Supplementary material for "Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation"

Authors: Alexis Vandenbon[1$]*, Yutaro Kumagai[2$], Mengjie Lin[3], Yutaka Suzuki[3], Kenta Nakai[4]*

[1] Immuno-Genomics Research Unit, Immunology Frontier Research Center (IFReC), Osaka University, Suita, 565-0871, Japan

[2] Quantitative Immunology Research Unit, Immunology Frontier Research Center (IFReC), Osaka University, Suita, 565-0871, Japan

[3]Department of Computational Biology and Medical Sciences, Graduate School of

Frontier Sciences, The University of Tokyo, Kashiwa 277-8561, Japan

[4] Laboratory of Functional Analysis in silico, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, 108-8639, Japan

$ Equal contribution

* To whom correspondence should be addressed. Email: alexisvdb@ifrec.osaka-u.ac.jp, knakai@ims.u-tokyo.ac.jp

Table of Contents

Supplementary Methods

Deciding the main TSS of promoter regions

In many TSS-seq based promoter regions, there are several candidate TSSs (bases with aligned TSS-seq reads). We used a log-likelihood model for deciding the representative TSS for each promoter region, inspired by position-specific scoring matrices used for modeling binding specificities of TFs (Stormo, 2000). In brief, we constructed a model capturing properties that are typical for TSSs (as compared to randomly selected loci), and use the model to score candidate TSSs. For each promoter with multiple candidate TSSs, the TSS with the highest score for our model was selected. More details are explained below.

For each biological feature f (here: histone modifications, Pol2 binding, RNA-seq reads, and TSS-seq reads) in our dataset, we averaged aligned read counts (ppm; reads per million reads) over the ten time points. Next, we calculated $ppm_{x,f,i}$, being the number of reads (in ppm) around the main TSS for each promoter x (i.e. the base with the highest number of aligned TSS-seq reads) in bin i, for 20 bins of 100 bps over the region -1 kb to + 1kb around the main TSS.

From the ppm values in all bins, we generated a position probability matrix M_f for each feature f, as follows:

$$M_f[i,j] = \frac{1}{n}\sum_{x \in P} I\left(ppm_{x,f,j} \in bin_i\right) \quad [1]$$

where $M_f[i,j]$ is the entry on row i and column j in M_f, n is the number of promoter regions in the genome-wide set of TSS-seq derived promoters P, and $I\left(ppm_{x,f,j} \in bin_i\right)$ is 1 if the value $ppm_{x,f,j}$ is within the range specified by bin_i (otherwise it is 0). To decide the ranges of bin_i, for each feature f, six equally populous bins were decided based on all $ppm_{x,f,j}$ values.

M_f captures the typical signal for feature f around TSSs. We similarly generated a control, M_f^{rand}, based on 10,000 randomly selected genomic regions, using the same bin values as used for the TSS-seq derived promoters. From M_f and M_f^{rand}, a position weight matrix was generated as follows:

$$PWM_f = ln\left(M_f \Big/ M_f^{rand}\right) \quad [2]$$

This PWM_f represents a log-likelihood model for feature f in promoters as compared to control sequences (randomly selected regions). Using these models, for each promoter region x, we scored all candidate TSSs t by summing over the relevant bin for each position j (the 20 bins of 100 bps):

$$score_{t,f} = \sum_{j=1}^{20} PWM_f\,[i,j] \quad [3]$$

where i is the index of bin_i range corresponding to the value of $ppm_{x,f,j}$. Thus we obtained a total score over all features f as follows:

$$score_{t,total} = \sum_{f \in features} score_{t,f} \quad [4]$$

where $features$ include all histone modifications, Pol2 binding, RNA-seq reads, and TSS-seq reads.

The final representative TSS $\hat{t}$ of promoter region x was defined as the TSS having the highest total score. In practice, for each promoter we considered at most 10 bases with the highest count of aligned TSS-seq reads as candidate TSSs. Moreover, candidate TSSs with few aligned TSS-seq reads (less than one tenth of the number of TSS-seq reads aligned to the base with the highest number of reads) were not considered. The thus decided representative TSSs were used as the "center" of each promoter for the downstream analysis of increases in histone modifications (see main manuscript).

In addition to deciding the representative TSS for each promoter region, this model was also used for removing low scoring promoters (e.g. regions with only a TSS-seq signal, lacking Pol2 binding signals, lacking histone modifications typical of promoters, etc). For this, we removed promoters with total scores < 0.

Finally, for the assignment of main promoters to genes, if a gene had multiple candidate promoter regions, we assigned to it the one with the highest total score for our model.

Primer sequences

Gene (forward/reverse)	**Primer sequence**
Actb (fw)	GGGAATGGGTCAGAAGGACT
Actb (rv)	CTTCTCCATGTCGTCCCAGT
Il6 (fw)	AGTTGCCTTCTTGGGACTGA
Il6 (rv)	ACAGGTCTGTTGGGAGTGGT
Tnf (fw)	CCCCAAAGGGATGAGAAGTT
Tnf (rv)	CACTTGGTGGTTTGCTACGA
Il1b (fw)	TGAAGCAGCTATGGCAACTG
Il1b (rv)	GGGTCCGTCAACTTCAAAGA
Cxcl1 (fw)	ACTCCAACACAGCACCATGA
Cxcl1 (rv)	ATGGTCTGCAGGCACTGAC
Cxcl10 (fw)	AAGTGCTGCCGTCATTTTCT
Cxcl10 (rv)	CCTATGGCCCTCATTCTCAC
Ifit1 (fw)	CAAGGCAGGTTTCTGAGGAG
Ifit1 (rv)	CATTCTCTCCCATGGTTGCT
Rsad2 (fw)	ACAGCCAAGACATCCTTCGT
Rsad2 (rv)	TCTTCTCCAAACCAGCCTGT
Nfkbiz (fw)	AGAAAGGGACCCGATCCTC
Nfkbiz (rv)	CGGTGATGTCACGAAGTGAG
Ccl5 (fw)	CGAGGGAGAGGTAGGCAAAG
Ccl5 (rv)	TCACCATCATCCTACCTGCA

Table S1: Table of primer sequences used for RT-qPCR.

Gene (forward/reverse)	**Primer sequence**
Tnf (fw)	GTGCCTATGTCTCAGCCTCT
Tnf (rv)	CCAGACACTCACCTCATCCC
Il1b (fw)	CACTGATGGACTTTGGGCTT
Il1b (rv)	TGTCAGCGGCTATACAGACA
Cxcl1 (fw)	CTATCGCCAATGAGCTGCG

Cxcl1 (rv)	GACTTCGGTTTGGGTGCAG
Nfkbiz (fw)	CGGCGAGCTCTAGAGAAAGA
Nfkbiz (rv)	CCCCAAGTACGTGAGAGCAT
Ifit1 (fw)	CAAGGCAGGTTTCTGAGGAG
Ifit1 (rv)	CCCTCAGAGTGGAGAACAGG
Rsad2 (fw)	AGCAGCCGAGCAGCTAGAG
Rsad2 (rv)	ATAAGCCCTTACAGGCAGCA
Cxcl10 (fw)	CACATGACCATTTCATGTCAGTT
Cxcl10 (rv)	AAAACCGTCCAATACCTTTTGT
Ccl5 (fw)	ACCTGCCTCACCATGTAAGT
Ccl5 (rv)	AGAAGGGGAGGTCTGGGTAT
Il6 (fw)	GAGGAGTGTGAGGCAGAGAG
Il6 (rv)	CTGCGTGGAGAAAAGGGAAA

Table S2: Table of primer sequences used for ChIP-qPCR.

Supplementary Figures

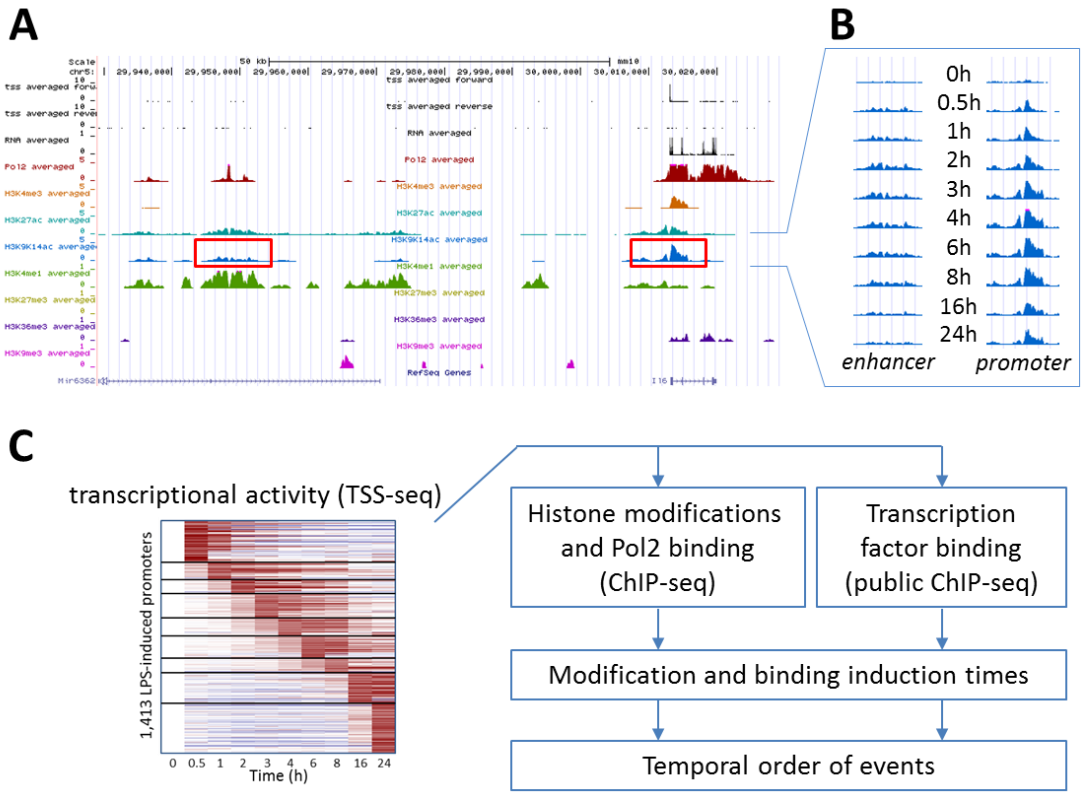


Fig. S1: Overview of data and analysis. (A). As an illustration of our dataset, averaged signals are shown for all features included in this study in the locus including the gene *Il6* and several potential enhancer regions upstream of *Il6*. (B) Detailed view of H3K9K14ac signals over time after LPS stimulation, around the Il6 promoter region and an upstream enhancer region. (C) Focusing on a set of 1,413 LPS-induced promoters, we defined significant induction of histone modifications, Pol2 binding, and TF binding over time. We analyzed the ordering of induction times of histone modifications and binding events, and generated hypotheses regarding underlying molecular mechanisms.

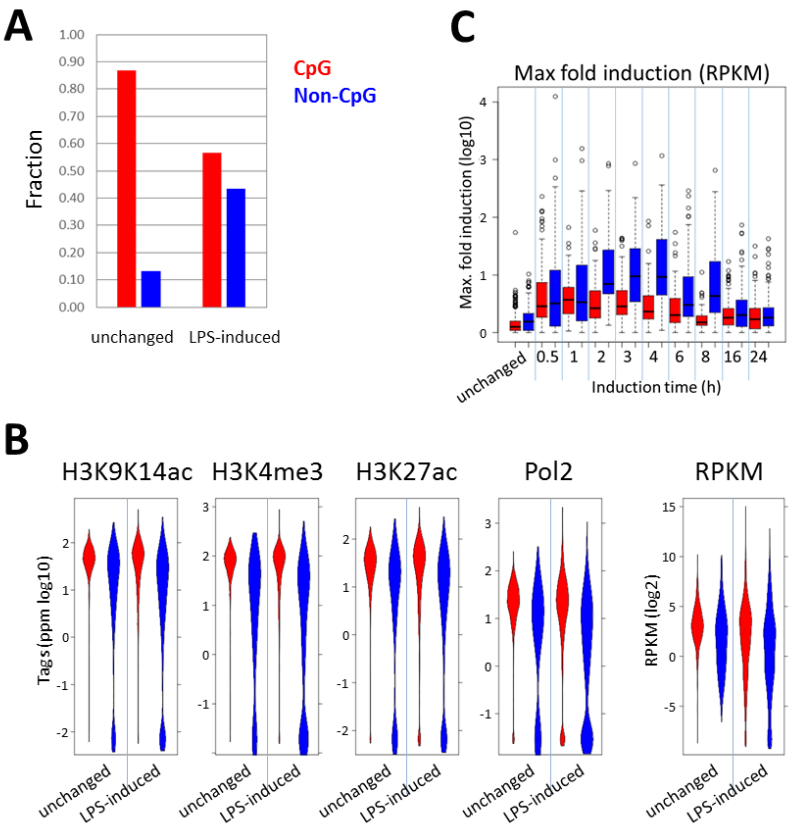
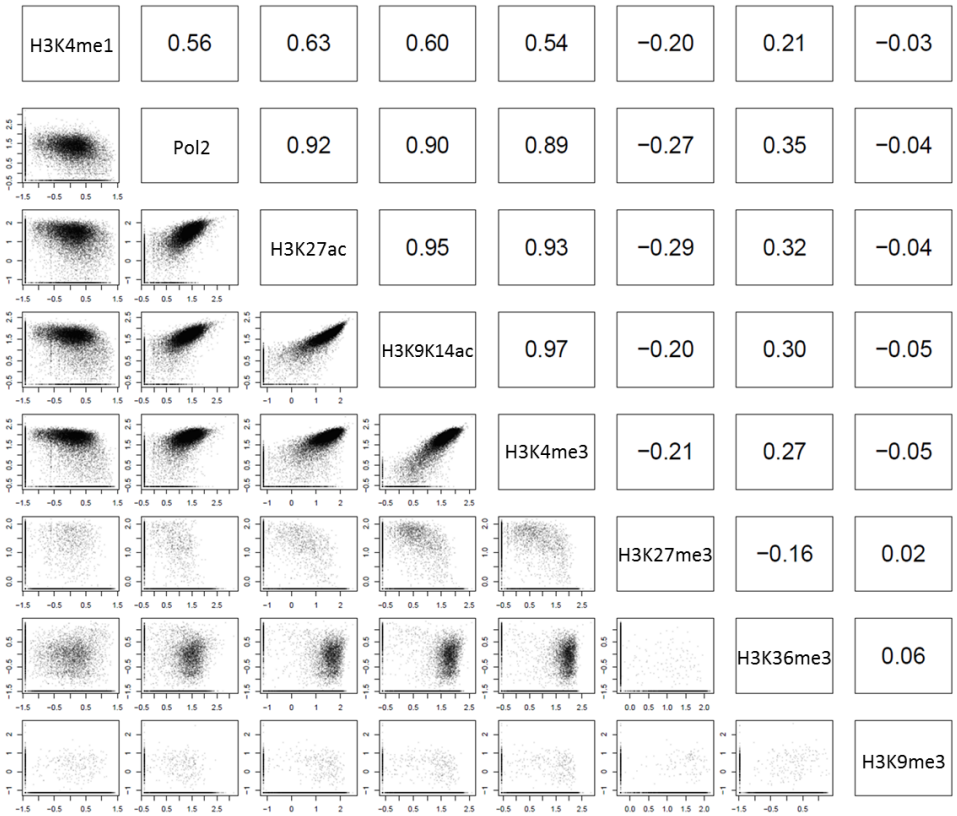


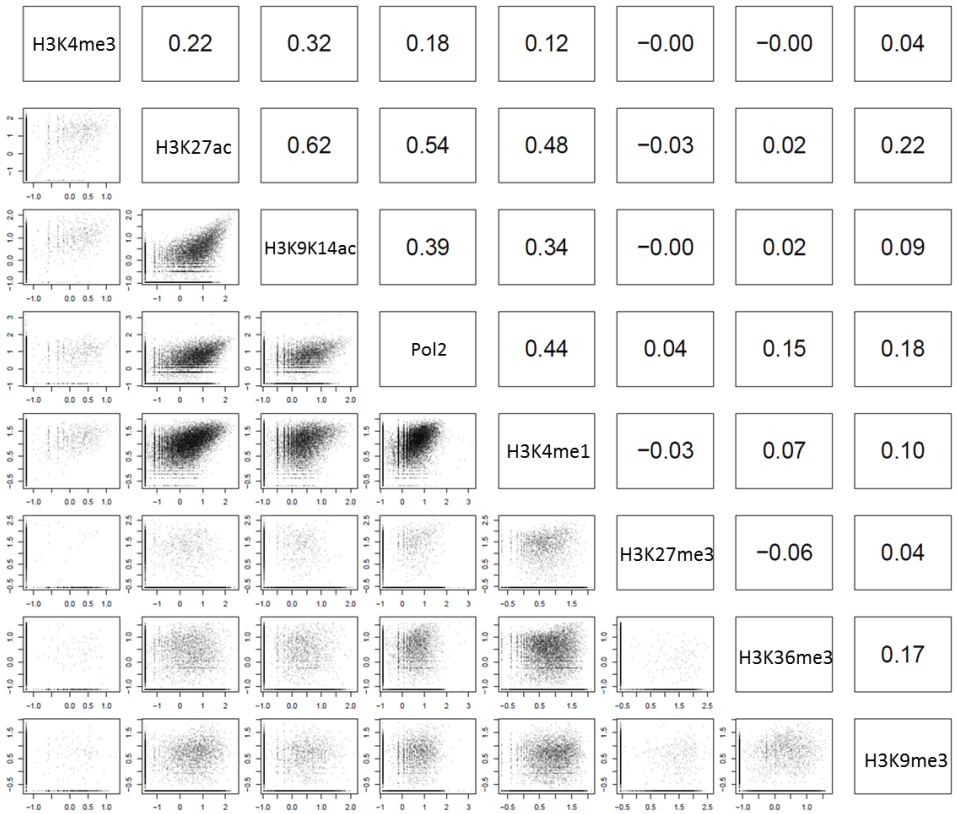
Fig. S2: Basal properties of LPS-induced CpG and non-CpG promoters. (A) Fraction of CpG-associated (red) and non-CpG (blue) promoters per induction time point, and for stable expressed promoters ("unchanged"). (B) Basal (at time 0h) properties of LPS-induced and stably expressed ("unchanged") promoters. A distinction is made between CpG-associated (red) and non-CpG (blue) promoters. Violin plots are shown for basal levels of H3K9K14ac, H3K4me3, H3K27ac, Pol2 binding, and gene expression (RPKMS). (C) Boxplot showing maximum fold-induction of gene expression (RPKM-based) compared to 0h.

Fig. S3: (next page) Correlations between basal features at promoters **(A)** and enhancers **(B)**. Features are named on the diagonal. The upper part shows the Pearson correlation coefficient between each pair of features, and the lower part plots pairs of features against each other (ppm in log scale). A small pseudocount was added to avoid problems with 0 values in log scales. At promoters, high correlations were seen especially between Pol2, H3K27ac, H3K9K14ac, and H3K4me3. At enhancers, basal correlations were in general less strong, but relatively high correlations were seen between H3K27ac and H3K9K14ac, Pol2 binding and H3K4me1.

A


H3K4me1
0.56
0.63
0.60
0.54
−0.20
0.21
−0.03
Pol2
0.92
0.90
0.89
−0.27
0.35
−0.04
H3K27ac
0.95
0.93
−0.29
0.32
−0.04
H3K9K14ac
0.97
−0.20
0.30
−0.05
H3K4me3
−0.21
0.27
−0.05
H3K27me3
−0.16
0.02
H3K36me3
0.06
H3K9me3

B


H3K4me3
0.22
0.32
0.18
0.12
−0.00
−0.00
0.04
H3K27ac
0.62
0.54
0.48
−0.03
0.02
0.22
H3K9K14ac
0.39
0.34
−0.00
0.02
0.09
Pol2
0.44
0.04
0.15
0.18
H3K4me1
−0.03
0.07
0.10
H3K27me3
−0.06
0.04
H3K36me3
0.17
H3K9me3

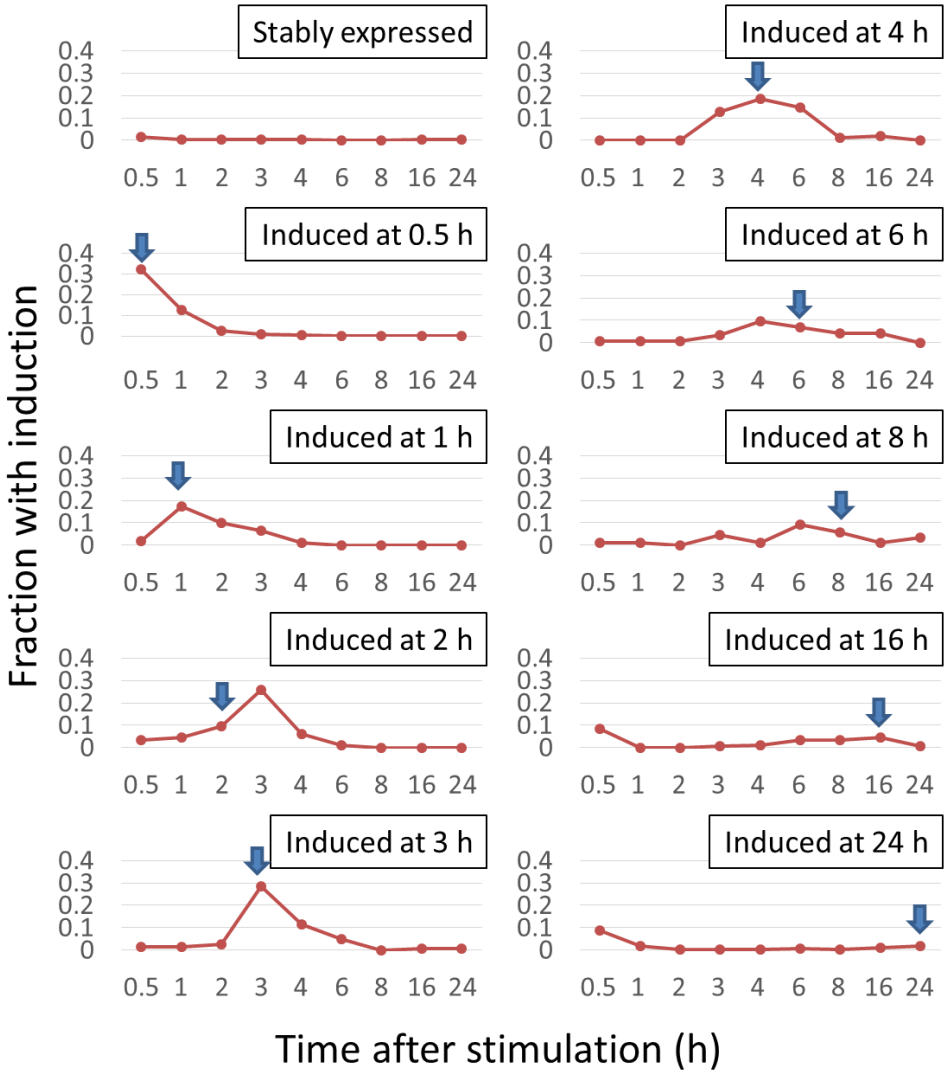


Fig. S4: Induction times of RNA (based on RNA-seq reads) mapped to LPS-induced promoters, in function of their transcription induction time (measured using TSS-seq data). Each graph shows for a set of promoters the induction times of RNA levels mapped to those promoters. Graphs are shown for stably expressed promoters, and for promoters induced by LPS at 0.5, 1, 2, 3, 4, 6, 8, 16, or 24 hours. Blue arrows indicate the time of transcription induction as defined using the TSS-seq data.

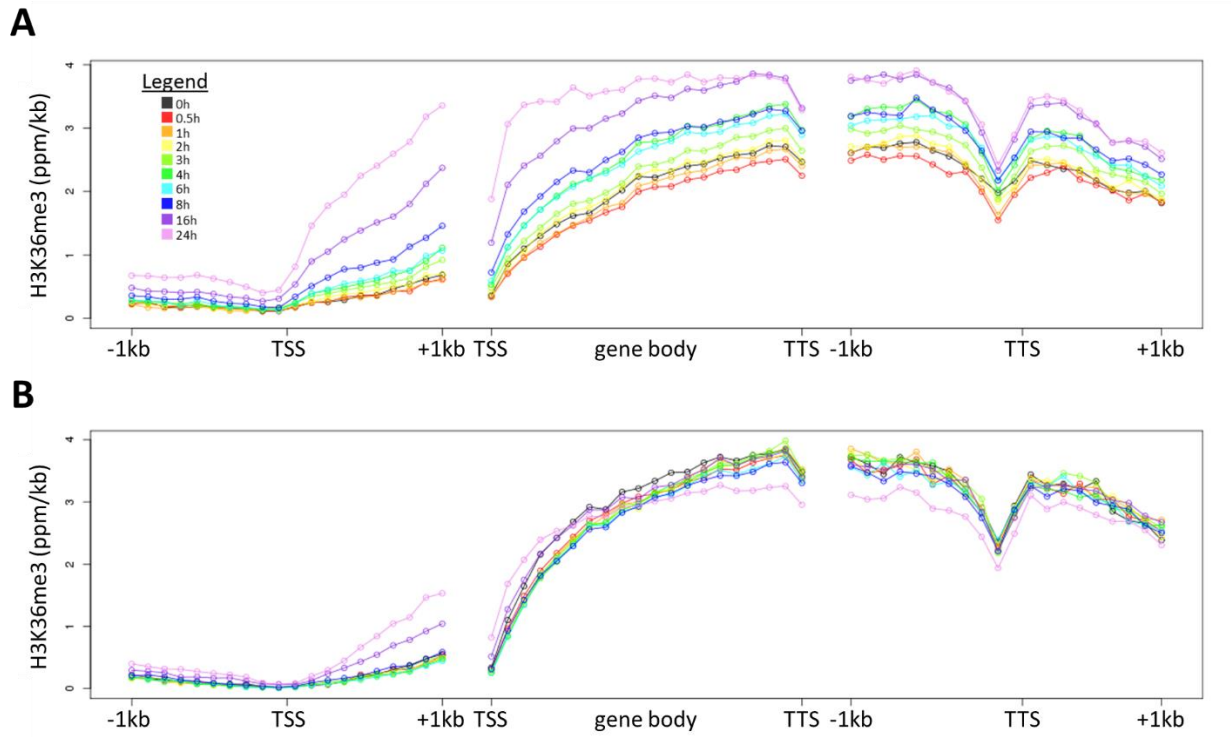


Fig. S5: Changes in H3K36me3 levels at genes' 5' end, gene bodies, and 3' ends. **(A)** Accumulation of H3K36me3 at genes with transcriptional induction after LPS stimulation. The average H3K36me3 signal (ppm/kb) is shown over time for the regions around the transcription start site (TSS) (left, -1kb to +1kb in bins of 100 bps), in the gene body (center, divided into 20 bins of equal size spanning the region from TSS to TTS), and around the transcription termination site (TTS) (right). **(B)** Same as **(A)** for genes with stable expression levels over the time course.

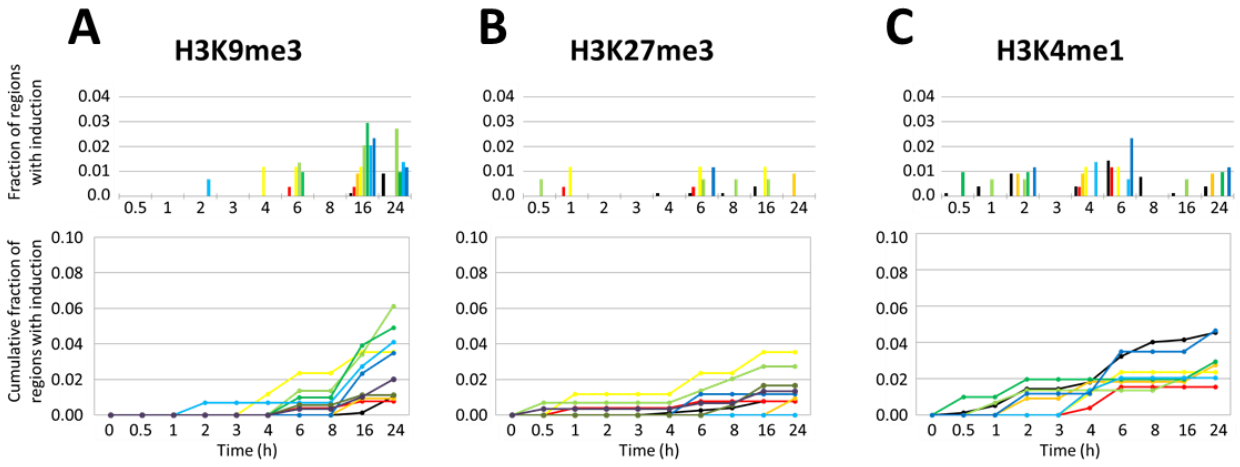


Fig. S6: Induction times of H3K9me3 (A), H3K27me3 (B), and H3K4me1 (C) at promoters in function of induction of transcriptional activation times. Plots are similar to those shown in Fig. 2.

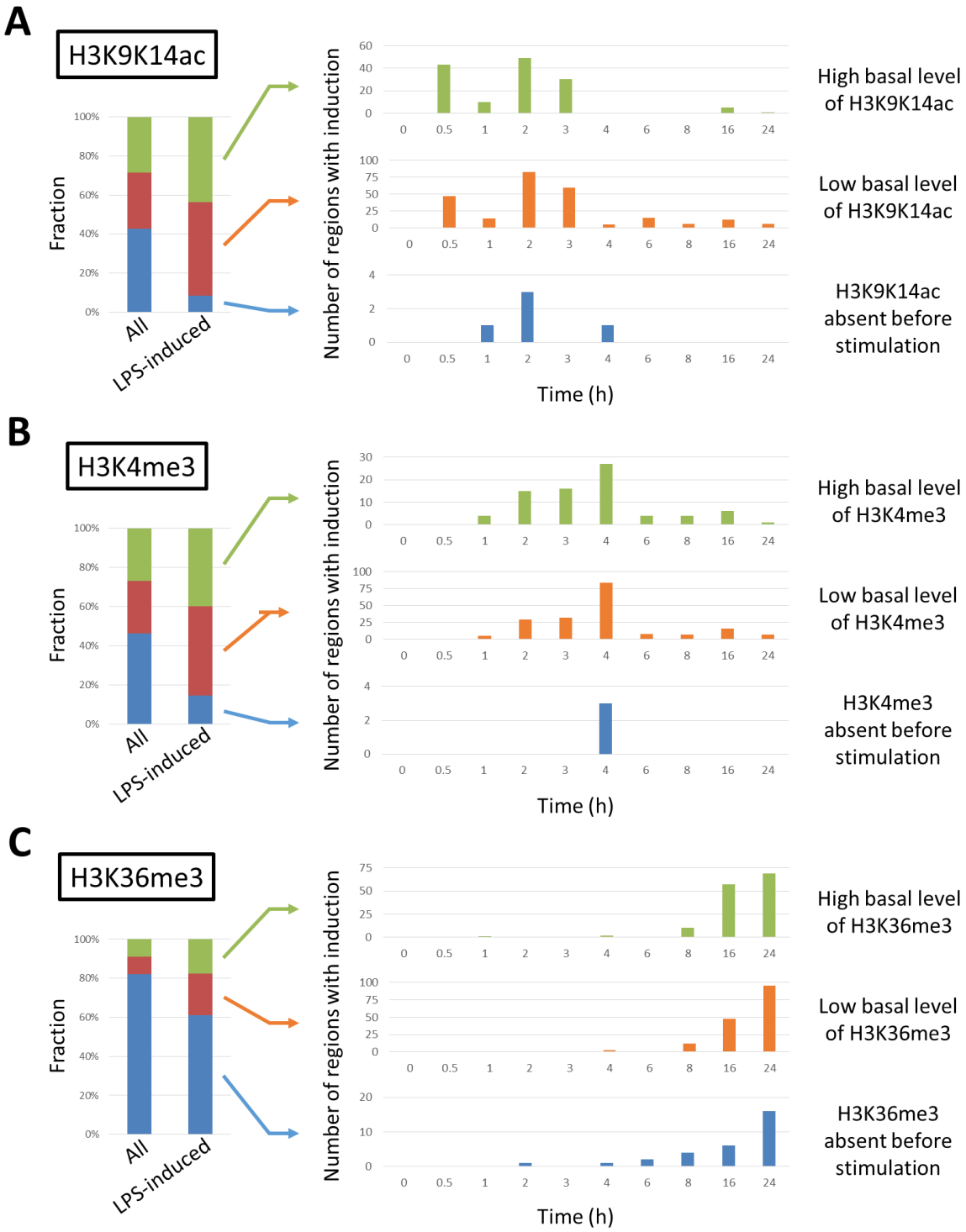


Fig. S7: Induction times of histone modifications in function of their basal levels. **(A)** The genome-wide set of promoters was divided by their pre-stimulation levels of H3K9K14ac into three classes (see Methods). The fraction of promoters in each class is shown at the left, for all promoters, and for LPS-induced promoters. At the right, for each of the three classes, induction times of H3K9K14ac are shown. The same plots are shown for H3K4me3 **(B)**, and H3K36me3 **(C)**.

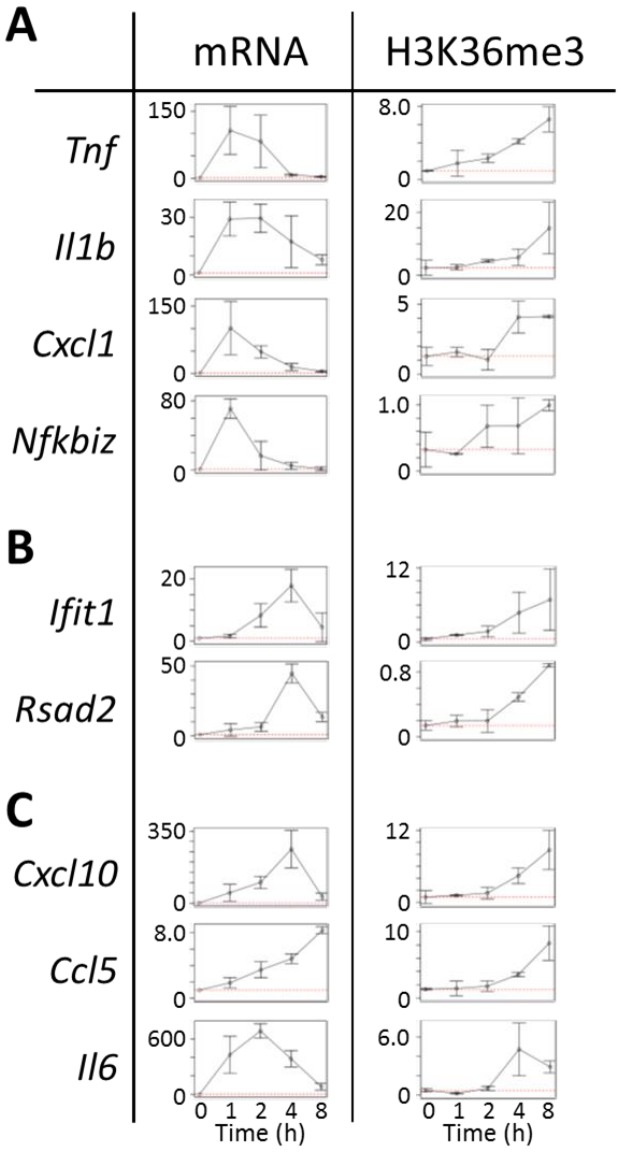


Fig. S8: Gene expression (mRNA) and H3K36me3 dynamics in WT DCs following LPS stimulation. We distinguished genes into the same three groups as described in Fig. 7, according to their dependence on TRIF, IRF3, and IFNR. In all 9 genes, accumulation of H3K36me3 is relatively late compared to induction of gene expression or accumulation of H3K9K14ac and H3K4me3 (see Fig. 7). Error bars represent the standard deviation based on duplicate experiments. The red dotted line in each graph represents the mean value at 0h. Y axes represent fold induction (for mRNA) and % input (for H3K36me3).

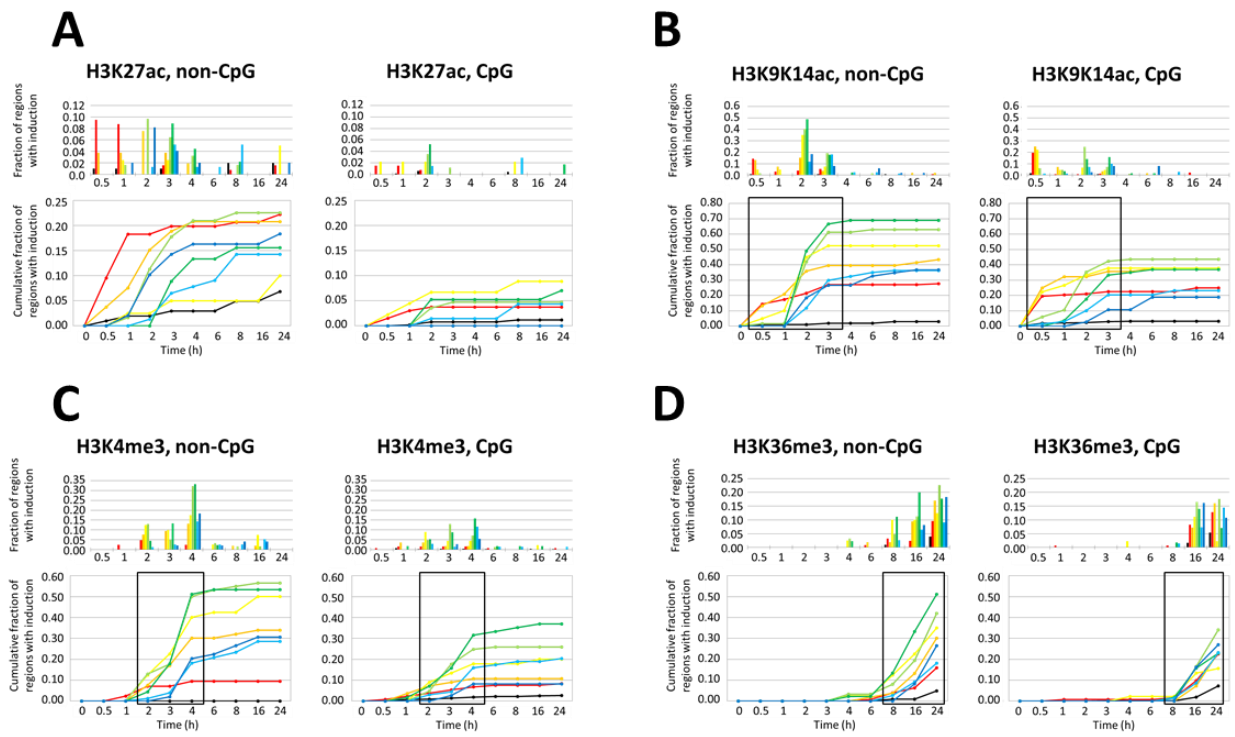


Fig. S9: Induction times of four histone modifications at non-CpG and CpG promoters. Plots are similar to those shown in Fig. 2. (A) H3K27ac, (B) H3K9K14ac, (C) H3K4me3, and (D) H3K36me3. For each pair of plots, the same Y axis scale is used to allow easier comparison of the differences between inductions at non-CpG and CpG promoters.

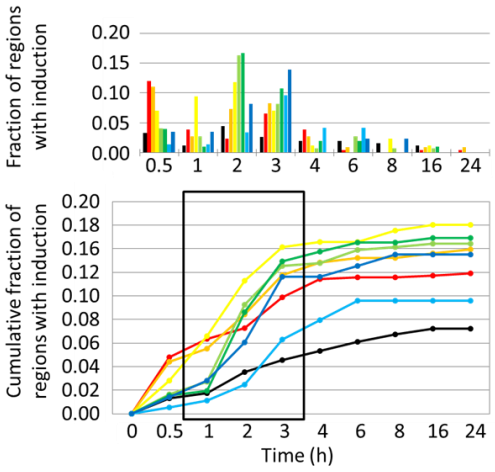


Fig. S10: Induction times of H3K9K14ac at enhancers of LPS-induced promoters. The plot is similar to those shown in Fig. 2. The Y axis shows the fraction of LPS-induced promoters having at least one assigned enhancer with induction in of H3K9K14ac.

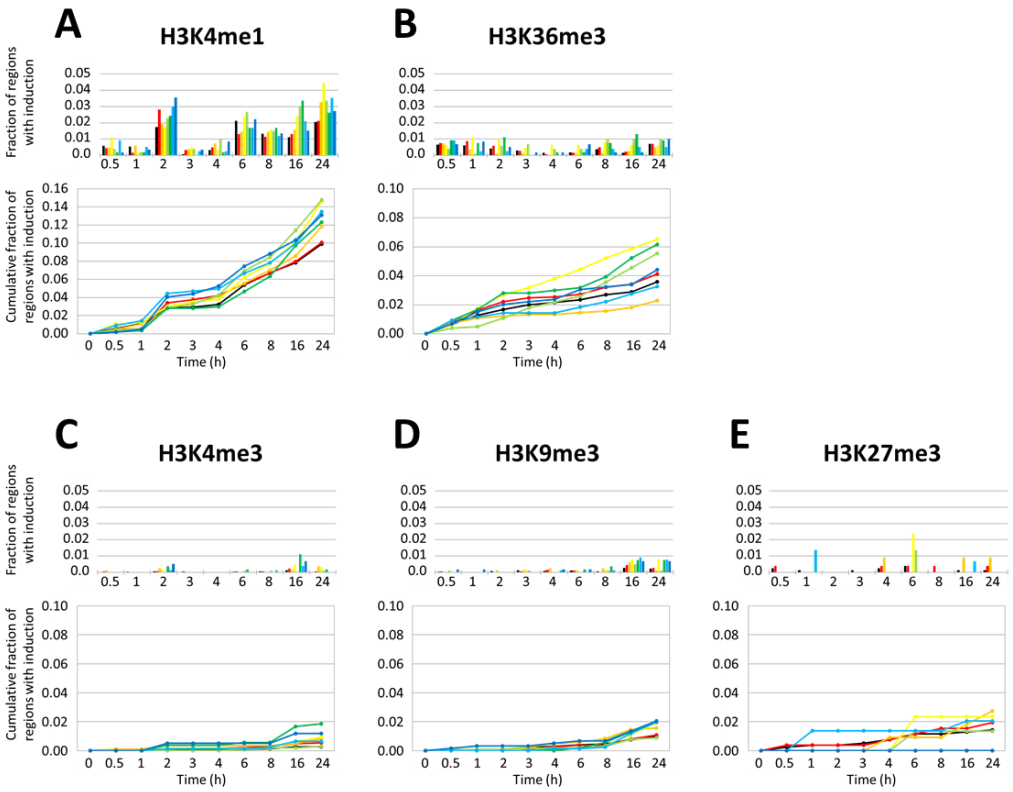


Fig. S11: Induction times of H3K4me1 (A), H3K36me3 (B), H3K4me3 (C), H3K9me3 (D), and H3K27me3 (E) at enhancers of LPS-induced promoters. The plots are similar to those shown in Fig. 3. The Y axis shows the fraction of enhancers having increases in histone modifications in function of time (X axis). Colors of lines represent the timing of transcriptional induction of the assigned promoters.

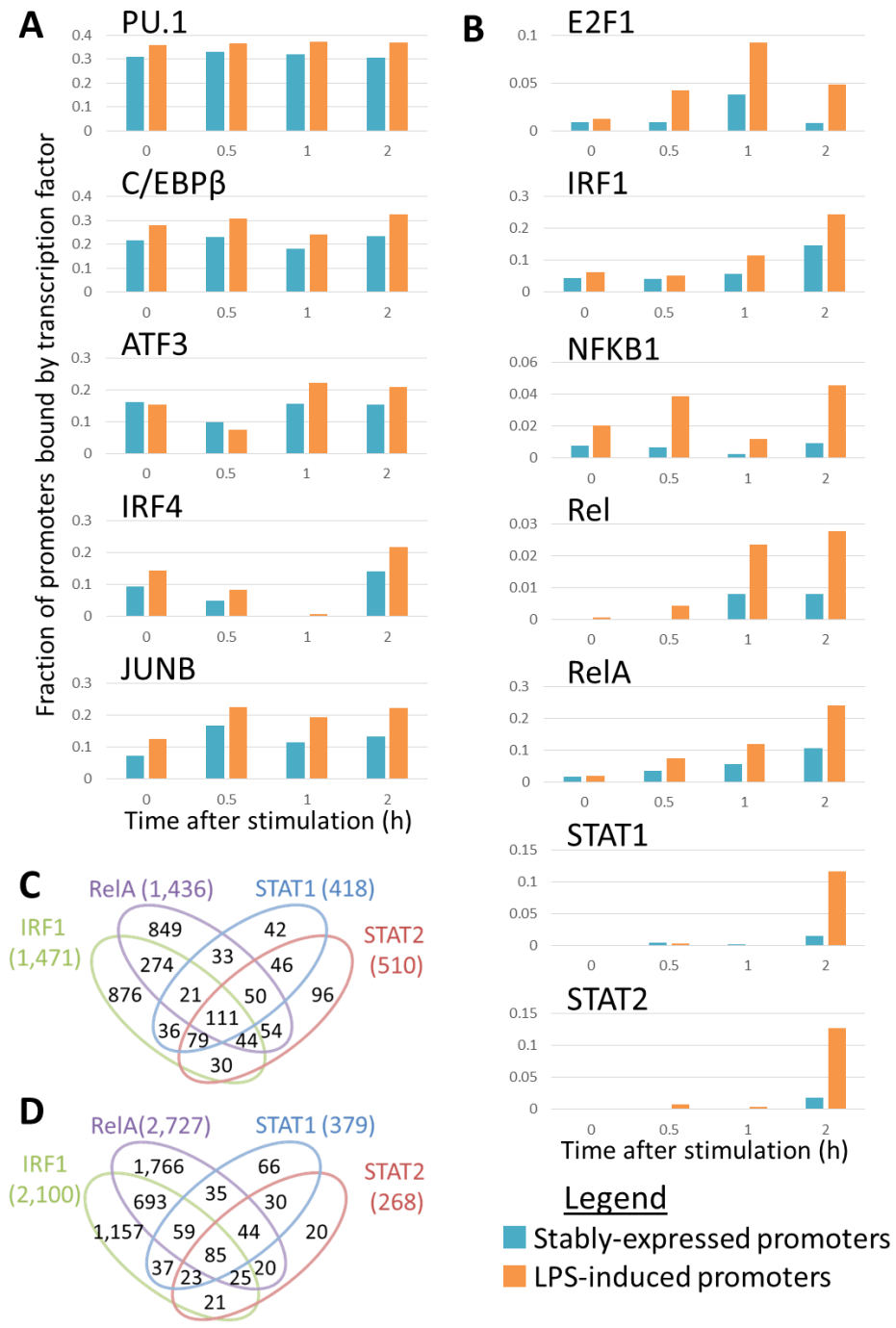


Fig. S12: TF binding in DCs following LPS stimulation. (A-B) For a number of TFs the fraction of promoters bound by each TF is shown in function of time after LPS stimulation. Cyan: stably expressed promoters; orange: LPS-induced promoters. TFs shown in (A) are highly expressed even before LPS stimulation. TFs shown in (B) are LPS-induced. (C-D) Overlap in regions that become bound by IRF1, RelA, STAT1, and STAT2 after LPS stimulation, at promoters (C) and enhancers (D). Numbers in the Venn diagram show the numbers of regions bound by combinations of the four TFs. All combinatorial binding was observed more frequently than expected. Numbers in parentheses show the total number of regions that become bound after stimulation for each TF.

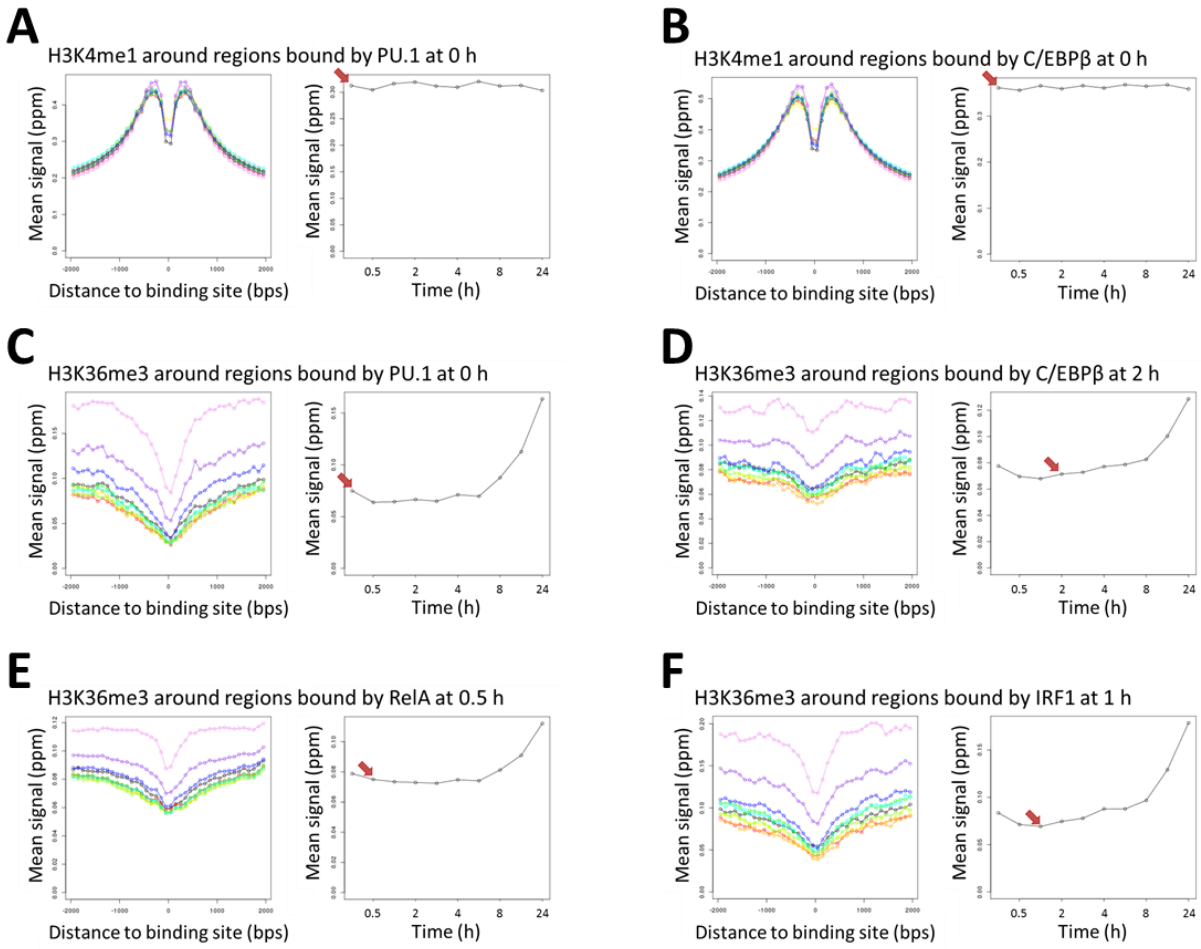


Fig. S13: Interaction between TF binding and histone modifications. (A) For all genomic regions that are pre-bound by PU.1 (0h), mean H3K4me1 signals are shown over time. Left: profile of mean values (y axis) over time in bins of 100 bps in function of distance (x axis) to the TF binding site. Right: mean values (y axis) summed over the region -2kb to +2kb over all bound regions, over time (x axis). The red arrow indicates the time at which these regions become bound by PU.1. (B) Same as (A) for C/EBPβ and H3K4me1. (C-F) Same as (A) for H3K36me3 around regions bound by PU.1, C/EBPβ, RelA, and IRF1.

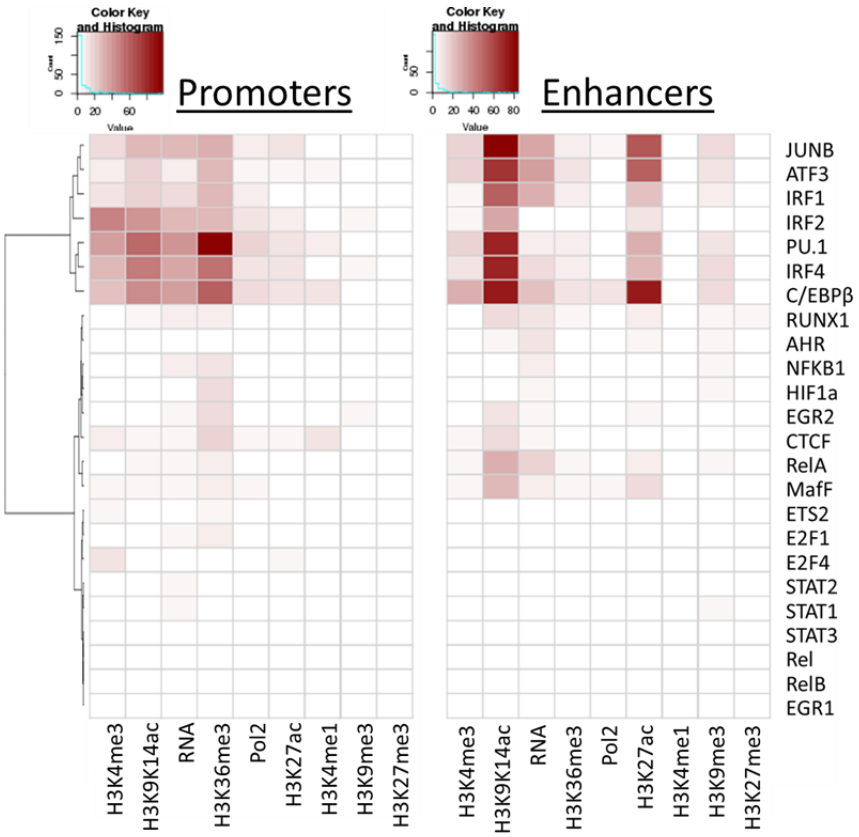


Fig. S14: Associations between pre-stimulation TF binding at promoters (left) and enhancers (right), and LPS-induced increases in histone modifications, Pol2 binding and transcription at the newly bound regions. Colors in the heatmap represent the degree of co-incidence (Fisher's exact test, $-\log_{10}$ p values) between pre-stimulus TF binding (rows) and post-stimulus increases (columns). TFs (rows) have been grouped through hierarchical clustering by similarity of their association pattern.

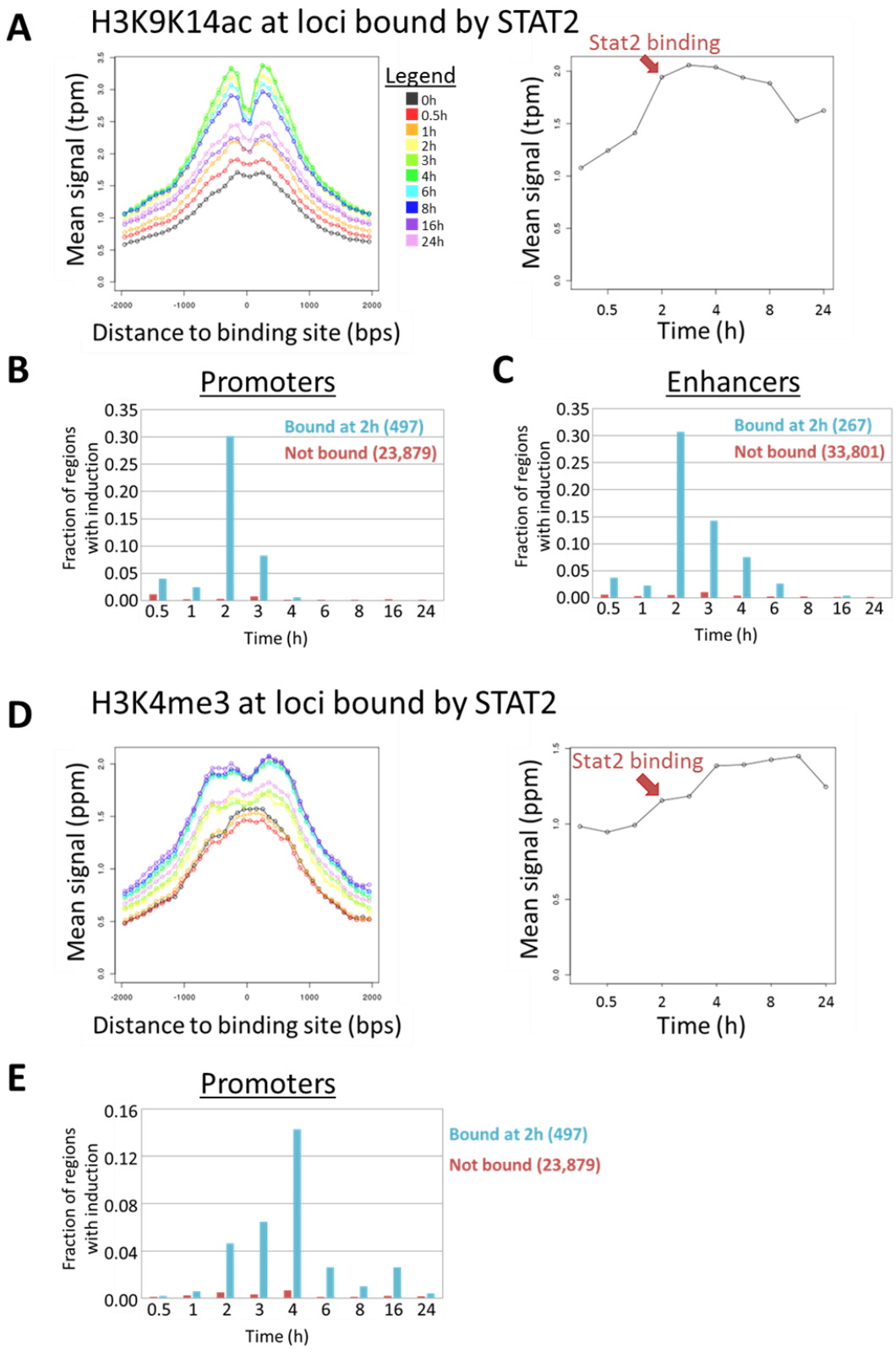


Fig. S15: Interaction between STAT2 binding and histone modifications over time. The plots shown are the same as shown in Fig. 5 for STAT1 bound regions.

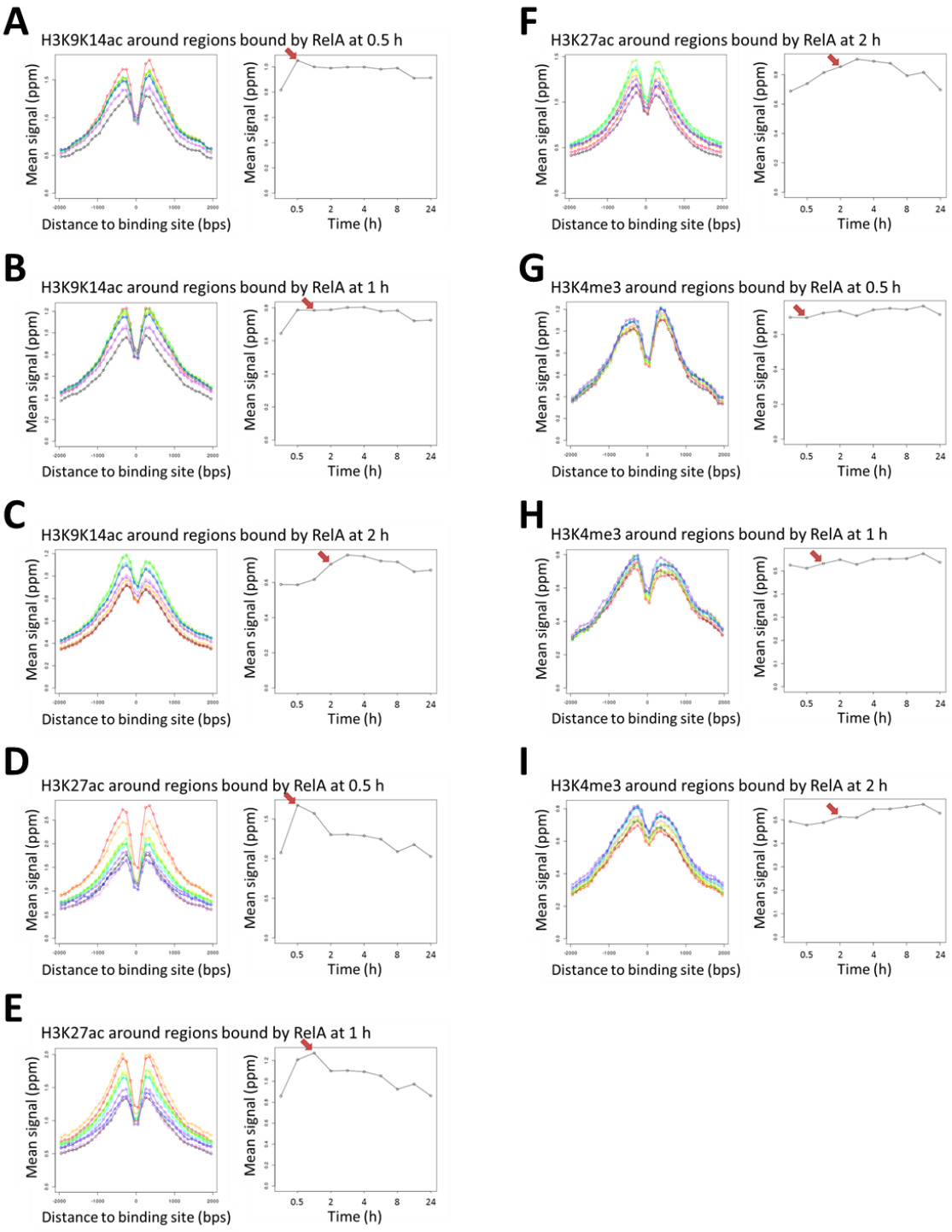


Fig. S16: Interaction between RelA binding and histone modifications. (A-C) For all genomic regions bound by RelA at 0.5h (A), 1h (B), and 2h (C) after LPS stimulation, mean H3K9K14ac signals are shown over time. Left: profile of mean values (y axis) over time in bins of 100 bps in function of distance (x axis) to the TF binding site. Right: mean values (y axis) summed over the region -2kb to +2kb over all bound regions, over time (x axis). The red arrow indicates the time at which these regions become bound by RelA. (D-F) Similar as (A-C) for H3K27ac. (G-I) Similar as (A-C) for H3K4me3.

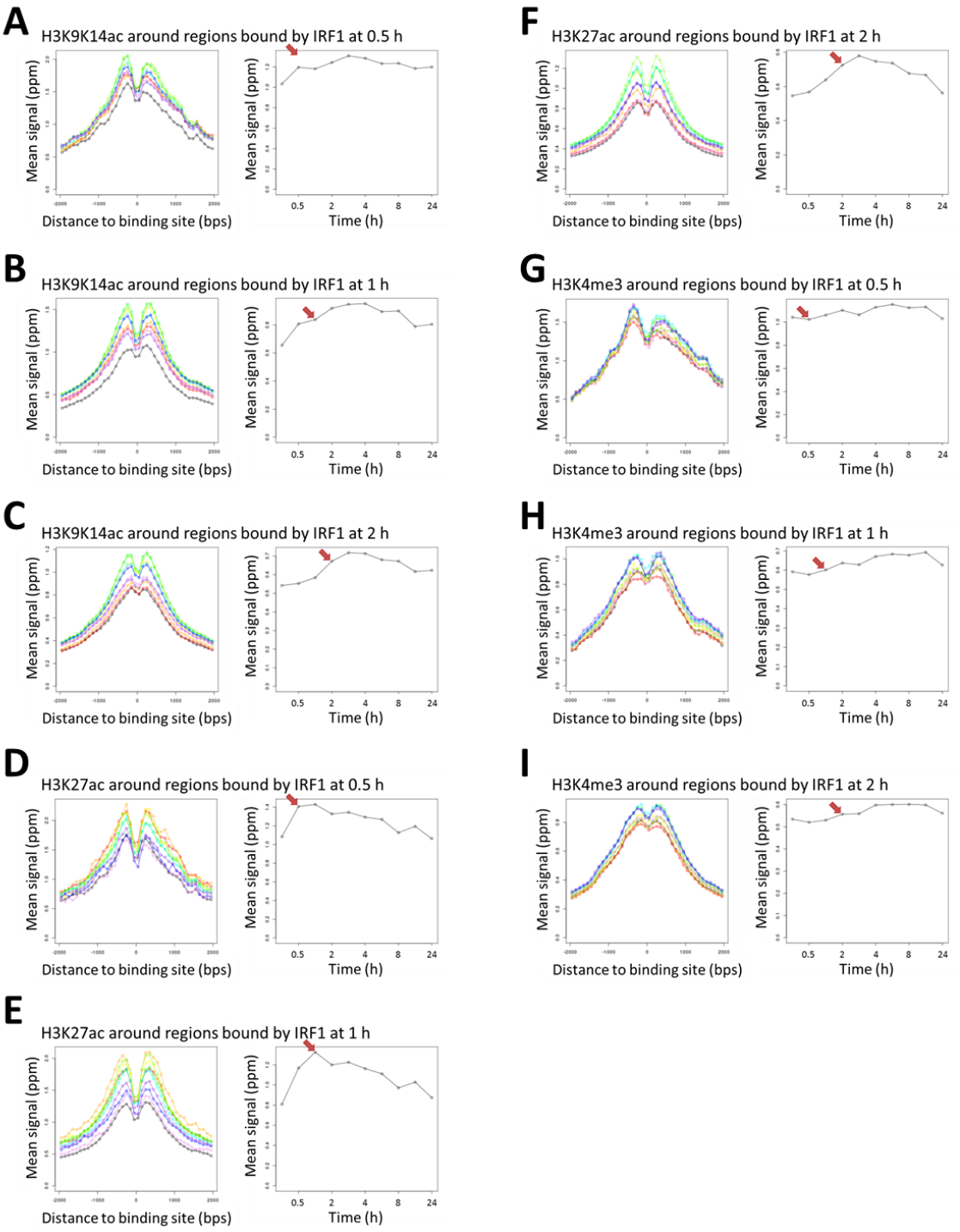


Fig. S17: Interaction between IRF1 binding and histone modifications. (A-C) For all genomic regions bound by IRF1 at 0.5h (A), 1h (B), and 2h (C) after LPS stimulation, mean H3K9K14ac signals are shown over time. Left: profile of mean values (y axis) over time in bins of 100 bps in function of distance (x axis) to the TF binding site. Right: mean values (y axis) summed over the region -2kb to +2kb over all bound regions, over time (x axis). The red arrow indicates the time at which these regions become bound by IRF1. (D-F) Similar as (A-C) for H3K27ac. (G-I) Similar as (A-C) for H3K4me3.

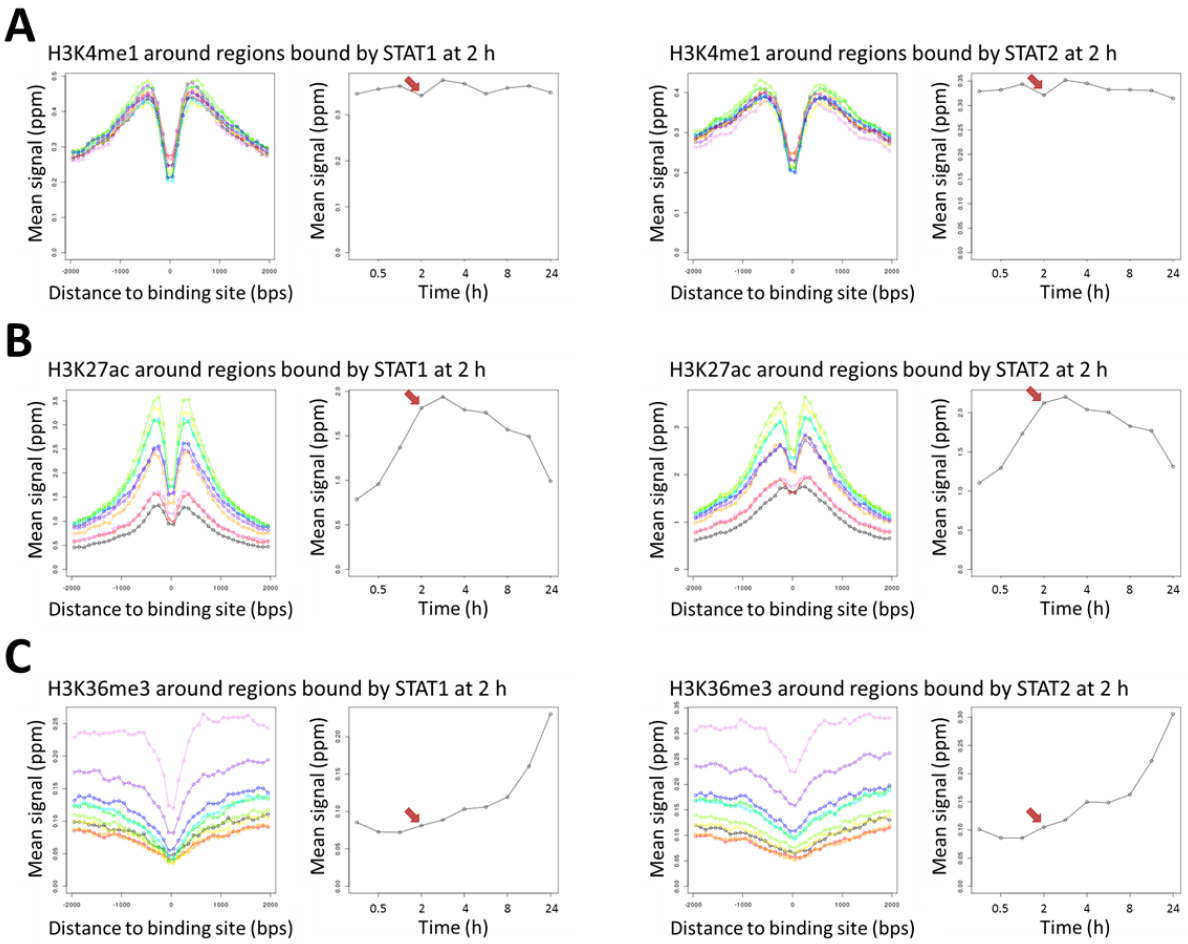


Fig. S18: Interaction between STAT1/STAT2 binding and histone modifications over time. The plots shown are similar to those shown in Fig. 5 for STAT1-bound regions, and in Fig. S15 for Stat2-bound regions. Plots are shown for H3K4me1 (A), H3K27ac (B), and H3K36me3 (C), for STAT1-bound (left) and STAT2-bound (right) regions.

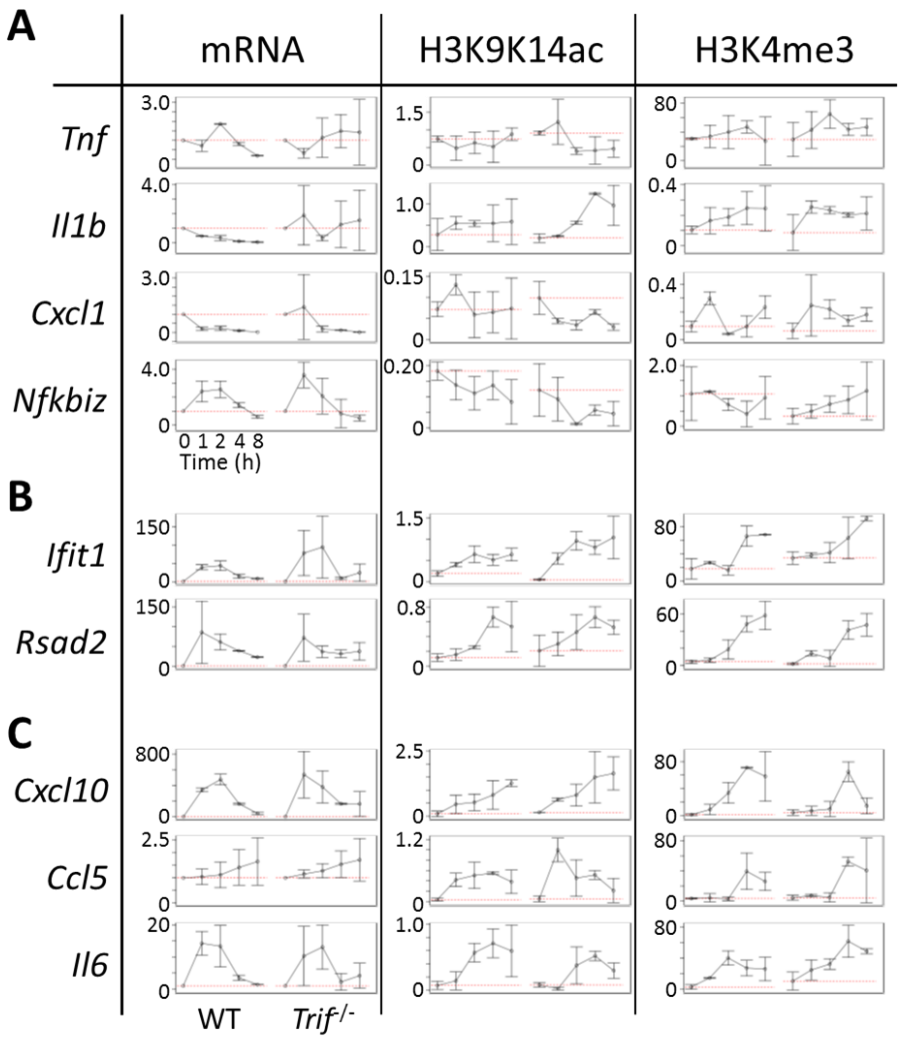


Fig. S19: Gene expression (mRNA), H3K9K14ac and H3K4me3 dynamics in WT, and *Trif*$^{-/-}$ cells following IFN-β stimulation. Genes are divided into three groups as in Fig. 7. Induction of expression and accumulation of H3K9K14ac and H3K36me3 is observed predominantly in genes under **(B)** and **(C)**, but not those under **(A)**. Accumulation of histone modifications is not affected in *Trif*$^{-/-}$ cells. Error bars represent the standard deviation based on duplicate experiments. The red dotted line in each graph represents the mean value at 0h. Y axes represent fold induction (for mRNA) and % input (for H3K9K14ac and H3K4me3).

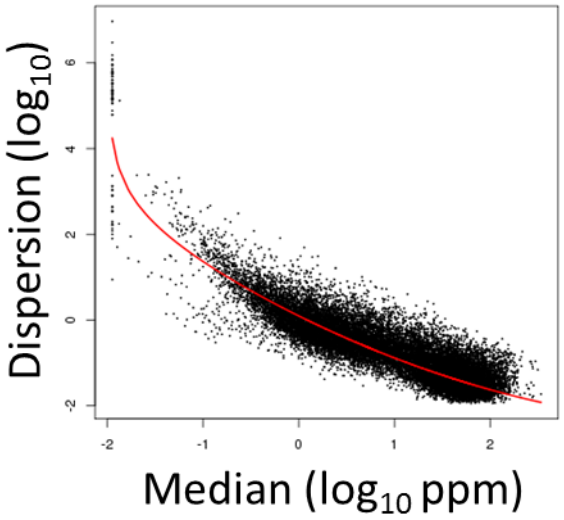


Fig. S20: Plot of median versus dispersion for H3K9K14ac reads (ppm) at the genome-wide set of promoters and enhancers. Each dot represents the median and dispersion of H3K9K14ac signals over the 10 time points, for 1 region. The red line is a plotted second order polynomial.