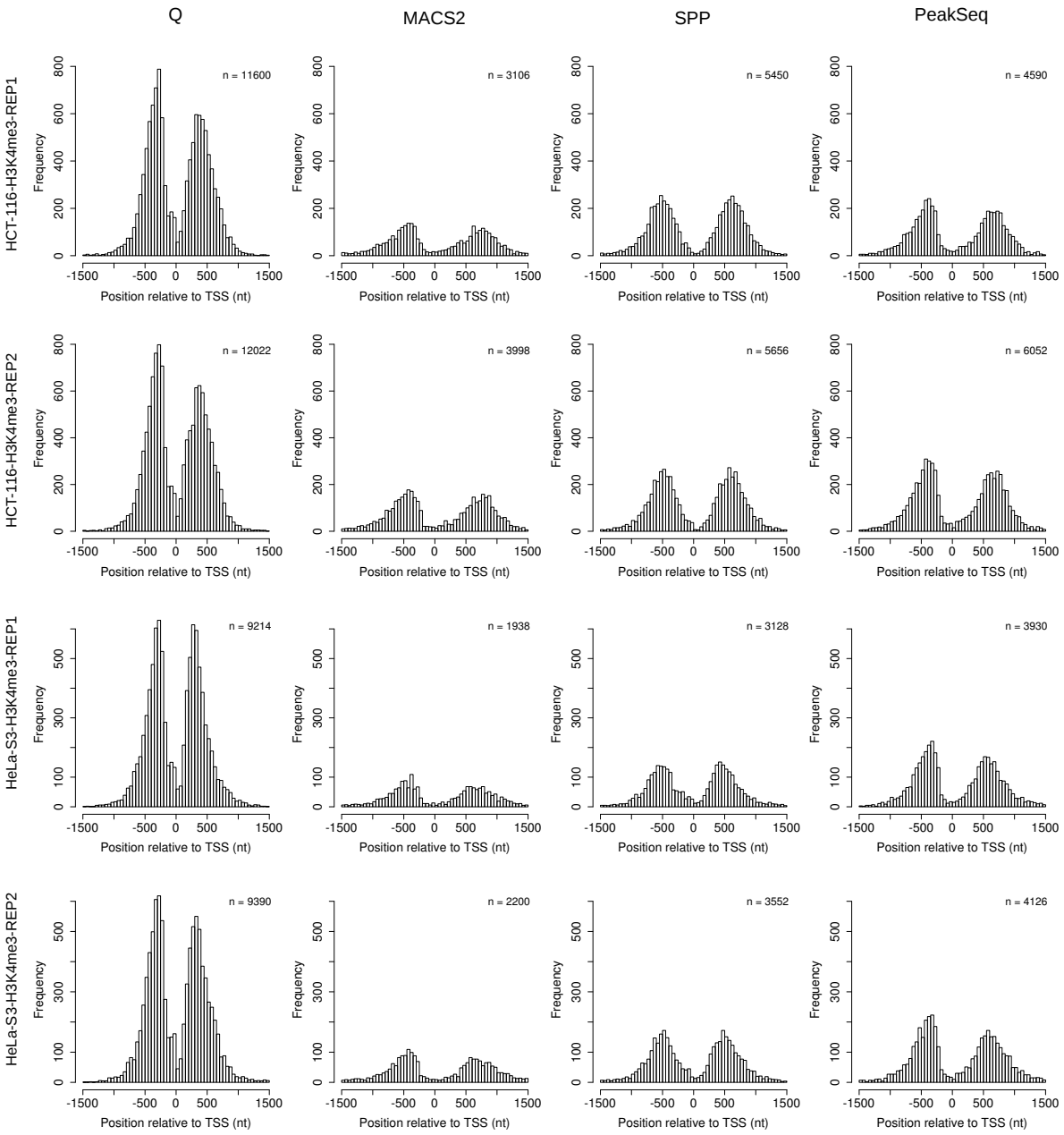

Figure S10. Distribution of TSS flanking double summits for H3K4me3. See the legend to Figure S8 for explanations. The subplot in the upper left is identical with Figure 6B in the main text.

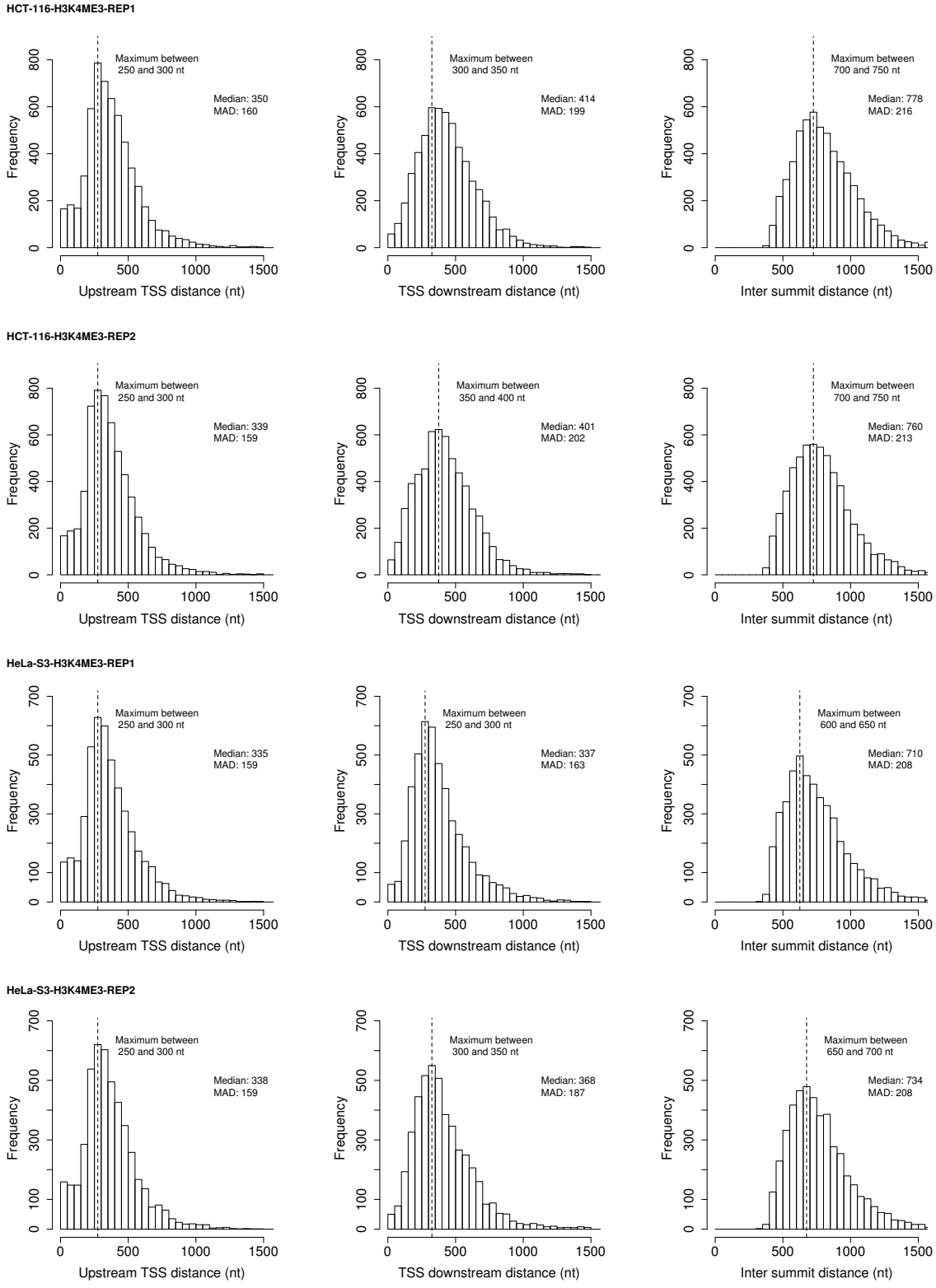

Figure S11. Distance distributions of TSS flanking double summits for H3K4me3. See the legend to Figure S10 for explanations. The plots are based on the same data as the first column (Q) in Figure S10.

Table S1. ChIP-seq datasets								
n	Cell line	Protein	Replicate	GEO	labExp ID	Control ID	# Hits Treatment	# Hits Control
1	GM12878	BATF	1	GSM803538	SL839	SL516	16,116,534	88,344,819
2	GM12878	BATF	2	GSM803538	SL985	SL517	16,971,028	26,781,377
3	GM12878	ETS1	1	GSM803510	SL1507	SL1613	14,140,724	26,426,487
4	GM12878	ETS1	2	GSM803510	SL1655	SL1613	18,811,239	26,426,487
5	GM12878	MEF2A	1	GSM803511	SL1425	SL1613	11,365,961	26,426,487
6	GM12878	MEF2A	2	GSM803511	SL1792	SL1613	17,549,982	26,426,487
7	GM12878	PAX5 (C20)	1	GSM803391	SL675	SL678	32,321,445	31,274,869
8	GM12878	PAX5 (C20)	2	GSM803391	SL735	SL517	13,227,334	26,781,377
9	GM12878	PAX5 (N19)	1	GSM803362	SL677	SL678	12,902,970	31,274,869
10	GM12878	PAX5 (N19)	2	GSM803362	SL848	SL516	23,889,850	88,344,819
11	GM12878	PU.1	1	GSM803531	SL612	SL516	9,666,126	88,344,819
12	GM12878	PU.1	2	GSM803531	SL649	SL516	23,357,263	88,344,819
13	H1-hESC	EGR1	1	GSM803430	SL1482	SL1398	23,961,734	63,913,543
14	H1-hESC	EGR1	2	GSM803430	SL1885	SL1398	23,990,931	63,913,543
15	H1-hESC	TAF1	1	GSM803450	SL853	SL969	13,511,515	63,913,543
16	H1-hESC	TAF1	2	GSM803450	SL964	SL969	12,530,448	63,913,543
17	H1-hESC	USF1	1	GSM803426	SL1159	SL1398	14,927,131	63,913,543
18	H1-hESC	USF1	2	GSM803426	SL1319	SL1398	13,135,507	63,913,543
19	HCT-116	POL2 (4H8)	1	GSM803474	SL3456	SL3457	16,290,358	28,967,090
20	HCT-116	POL2 (4H8)	2	GSM803474	SL3830	SL3457	27,110,423	28,967,090
21	HeLa-S3	POL2	1	GSM803533	SL631	SL592	21,569,982	29,454,439
22	HeLa-S3	POL2	2	GSM803533	SL672	SL592	23,494,468	29,454,439
23	HepG2	FOXA2	1	GSM803403	SL2196	SL1781	21,403,947	31,364,318
24	HepG2	FOXA2	2	GSM803403	SL3175	SL1781	32,793,338	31,364,318
25	HepG2	RAD21	1	GSM803517	SL3179	SL1781	22,251,303	31,364,318
26	HepG2	RAD21	2	GSM803517	SL3586	SL1781	24,232,455	31,364,318
27	HepG2	SIN3A	1	GSM803530	SL583	SL593	13,366,326	19,754,550
28	HepG2	SIN3A	2	GSM803530	SL637	SL593	26,093,615	19,754,550
29	K562	EGR1	1	GSM803414	SL3164	SL2455	17,769,679	11,573,554
30	K562	EGR1	2	GSM803414	SL3497	SL2455	30,990,458	11,573,554
31	K562	MAX	1	GSM803523	SL2945	SL1396	17,109,585	64,876,189
32	K562	MAX	2	GSM803523	SL3070	SL1397	30,518,637	24,507,401
33	K562	NRSF	1	GSM803440	SL3821	SL2455	20,648,251	11,573,554
34	K562	NRSF	2	GSM803440	SL3822	SL2455	18,684,487	11,573,554
35	K562	ZBTB7A	1	GSM803473	SL2265	SL2455	15,848,419	11,573,554
36	K562	ZBTB7A	2	GSM803473	SL3183	SL2455	23,721,201	11,573,554
37	SK-N-SH_RA	P300	1	GSM803495	SL2910	SL2905	20,343,215	26,169,645
38	SK-N-SH_RA	P300	2	GSM803495	SL2918	SL2905	14,389,846	26,169,645
39	HCT-116	H3K4me3	1	-	DS16056	EH000950	22,956,881	21,115,507
40	HCT-116	H3K4me3	2	-	DS16055	EH000950	26,193,423	21,115,507
41	HeLa-S3	H3K4me3	1	-	DS11555	EH000469	11,927,832	14,921,520
42	HeLa-S3	H3K4me3	2	-	DS11553	EH000469	14,741,696	14,921,520

Table S1. Datasets used for validation. The first column shows the number used to specify the dataset in the plots of the main manuscript. Other columns show the cell line, the protein of interest, the replicate number, the NCBI GEO sample accession number, and the ID for the control dataset used in the experiment. The final two columns show the numbers of non-redundant reads for treatment and control. The datasets can be downloaded from http://genome.ucsc.edu/cgi-bin/hgFileSearch?db=hg19.

Table S2. Runtime - Fragment length estimation				
CPU	**Remove dup.**	**N-Fold improvement**	**SPP (m)**	**Q (m)**
AMD	no	3.61	138.76	44.25
AMD	yes	4.74	144.69	31.33
Intel	no	3.12	45.56	15.28
Intel	yes	3.09	45.12	15.11

Table S2. Runtime comparison: Pearson cross-correlation vs. Hamming distance. A total of 38 datasets were analysed using SPP [4] to determine the Pearson cross-correlation or with Q to determine the Hamming distance. The calculations were performed on a desktop computer with an Intel Core i7-3770 (3.4 GHz) processor and on a rack server with an AMD Opteron(tm) Processor 6172 (2.1 GHz). The average time in minutes (user time plus system time) is indicated together with the fold improvement for Q as compared to SPP.

Table S3 Peak overlaps					
Dataset	Q	MACS2	SPP	PeakSeq	Row mean
1	35,088	32,243	30,115	30,985	32,108
2	62,482	58,289	54,759	58,082	58,403
3	7,669	7,777	6,020	5,991	6,864
4	16,593	8,710	10,216	7,148	10,667
5	39,068	32,197	33,175	22,185	31,656
6	28,765	18,387	21,954	13,163	20,567
7	42,693	36,811	33,569	50,779	40,963
8	39,465	34,701	34,302	28,339	34,202
9	27,561	23,651	22,937	22,382	24,133
10	49,868	44,028	37,757	50,623	45,569
11	33,478	31,049	31,220	21,791	29,385
12	66,363	62,799	58,860	65,271	63,323
13	14,042	14,238	10,943	10,746	12,492
14	25,319	23,621	19,170	20,426	22,134
15	41,773	26,436	29,595	34,991	33,199
16	34,381	22,936	26,270	25,423	27,253
17	38,490	35,323	32,099	40,151	36,516
18	43,357	41,194	38,516	47,853	42,730
19	53,590	43,421	43,412	42,412	45,709
20	58,403	49,107	46,718	46,563	50,198
21	60,450	46,976	45,759	45,022	49,552
22	61,604	52,266	49,064	52,484	53,855
23	22,957	22,760	22,828	18,563	21,777
24	69,342	65,508	62,726	60,968	64,636
25	66,398	63,299	61,613	63,657	63,742
26	62,509	57,581	57,337	58,989	59,104
27	25,021	12,775	16,584	8,425	15,701
28	29,204	19,952	25,189	15,290	22,409
29	48,601	42,913	41,310	43,409	44,058
30	29,865	25,972	23,902	20,327	25,017
31	56,221	47,385	49,618	57,324	52,637
32	70,663	63,311	58,838	69,232	65,511
33	32,775	29,748	28,702	23,882	28,777
34	23,677	21,270	20,009	14,088	19,761
35	22,326	17,030	17,558	9,811	16,681
36	42,189	34,611	35,586	37,417	37,451
37	66,506	62,569	60,383	65,971	63,857
38	62,374	61,369	56,556	62,212	60,628

Table S3. Peak overlap. The number of peaks among the 100,000 top ranked peaks of pseudoreplicates that overlap. This data is visualized as a radarplot in Figure 4A of the main manuscript. Mean normalized values in Figure 4B were derived from this table by subtracting the row mean (shown in the last column) for each individual experiment.

Table S4 Pearson correlation coefficients				
Dataset	Q	MACS2	SPP	PeakSeq
1	0.94	0.85	0.89	0.91
2	0.96	0.89	0.93	0.92
3	0.95	0.90	0.93	0.97
4	0.83	0.71	0.76	0.88
5	0.90	0.79	0.82	0.85
6	0.91	0.83	0.85	0.89
7	0.90	0.80	0.81	0.80
8	0.93	0.85	0.87	0.85
9	0.89	0.79	0.80	0.86
10	0.93	0.86	0.86	0.84
11	0.95	0.88	0.93	0.95
12	0.96	0.92	0.95	0.94
13	0.82	0.70	0.72	0.88
14	0.91	0.83	0.84	0.84
15	0.96	0.87	0.88	0.92
16	0.95	0.85	0.86	0.90
17	0.96	0.90	0.95	0.96
18	0.97	0.92	0.96	0.95
19	0.95	0.84	0.86	0.88
20	0.94	0.83	0.82	0.87
21	0.97	0.88	0.88	0.90
22	0.96	0.87	0.86	0.91
23	0.88	0.75	0.75	0.83
24	0.95	0.88	0.92	0.93
25	0.96	0.88	0.95	0.92
26	0.95	0.87	0.94	0.92
27	0.78	0.66	0.65	0.79
28	0.84	0.67	0.60	0.74
29	0.94	0.85	0.89	0.90
30	0.87	0.74	0.78	0.84
31	0.93	0.84	0.84	0.87
32	0.95	0.89	0.85	0.92
33	0.96	0.87	0.93	0.92
34	0.98	0.90	0.96	0.96
35	0.80	0.68	0.69	0.82
36	0.87	0.73	0.67	0.82
37	0.94	0.85	0.84	0.93
38	0.93	0.84	0.85	0.92

Table S4. Pearson correlation coefficients for the signal values of peak overlaps of the $100,000$ top ranked peaks of pseudoreplicates (Table S3). P-values relative to Q were calculated using two-sample two-sided Wilcoxon tests: Q vs. MACS2: 2.29×10^{-8}; Q vs. SPP: 3.45×10^{-5}; Q vs. PeakSeq: 1.97×10^{-3}.

Table S5 CWOP Analysis				
Dataset	**Q**	**MACS2**	**SPP**	**PeakSeq**
1	0.00	0.02	0.00	0.03
2	0.01	0.03	0.01	0.00
3	0.00	0.00	0.00	**1.00**
4	0.00	**0.72**	**0.55**	**1.00**
5	0.00	0.09	0.08	**0.39**
6	0.00	**0.43**	0.02	**0.70**
7	0.00	0.04	0.01	0.00
8	0.00	**0.16**	0.00	**0.16**
9	0.00	**0.28**	0.05	**0.11**
10	0.00	0.03	0.00	0.00
11	0.00	0.00	0.00	**0.21**
12	0.00	0.00	0.00	0.00
13	**0.20**	**0.54**	0.00	**0.91**
14	0.01	**0.12**	0.02	**0.37**
15	0.00	0.04	0.00	0.01
16	0.00	0.08	0.01	0.07
17	0.00	0.04	0.01	0.00
18	0.00	0.03	0.00	0.00
19	0.00	0.02	0.00	0.00
20	0.00	0.01	0.00	0.00
21	0.01	0.02	0.00	0.00
22	0.00	0.00	0.00	0.00
23	0.00	**0.40**	0.08	0.04
24	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
26	0.00	0.01	0.00	0.00
27	0.00	**0.68**	**0.57**	0.00
28	0.00	**0.31**	0.01	**0.42**
29	0.00	0.06	0.00	0.00
30	0.09	**0.16**	0.09	**0.34**
31	0.00	0.01	0.01	0.00
32	0.00	0.00	0.00	0.00
33	0.00	0.06	0.04	**0.22**
34	0.09	**0.22**	0.03	**0.72**
35	0.04	**0.41**	**0.24**	**1.00**
36	0.00	**0.15**	0.03	0.01
37	0.00	0.00	0.00	0.00
38	0.00	0.01	0.00	0.00
Number of datasets with CWOPs > 0	7	30	18	19
Number of datasets with CWOPs > 0.1	1	13	3	14

Table S5. CWOP Analysis. Proportions of low quality peaks with IDR$\leq$ 0.01. Low quality peaks are here defined as overlapping peaks for which both pseudoreplicates signal have scores lower than the mean of the signal scores with IDR$>$ 0.01. Those with a proportion greater than 0.10 are shown in bold.

Table S6 Peak overlaps for peaks with IDR≤ 0.01					
Dataset	Q	MACS2	SPP	PeakSeq	Row mean
1	11,944	11,283	9,092	10,665	10,746
2	24,489	21,474	17,953	20,438	21,089
3	1,106	805	781	5,320	2,003
4	2,299	1,981	2,311	4,176	2,692
5	7,792	5,040	5,209	7,009	6,263
6	5,362	5,399	3,743	4,940	4,861
7	8,775	5,683	4,692	13,451	8,150
8	7,191	6,077	4,383	5,787	5,860
9	5,192	4,049	2,971	4,571	4,196
10	12,081	10,194	6,385	16,681	11,335
11	12,212	11,276	10,580	9,904	10,993
12	28,875	26,834	24,316	26,241	26,567
13	1,828	1,774	1,238	3,901	2,185
14	4,926	4,115	2,858	5,513	4,353
15	18,028	13,194	9,715	15,878	14,204
16	13,236	10,507	7,011	13,154	10,977
17	11,730	10,148	9,719	11,229	10,707
18	13,524	12,070	11,414	13,239	12,562
19	23,543	17,470	14,429	17,600	18,261
20	27,323	20,751	14,609	18,849	20,383
21	27,284	17,015	11,618	16,318	18,059
22	31,244	20,185	15,929	19,863	21,805
23	5,057	5,252	3,214	6,250	4,943
24	29,834	26,783	24,873	22,794	26,071
25	35,087	34,742	32,681	30,594	33,276
26	33,643	32,492	30,630	30,400	31,791
27	3,133	2,078	2,074	5,409	3,174
28	8,599	3,308	1,754	3,722	4,346
29	18,977	15,581	14,492	17,566	16,654
30	11,471	8,396	7,178	8,362	8,852
31	19,227	21,301	11,181	22,461	18,543
32	32,006	26,608	19,178	30,035	26,957
33	7,279	6,338	5,578	6,216	6,353
34	5,762	5,651	3,972	6,126	5,378
35	3,134	2,539	1,708	3,824	2,801
36	13,238	10,270	5,250	15,756	11,129
37	23,824	19,079	16,394	21,118	20,104
38	19,059	14,964	13,564	17,777	16,341

Table S6. Peak overlaps for peaks with IDR≤ 0.01. Counts of peaks among the 100,000 top ranked peaks of pseudoreplicates (Table S3) with IDR≤ 0.01. The same data is presented as a radarplot in the main manuscript as Fig. 4D, where all datasets for which one or more peak caller had 10% or more CWOPs (Table S5) were excluded. Mean normalized values in Figure 4E were derived from this table by subtracting the row mean for each individual experiment.

Table S7 Motif content analysis					
Dataset	Q	MACS2	SPP	PeakSeq	Row Mean
1	13,556	13,745	12,864	12,430	13,149
2	19,279	19,372	18,972	19,039	19,166
3	6,036	5,090	5,458	4,858	5,361
4	15,676	11,249	14,148	8,167	12,310
5	18,973	17,528	17,967	16,444	17,728
6	20,991	18,165	20,178	15,464	18,700
7	16,128	14,715	14,641	15,118	15,151
8	19,538	18,112	17,940	18,629	18,555
9	16,194	14,456	15,114	15,345	15,277
10	20,657	19,914	19,247	19,667	19,871
11	28,210	27,334	27,306	26,773	27,406
12	31,051	30,891	30,769	30,655	30,842
13	17,823	14,194	14,447	13,285	14,937
14	27,935	23,766	24,402	20,642	24,186
15	24,129	19,281	22,478	19,930	21,455
16	21,689	16,333	18,889	17,897	18,702
17	22,901	22,577	22,182	21,539	22,300
18	24,117	24,039	23,499	22,203	23,465
19	23,880	19,865	22,502	20,017	21,566
20	20,619	18,174	19,543	17,438	18,944
21	15,064	11,441	13,993	11,663	13,040
22	14,946	10,983	13,744	11,226	12,725
23	21,523	21,284	22,738	20,204	21,437
24	15,606	15,387	15,406	15,187	15,397
25	28,579	28,729	28,539	28,555	28,601
26	29,005	29,138	28,938	28,761	28,961
27	14,300	10,342	12,218	7,332	11,048
28	17,366	12,340	14,035	10,589	13,583
29	33,073	31,706	31,908	31,641	32,082
30	30,054	26,422	26,769	23,961	26,802
31	24,326	21,994	22,598	22,061	22,745
32	26,973	24,590	25,627	24,302	25,373
33	20,392	19,319	19,022	19,657	19,598
34	17,173	14,310	15,992	12,546	15,005
35	27,868	24,115	24,928	17,715	23,657
36	32,719	29,362	30,047	29,800	30,482
37	16,339	16,200	15,915	16,114	16,142
38	18,632	18,471	18,081	18,549	18,433

Table S7. Motif content analysis. Number of peaks amongst the 50,000 top ranked peaks containing at least one significant occurence of a reference motif, also visualized in Figure 5A The mean-normalized values in Figure 5B were derived from this table by subtracting the row mean for each individual experiment.

.

Table S8 TSS flanking double summits for RNAPII				
HCT-116-POL24H8-REP1	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	53,590	43,421	43,412	42,412
Summits in promoters	15,888	9,898	11,328	11,012
Promoters with $\geq$ 1 summit	8,275	8,560	8,306	8,401
Promoters with TFDS	39.49% (3,268)	10.44% (894)	19.09% (1,586)	16.87% (1,417)
HCT-116-POL24H8-REP2	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	58,403	49,107	46,718	46,563
Summits in promoters	16,284	10,096	11,206	10,968
Promoters with $\geq$ 1 summit	8,148	8,516	8,111	8,302
Promoters with TFDS	42.82% (3,489)	12.27% (1,045)	19.74% (1,601)	17.74% (1,473)
HeLa-S3-POL2-REP1	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	60,450	46,976	45,759	45,022
Summits in promoters	16,807	9,336	10,348	10,014
Promoters with $\geq$ 1 summit	7,376	7,662	7,376	7,462
Promoters with TFDS	44.47% (3,280)	14.00% (1,073)	20.31% (1,498)	18.32% (1,367)
HeLa-S3-POL2-REP2	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	61,604	52,266	49,064	52,484
Summits in promoters	18,324	9,632	11,329	10,993
Promoters with $\geq$ 1 summit	7,602	7,948	7,709	7,801
Promoters with TFDS	48.41% (3,680)	13.29% (1,056)	23.49% (1,811)	20.63% (1,609)

Table S8. TSS flanking double summits for RNAPII. We performed an analysis of TSS flanking double summits (TFDS) on four RNAPII datasets from the two different cell lines, HCT-116 and HeLa-S3. We analysed the summits of the overlapping peaks derived in the reproducibility analysis (Table S3). For each dataset and peak caller, we counted the total number of summits in promoters (TSS$\pm$1500), of promoters with at least one summit and promoters with TFDS. The percentages for promoters with TFDS are with respect to the number of promoters with at least one summit.

.

Table S9 TSS flanking double summits for H3K4me3				
HCT-116-H3K4ME3-REP1	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	40,707	28,858	28,106	25,851
Summits in promoters	17,635	10,398	11,479	10,519
Promoters with $\geq$ 1 summit	8,529	8,593	8,487	7,871
Promoters with TFDS	68.00% (5,800)	18.07% (1,553)	32.11% (2,725)	29.16% (2,295)
HCT-116-H3K4ME3-REP2	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	42,475	27,780	27,524	27,874
Summits in promoters	18,701	10,878	11,541	11,853
Promoters with $\geq$ 1 summit	(8,512)	(8,543)	(8,412)	(8,286)
Promoters with TFDS	70.62% (6,011)	23.40% (1,999)	33.62% (2,828)	36.52% (3,026)
HeLa-S3-H3K4ME3-REP1	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	35,397	19,609	22,551	19,467
Summits in promoters	16,160	8,826	9,730	9,830
Promoters with $\geq$ 1 summit	7,779	7,727	7,745	7,529
Promoters with TFDS	59.22% (4,607)	12.54% (969)	20.19% (1,564)	26.10% (1,965)
HeLa-S3-H3K4ME3-REP2	**Q**	**MACS2**	**SPP**	**PeakSeq**
Summits	36,158	21,484	23,516	19,758
Summits in promoters	16,081	8,971	9,918	9,894
Promoters with $\geq$ 1 summit	7,770	7,739	7,722	7,512
Promoters with TFDS	60.42% (4,695)	14.21% (1,100)	23.00% (1,776)	27.46% (2,063)

Table S9. TSS flanking double summits for H3K4me3. See the legend to Figure S8 for explanations.

.

Table S10. TFDS - Reproducibility			
	TFDS REP1	TFDS REP2	Overlap
HCT-116-POL24H8	3,268	3,489	2,552 (78.09%)
HeLa-S3-POL2	3,280	3,680	2,718 (82.87%)
HCT-116-H3K4ME3	5,800	6,011	5,255 (90.60%)
HeLa-S3-H3K4ME3	4,607	4,695	3,928 (85.26%)

Table S10. TFDS - Reproducibility. For biological replicates we determined promoters (TSS±1500) with TFDS identified by Q for both replicates using bedtools [5].

Table S11. TFDS - Simulation study						
Cell line	RNAPII	H3K4me3	Overlap	Pattern observed	Simulated avg	P-value
HCT-116	POL24H8-R1	H3K4ME3-R1	2,790	1,807 (64.77%)	1,536 (55.05%)	$< 10^{-4}$
HCT-116	POL24H8-R1	H3K4ME3-R2	2,833	1,798 (63.47%)	1,515 (53.48%)	$< 10^{-4}$
HCT-116	POL24H8-R2	H3K4ME3-R1	2,998	2,041 (68.08%)	1,761 (58.74%)	$< 10^{-4}$
HCT-116	POL24H8-R2	H3K4ME3-R2	3,037	2,016 (66.38%)	1,722 (56.70%)	$< 10^{-4}$
HeLa-S3	POL2-R1	H3K4ME3-R1	2,578	1,648 (63.93%)	1,322 (51.28%)	$< 10^{-4}$
HeLa-S3	POL2-R1	H3K4ME3-R2	2,615	1,703 (65.12%)	1,386 (53.00%)	$< 10^{-4}$
HeLa-S3	POL2-R2	H3K4ME3-R1	2,837	1,939 (68.35%)	1,628 (57.38%)	$< 10^{-4}$
HeLa-S3	POL2-R2	H3K4ME3-R2	2,904	1,995 (68.70%)	1,697 (58.44%)	$< 10^{-4}$

Table S11. TFDS - Simulation study. We performed simulations for all pairs of replicates within one cell line. The numbers in the column *Overlap* refer to promoter overlaps, i.e. those promoters having a TFDS for RNAPII and H3K4me3. By the definition of TFDS also the corresponding intersummit regions must overlap. The numbers in the column *Pattern observed* refer to the number of cases for which the intersummit region of the TFDS for RNAPII is completely contained in that of the TFDS of H3K4me3. In the column *Simulated avg* the average observed number of patterns for $10,000$ simulations are shown. In each simulation we randomly shuffled the intersummit regions of the TFDS for H3K4me3 and determined the how often the pattern occured. Percentages refer to the corresponding number of overlapping promoters (column *Overlap*). The last column shows upper bounds for the empirical P-value.

Table S12. Runtime - Peak calling			
Q (m)	MACS2 (m)	SPP (m)	PeakSeq (m)
2.06	10.92	38.11	7.27

Table S12. Runtime comparison - Peak calling. A total of 38 datasets were analysed using Q, MACS2, SPP and PeakSeq. The calculations were performed on a rack server with an AMD Opteron(tm) Processor 6172 (2.1 GHz). The average runtimes are given in minutes (user time plus system time).

References (Online Supplementary Material)

[1] Qunhua Li, James Brown, Haiyan Huang, and Peter Bickel. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, 5(3):1752–1779, 2011.

[2] Yajie Yang, Justin Fear, Jianhong Hu, Irina Haecker, Lei Zhou, Rolf Renne, David Bloom, and Lauren M McIntyre. Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Computational and structural biotechnology journal*, 9(13):e201401002, January 2014.

[3] Daniel M. Ibrahim, Peter Hansen, Christian Rödelsperger, Asita C. Stiege, Sandra C. Doelken, Denise Horn, Marten Jäger, Catrin Janetzki, Peter Krawitz, Gundula Leschik, Florian Wagner, Till Scheuer, Mareen Schmidt-von Kegler, Petra Seemann, Bernd Timmermann, Peter N. Robinson, Stefan Mundlos, and Jochen Hecht. Distinct global shifts in genomic binding profiles of limb malformation-associated hoxd13 mutations. *Genome Res*, 23(12):2091–2102, Dec 2013.

[4] Peter V. Kharchenko, Michael Y. Tolstorukov, and Peter J. Park. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*, 26(12):1351–1359, Dec 2008.

[5] Aaron R Quinlan and Ira M Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841–2, March 2010.