Molecular Cell, Volume *47*

Supplemental Information

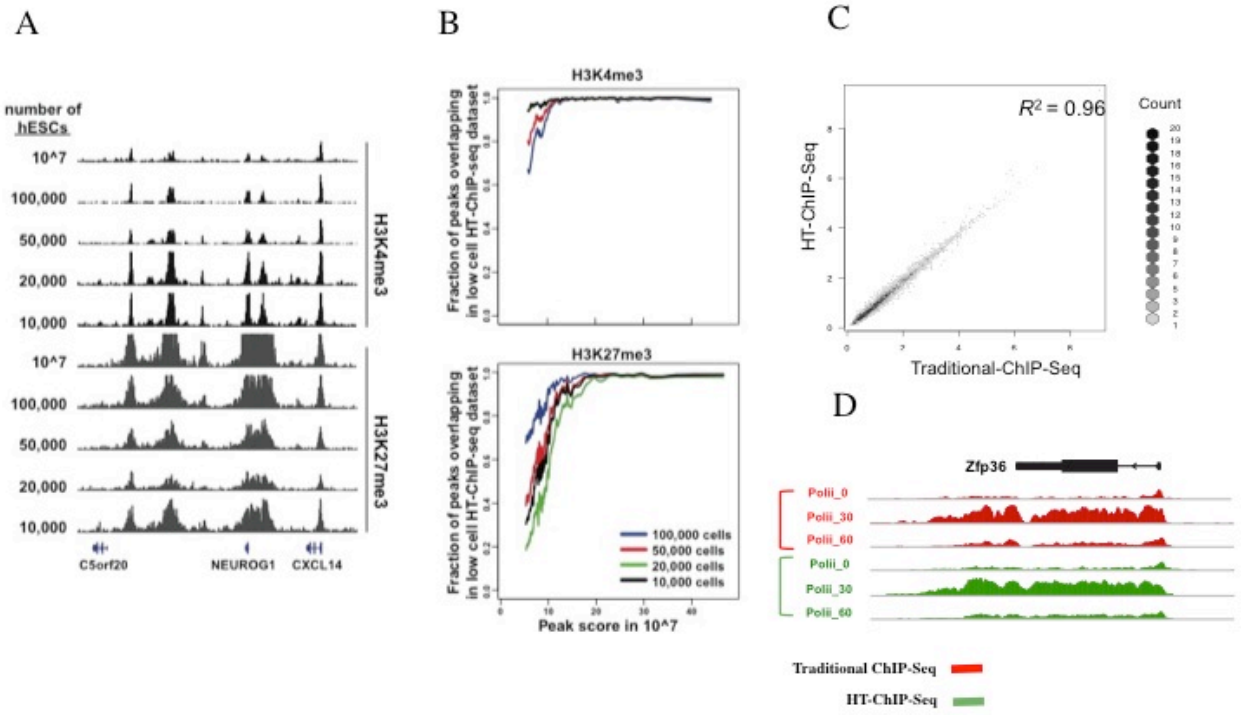
A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals

Manuel Garber, Nir Yosef, Alon Goren, Raktima Raychowdhury, Anne Thielke, Mitchell Guttman, James Robinson, Brian Minie, Nicolas Chevrier, Zohar Itzhaki, Ronnie Blecher-Gonen, Chamutal Bornstein, Daniela Amann-Zalcenstein, Assaf Weiner, Dennis Friedrich, James Meldrim, Oren Ram, Christine Cheng, Andreas Gnirke, Sheila Fisher, Nir Friedman, Bang Wong, Bradley E. Bernstein, Chad Nusbaum, Nir Hacohen, Aviv Regev, and Ido Amit

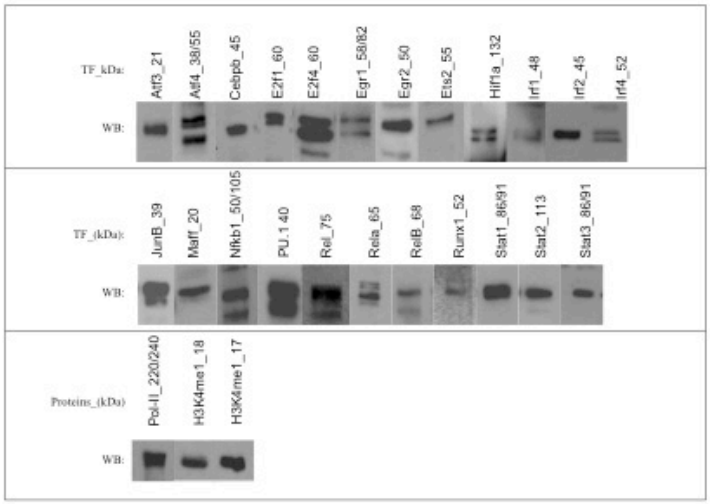
Supplemental Figures

Images

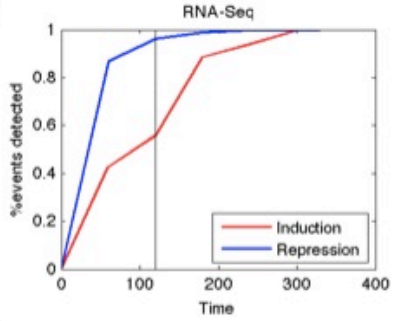
Figure S1



E


TF_kDa:
Atf3_21
Atf4_38/55
Cebpb_45
E2f1_60
E2f4_60
Egr1_58/82
Egr2_50
Ets2_55
Hif1a_132
Irf1_48
Irf2_45
Irf4_52
WB:
TF_(kDa):
JunB_39
Maff_20
Nfkb1_50/105
PU.1 40
Rel_75
Rela_65
RelB_68
Runx1_52
Stat1_86/91
Stat2_113
Stat3_86/91
WB:
Proteins_(kDa)
Pol-II_220/240
H3K4me1_18
H3K4me1_17
WB:

F


RNA-Seq
%events detected
Time
Induction
Repression
0
0.2
0.4
0.6
0.8
1
0
100
200
300
400

G

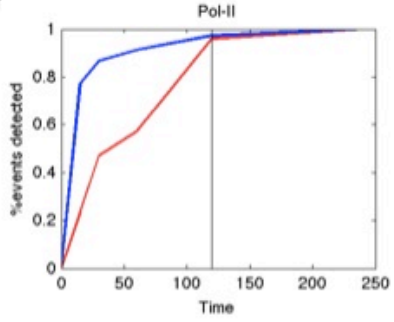
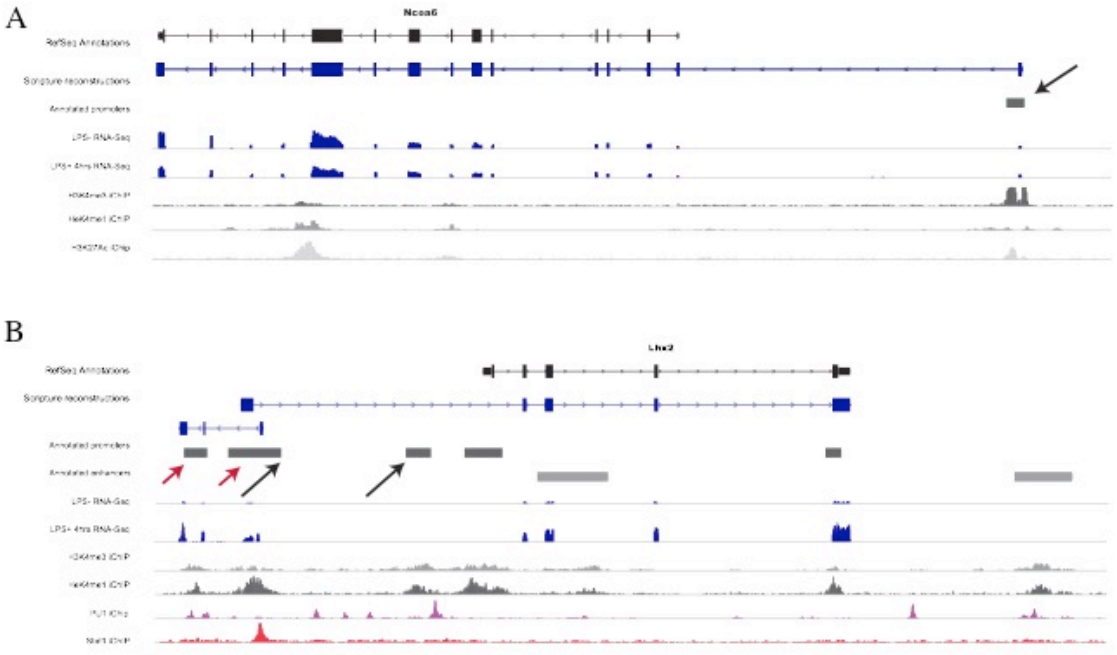
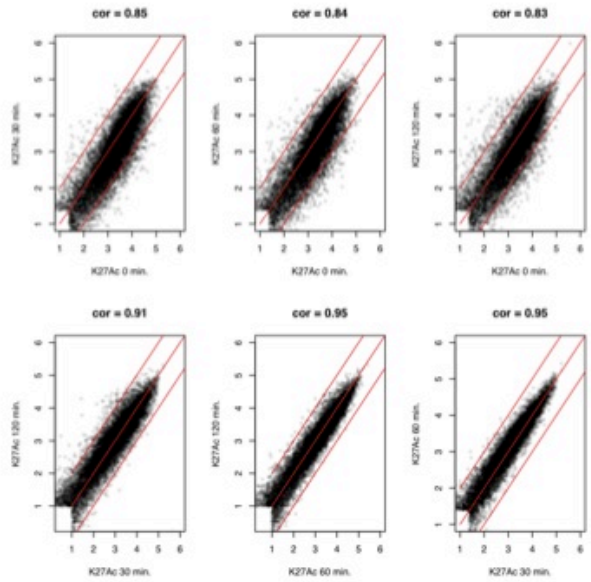

Pol-II
%events detected
Time
0
0.2
0.4
0.6
0.8
1
0
50
100
150
200
250

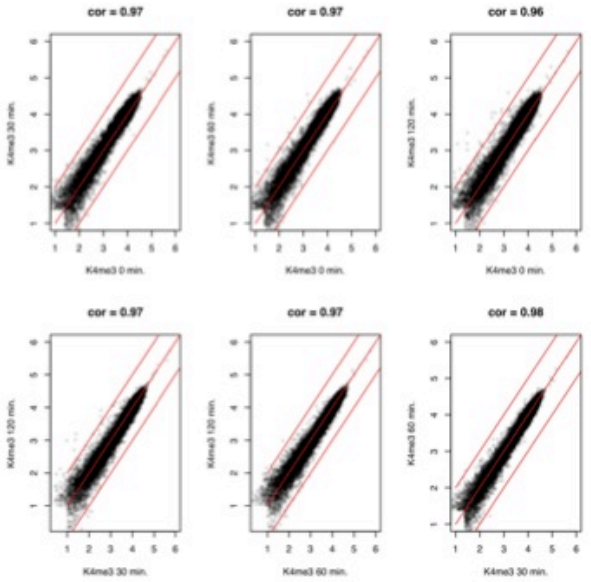
Figure S2


A
Ncoa6
RefSeq Annotations
Scripture reconstructions
Annotated promoters
LPS- RNA-Seq
LPS+ 4hrs RNA-Seq
H3K4me3 ChIP
H3K4me1 ChIP
H3K27Ac ChIP
B
Lhx2
RefSeq Annotations
Scripture reconstructions
Annotated promoters
Annotated enhancers
LPS- RNA-Seq
LPS+ 4hrs RNA-Seq
H3K4me3 ChIP
H3K4me1 ChIP
PU1 ChIp
Stat1 ChIP

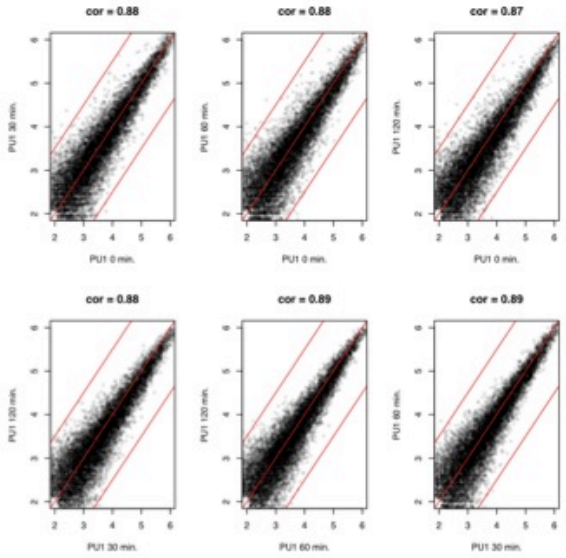
C


cor = 0.85
cor = 0.84
cor = 0.83
K27Ac 30 min.
K27Ac 60 min.
K27Ac 120 min.
K27Ac 0 min.
K27Ac 0 min.
K27Ac 0 min.
cor = 0.91
cor = 0.95
cor = 0.95
K27Ac 120 min.
K27Ac 120 min.
K27Ac 60 min.
K27Ac 30 min.
K27Ac 60 min.
K27Ac 30 min.

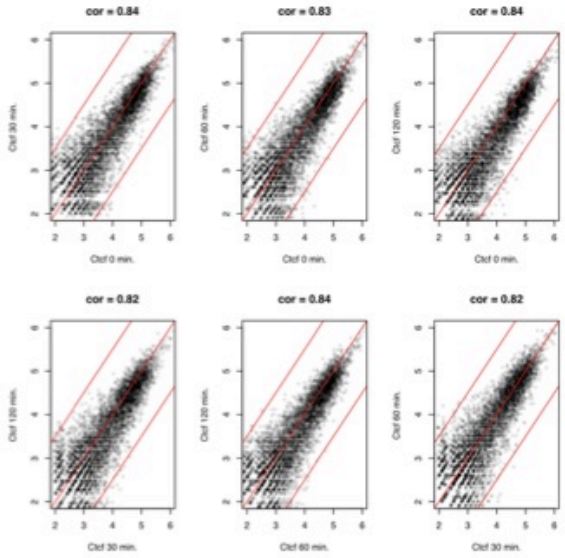
D


cor = 0.97
cor = 0.97
cor = 0.96
K4me3 30 min.
K4me3 60 min.
K4me3 120 min.
K4me3 0 min.
K4me3 0 min.
K4me3 0 min.
cor = 0.97
cor = 0.97
cor = 0.98
K4me3 120 min.
K4me3 120 min.
K4me3 60 min.
K4me3 30 min.
K4me3 60 min.
K4me3 30 min.

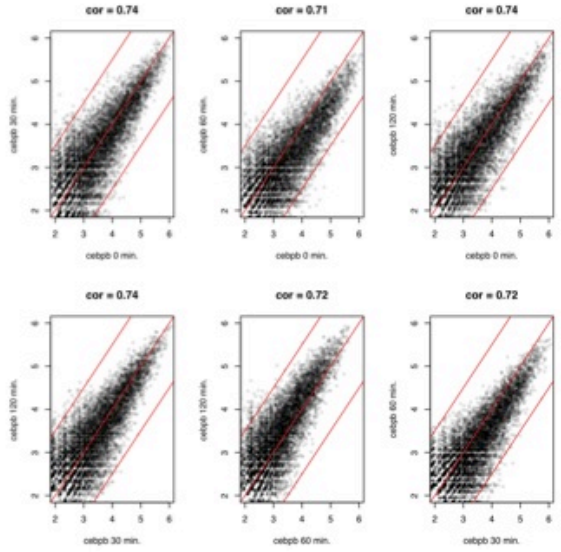
E


cor = 0.88
PU1 30 min.
PU1 0 min.
cor = 0.88
PU1 60 min.
PU1 0 min.
cor = 0.87
PU1 120 min.
PU1 0 min.
cor = 0.88
PU1 120 min.
PU1 30 min.
cor = 0.89
PU1 120 min.
PU1 60 min.
cor = 0.89
PU1 60 min.
PU1 30 min.

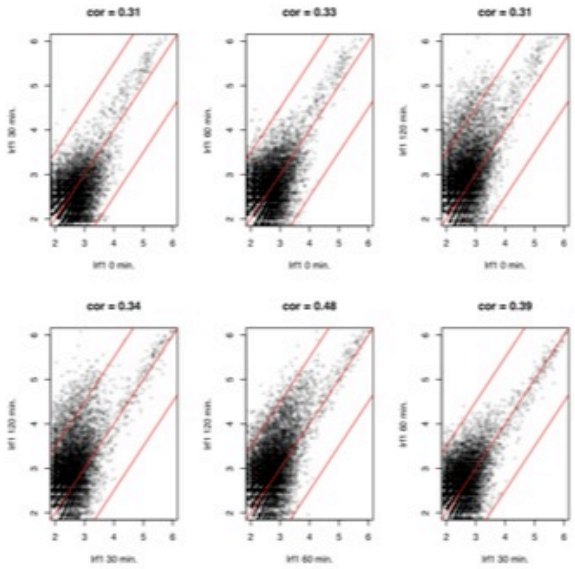
F


cor = 0.84
Ctcf 30 min.
Ctcf 0 min.
cor = 0.83
Ctcf 60 min.
Ctcf 0 min.
cor = 0.84
Ctcf 120 min.
Ctcf 0 min.
cor = 0.82
Ctcf 120 min.
Ctcf 30 min.
cor = 0.84
Ctcf 120 min.
Ctcf 60 min.
cor = 0.82
Ctcf 60 min.
Ctcf 30 min.

G


cor = 0.74
cor = 0.71
cor = 0.74
cor = 0.74
cor = 0.72
cor = 0.72
cebpb 0 min.
cebpb 30 min.
cebpb 60 min.
cebpb 120 min.

H


cor = 0.31
cor = 0.33
cor = 0.31
cor = 0.34
cor = 0.48
cor = 0.39
irf1 0 min.
irf1 30 min.
irf1 60 min.
irf1 120 min.

I

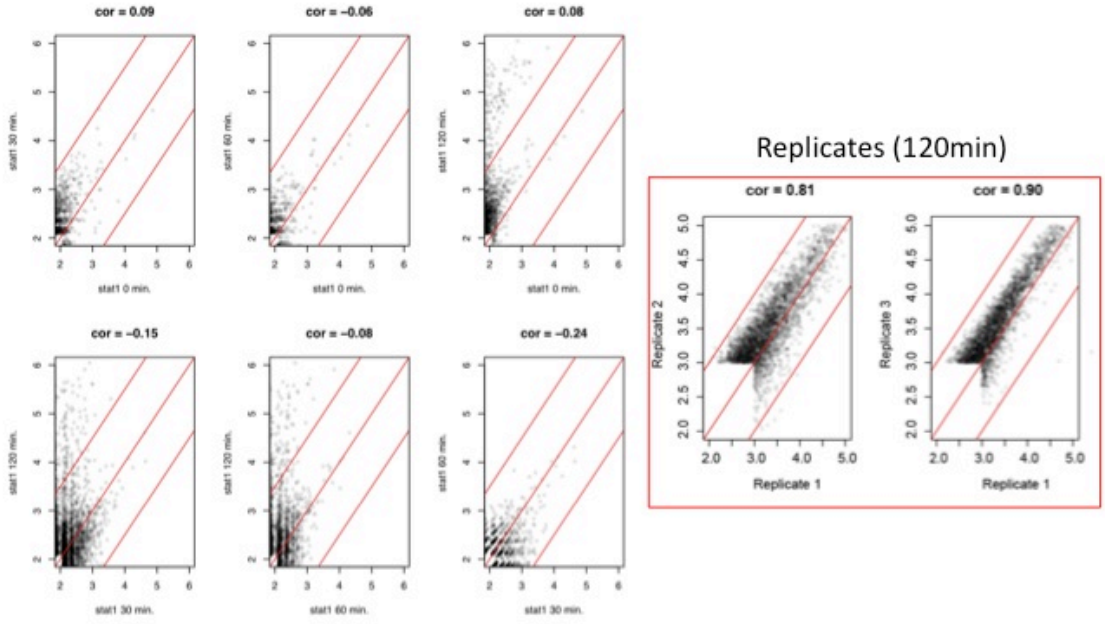
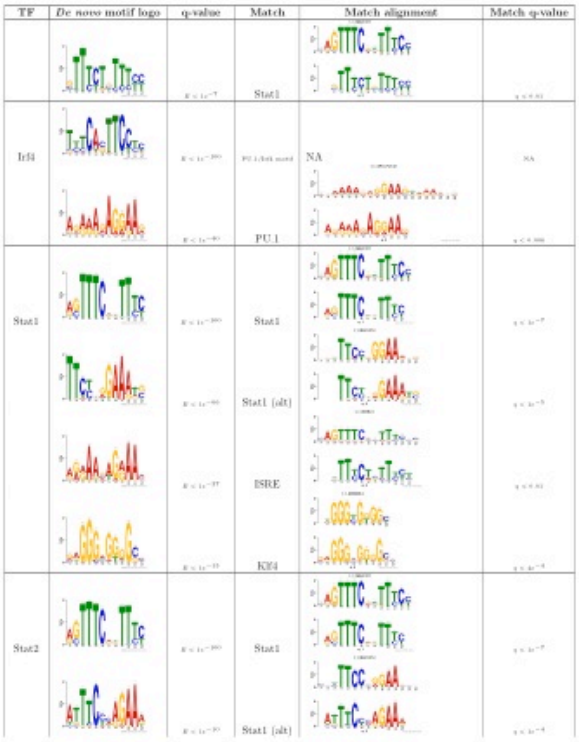

cor = 0.09
cor = −0.06
cor = 0.08
cor = −0.15
cor = −0.08
cor = −0.24
stat1 30 min.
stat1 60 min.
stat1 120 min.
stat1 0 min.
Replicates (120min)
cor = 0.81
cor = 0.90
Replicate 2
Replicate 3
Replicate 1

Figure S2J (Part I)



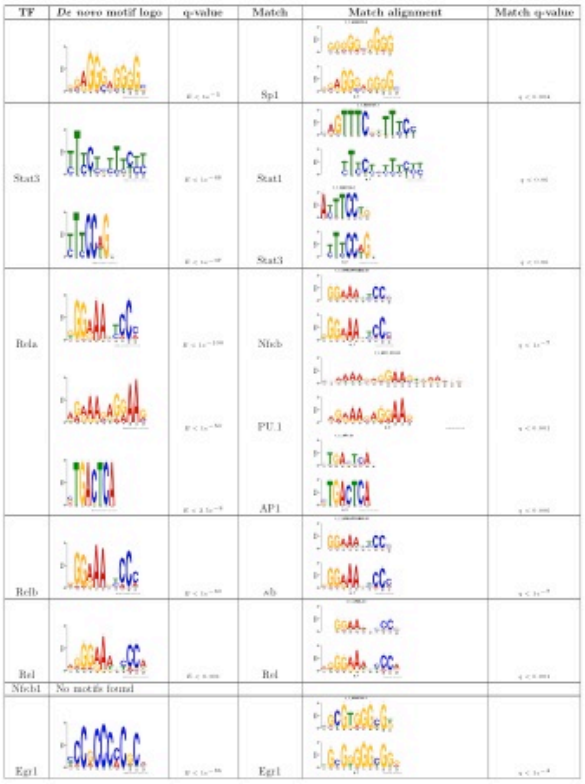


Figure S2J (Part II)

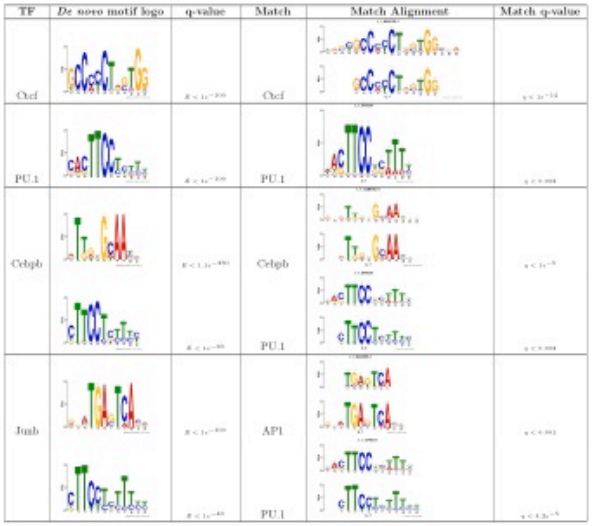
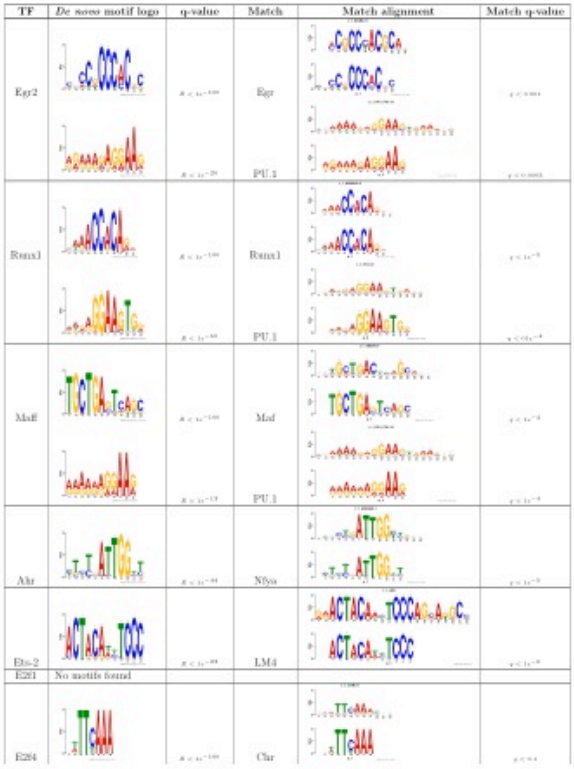
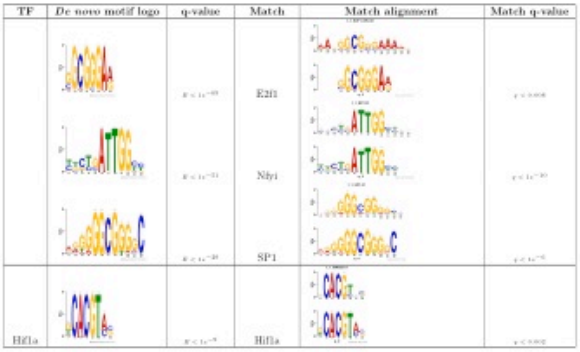


Figure S2J (Part III)



TF	*De novo* motif logo	q-value	Match	Match alignment	Match q-value
Egr2			Egr		
			PU.1		
Runx1			Runx1		
			PU.1		
Maff			Mad		
			PU.1		
Ahr			Nfya		
Ets-2			LM4		
E2f1	No motifs found				
E2f4			Clx		



TF	*De novo* motif logo	q-value	Match	Match alignment	Match q-value
			E2f1		
			Nfyi		
			SP1		
Hif1a			Hif1a		

Figure S3

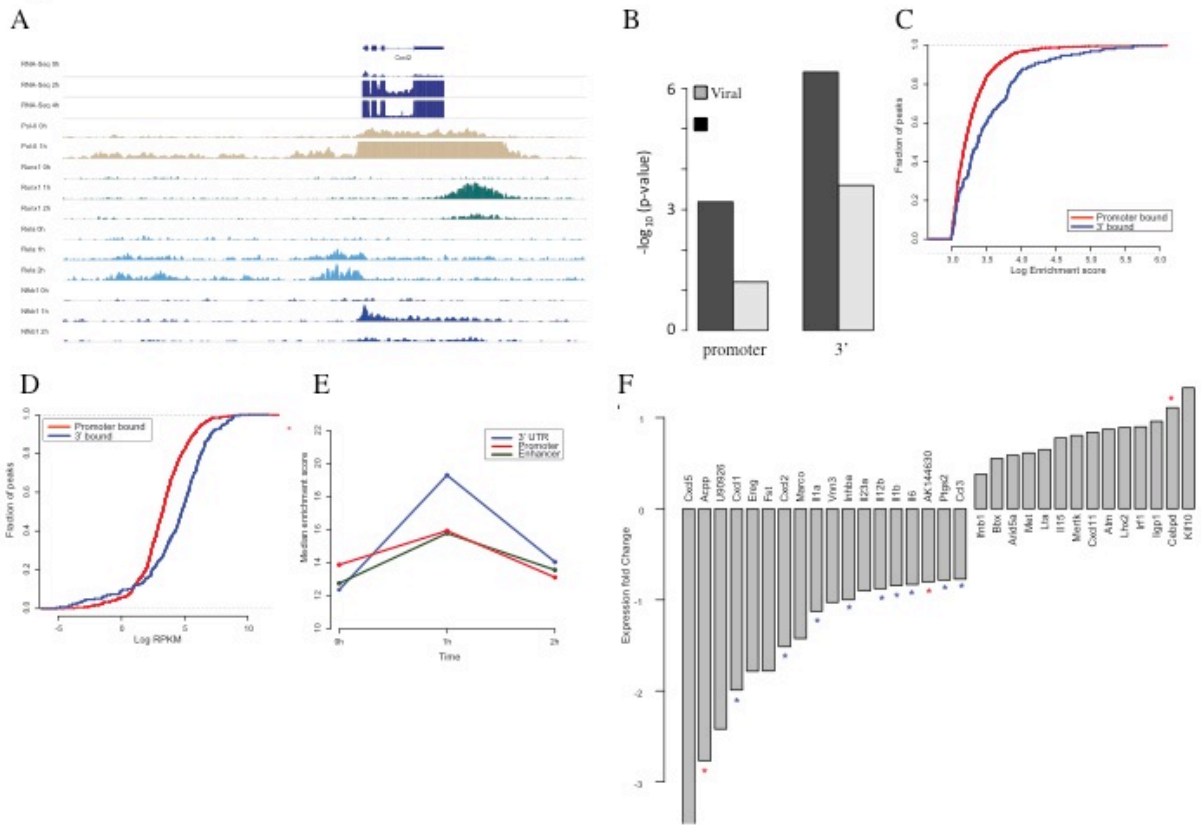
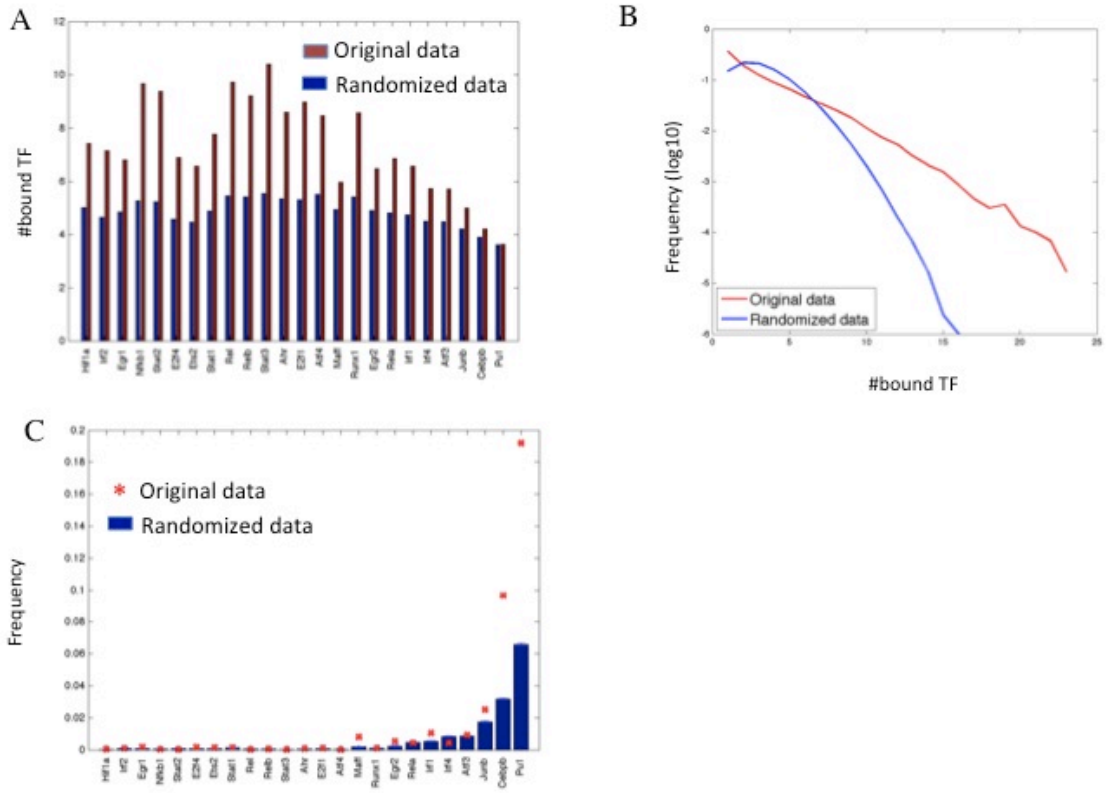

A
B
C
D
E
F
Viral
promoter
3'
-log10 (p-value)
Promoter bound
3' bound
Fraction of peaks
Log Enrichment score
Log RPKM
3' UTR
Promoter
Enhancer
Median enrichment score
Time
Expression fold Change
Cxcl5
Acpp
U90926
Cxcl1
Ereg
Fst
Cxcl2
Marco
Il1a
Vnn3
Inhba
Il23a
Il12b
Il1b
Il6
AK144630
Ptgs2
Ccl3
Rnb1
Btx
Aldh3a
Mat
Lta
Il15
Mertk
Cxcl11
Atm
Lhx2
Irf1
Igp1
Cebpd
Kif10

Figure S4


A
Original data
Randomized data
#bound TF
B
Frequency (log10)
Original data
Randomized data
#bound TF
C
Original data
Randomized data
Frequency

D

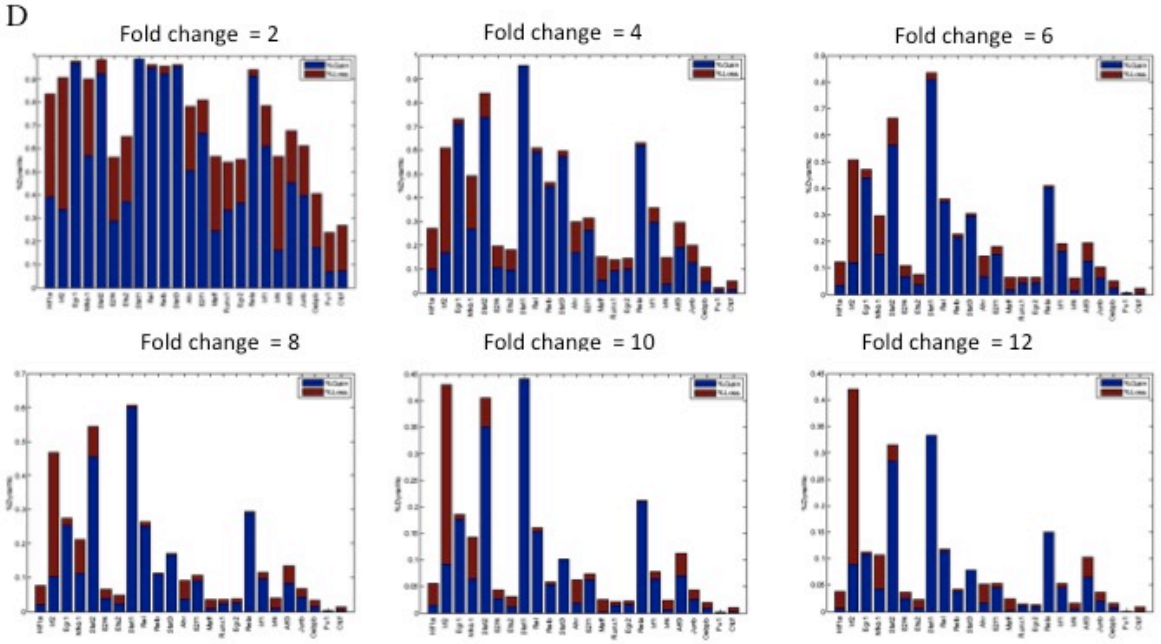

Fold change = 2
Fold change = 4
Fold change = 6
Fold change = 8
Fold change = 10
Fold change = 12

Figure S5

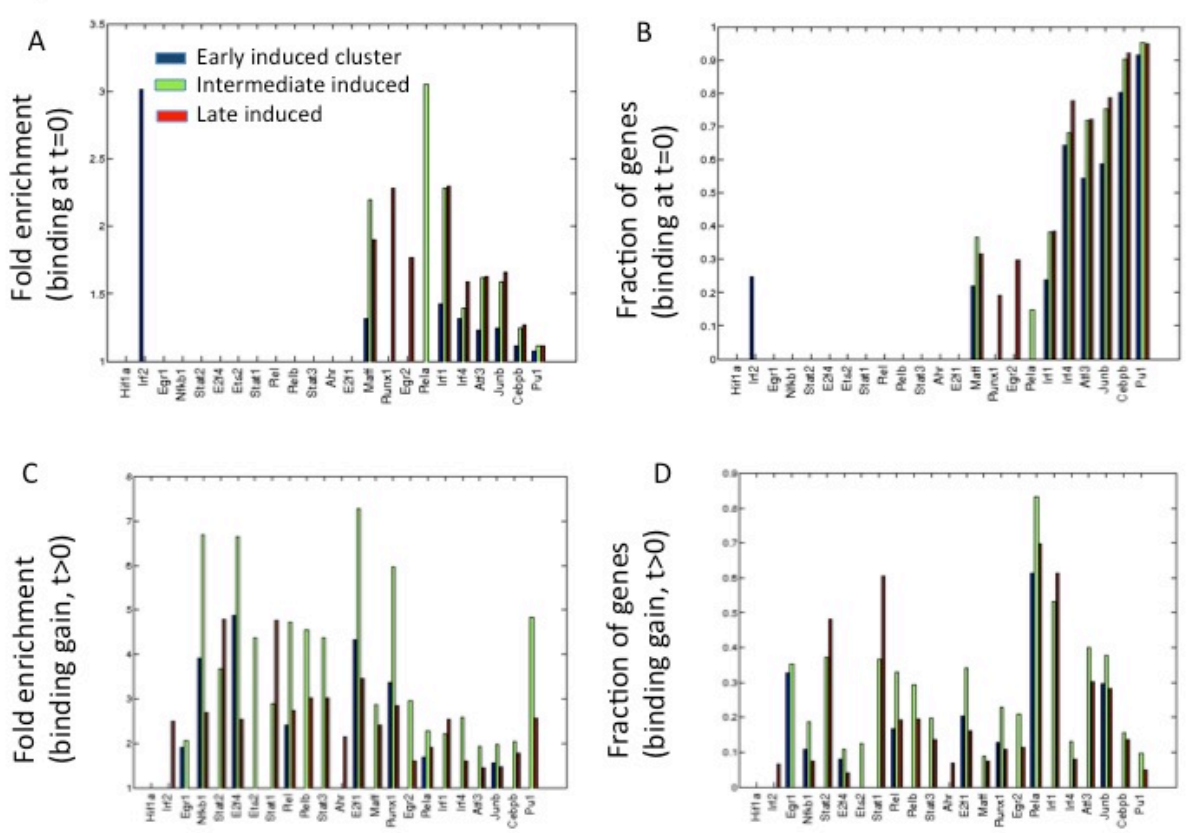

A
Fold enrichment
(binding at t=0)
Early induced cluster
Intermediate induced
Late induced
B
Fraction of genes
(binding at t=0)
C
Fold enrichment
(binding gain, t>0)
D
Fraction of genes
(binding gain, t>0)

Figure S5E (part I)

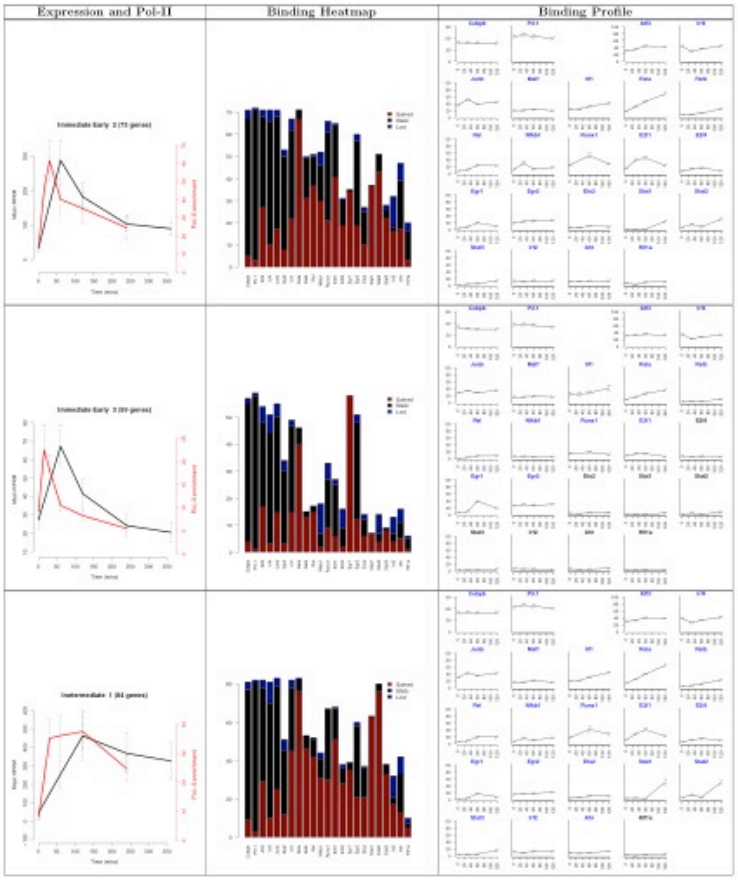


Figure S5E (part II)

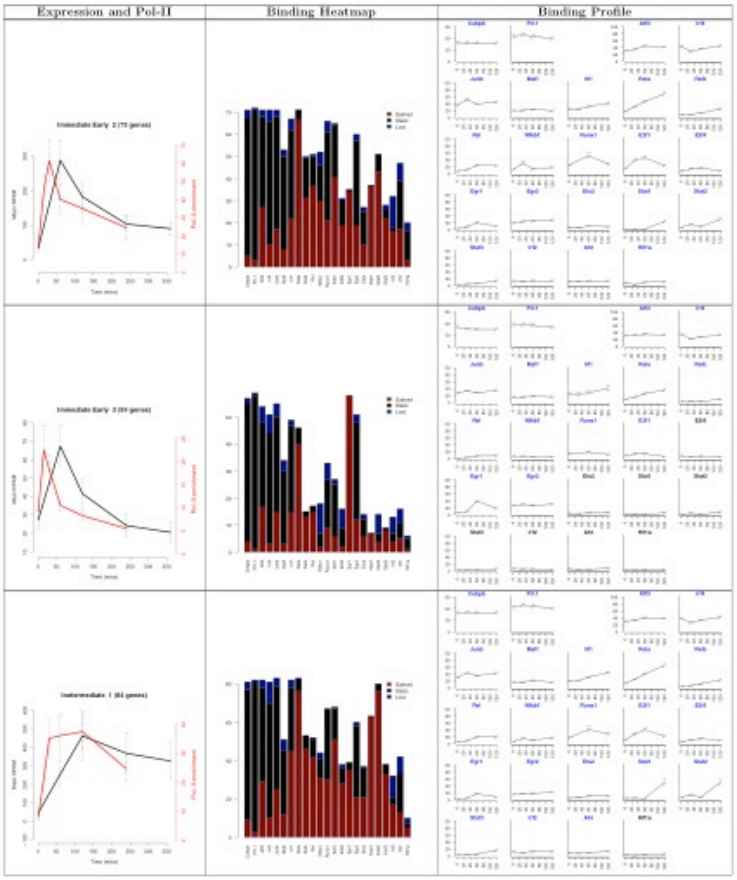


Figure S5E (part III)

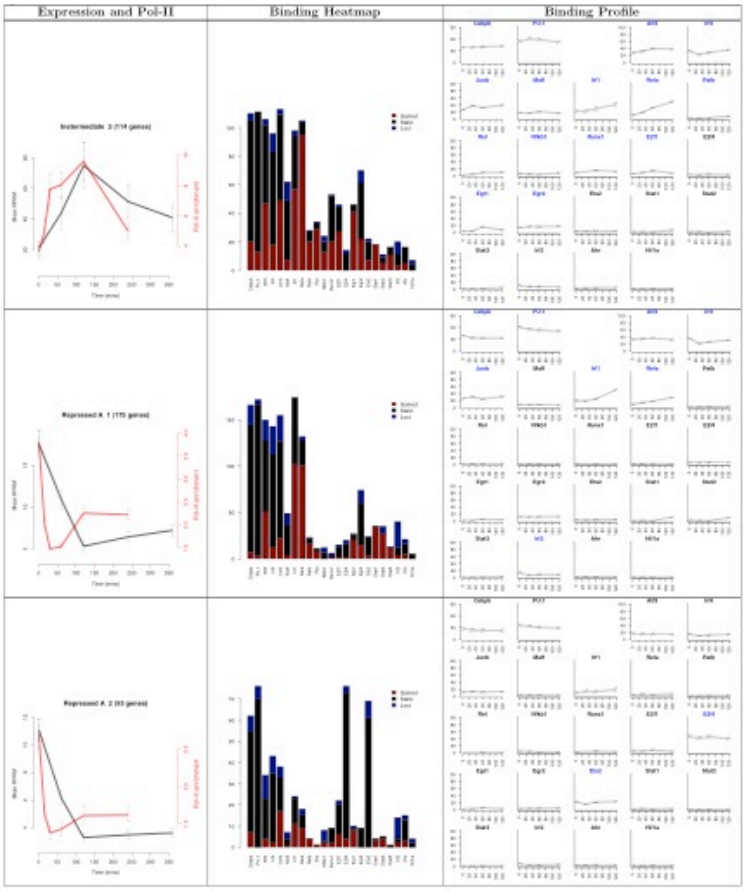


Figure S5E (part IV)

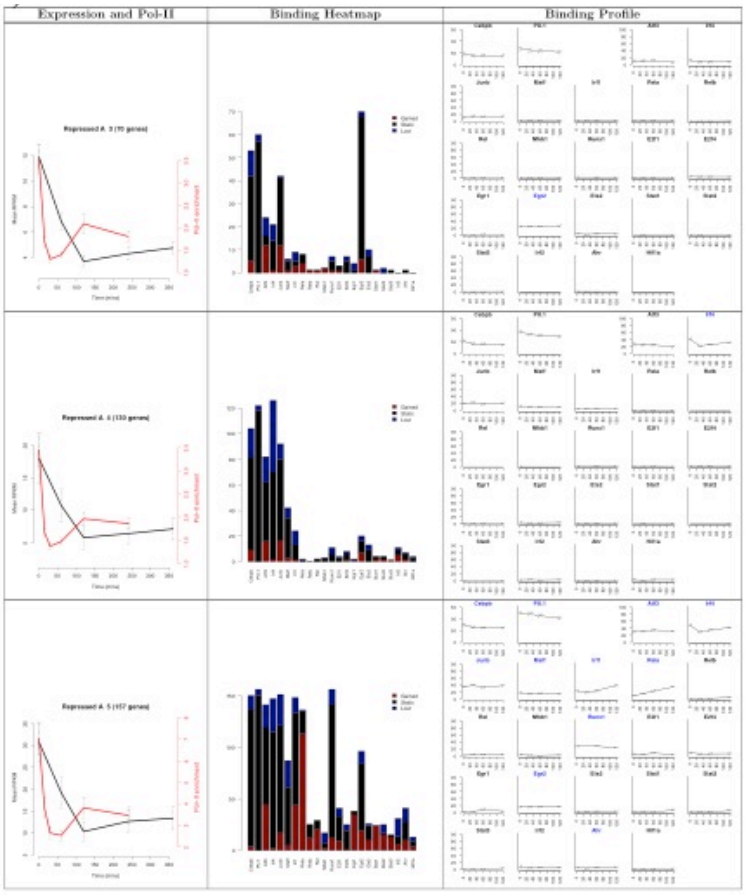


Figure S5E (part V)

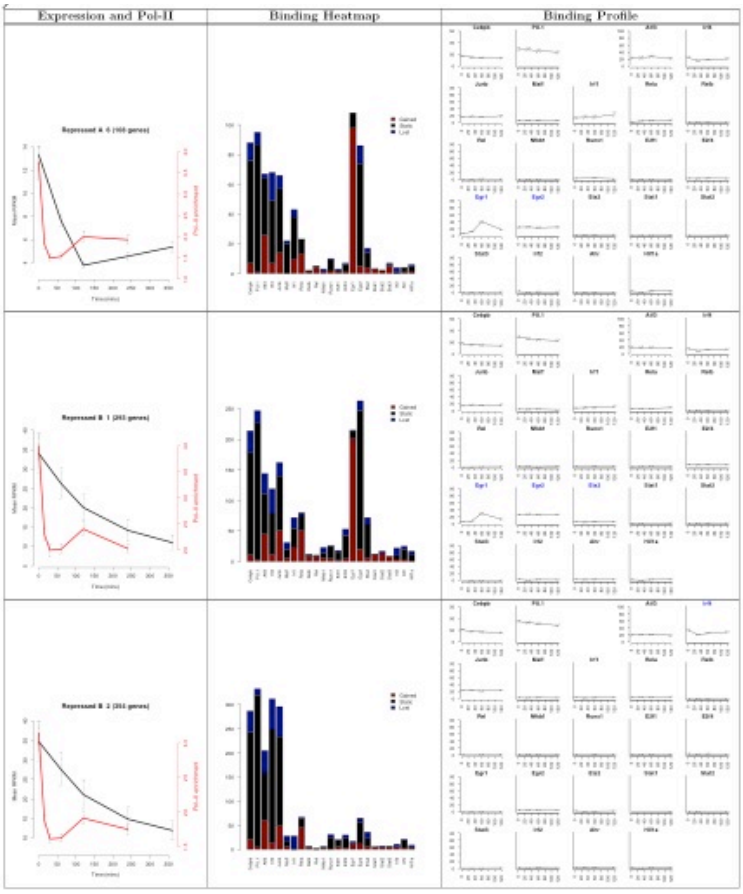


Figure S5E (part VI)

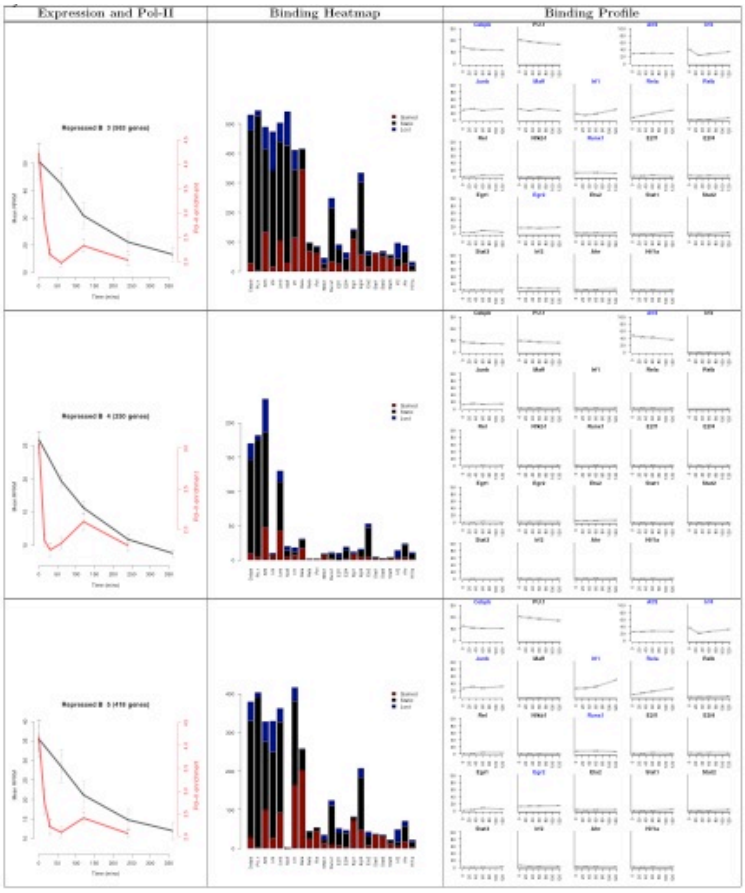


Figure S5E (part VII)

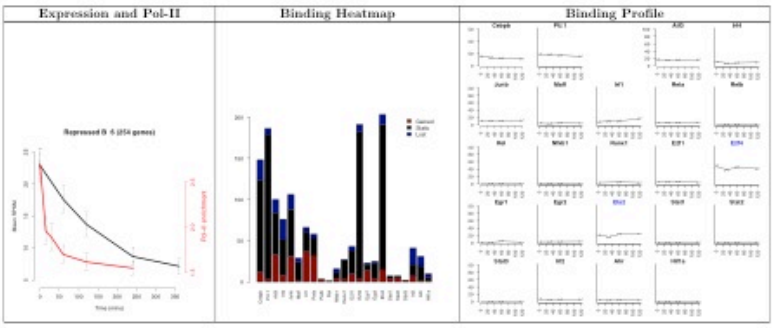


Figure S5F

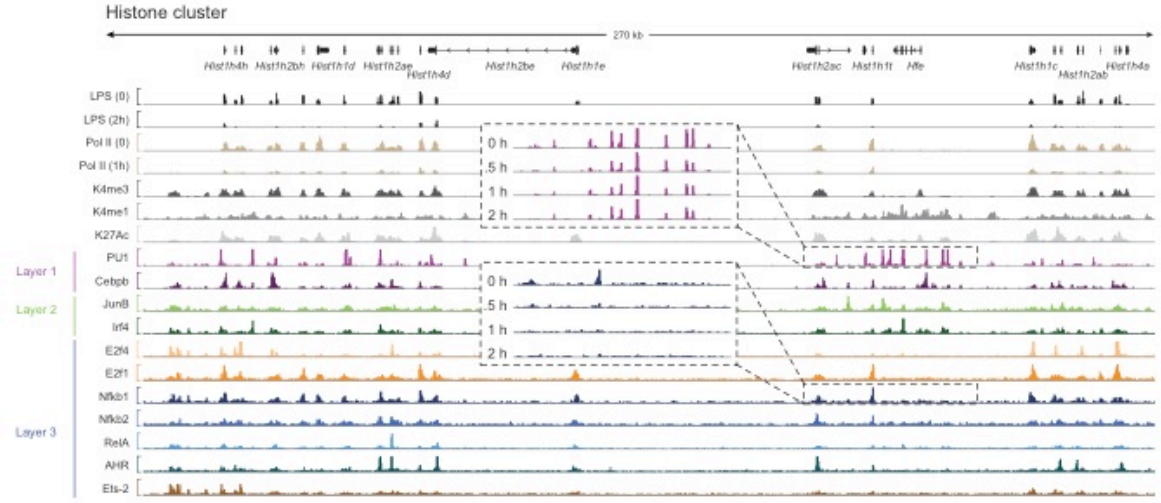
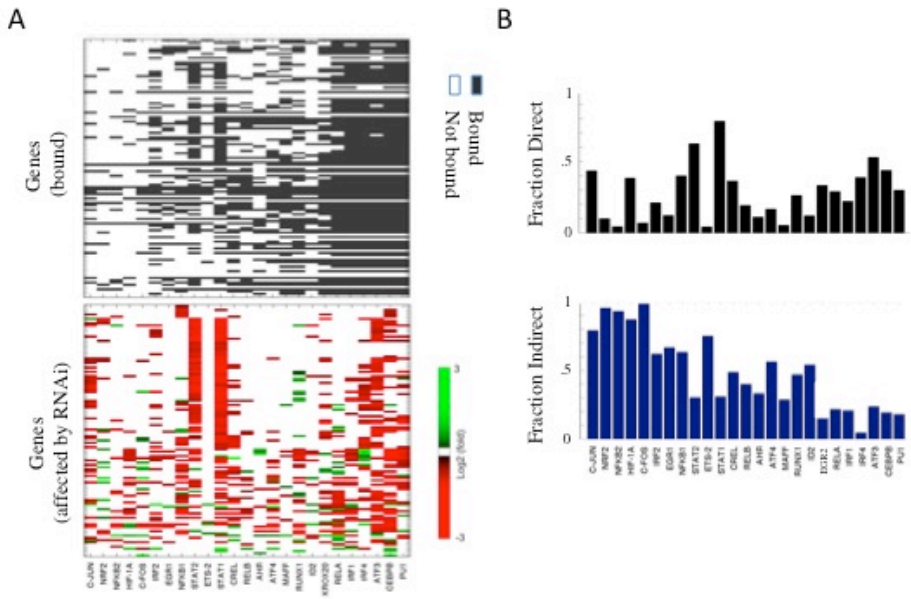

Histone cluster
270 kb
Hist1h4h
Hist1h2bh
Hist1h1d
Hist1h2ae
Hist1h4d
Hist1h2be
Hist1h1e
Hist1h2ac
Hist1h1t
Hfe
Hist1h1c
Hist1h2ab
Hist1h4a
LPS (0)
LPS (2h)
Pol II (0)
Pol II (1h)
K4me3
K4me1
K27Ac
PU1
Cebpb
JunB
Irf4
E2f4
E2f1
Nfkb1
Nfkb2
RelA
AHR
Ets-2
0 h
.5 h
1 h
2 h
Layer 1
Layer 2
Layer 3

Figure S6


A
B
Genes (bound)
Genes (affected by RNAi)
Bound
Not bound
Log2 (fold)
3
-3
Fraction Direct
Fraction Indirect
1
.5
0

Supplemental Figure Legends

Figure S1. Traditional ChIP versus HT-ChIP and timepoint selection, corresponds to main Figure 1

A. Representative genomic views (Chromosome 5:134,756,871-134,934,050) of H3K4me3 and H3K27me3 ChIP-Seq data derived from 10 million (10^7), 100,000, 50,000, 20,000 and 10,000 hESCs. **B.** Evaluation of the enrichments derived from 10,000, 20,000, 50,000 and 100,000 hESCs (Y-axis), as a function of the score of each peak in the 10 million cells dataset (X-axis). **C.** Pol-II precipitated DNA (1 ng) was split in two samples, one was prepared using the traditional ChIP-Seq protocol and the other using the HT-ChIP library production process, shown is a scatter plot of peaks scored with each library. **D.** IGV tracks showing the Zfp36 locus for the traditional ChIP-Seq (red) and HT-ChIP (green) library production process**. E.** Cell lysates from DC activated for 2 hours with LPS where subjected to Western Blotting (WB) using the indicated antibodies**. F,G.** We define a gene to be induced (repressed) if its mRNA or Pol-II occupancy levels increases (decreased) by at least 2-fold at one point after stimulation when compared to basal expression. The cumulative plots depict, for every time point (x-axis) the percentage of induced genes that have reached the 2-fold change threshold by the time point (y-axis). Panel F. was computed with the RNA-Seq data (RPKM) and panel G. with Pol-II binding data (Enrichment over background Pol-II occupancy).

Figure S2. Binding landscape and reproducibility across time, corresponds to Figure 2

A,B. Combining RNA-Seq reconstruction with HT-ChIP of chromatin marks reveals start sites, promoter regions, and enhancer regions. **A.** The Ncoa6 gene. The tracks in top-down order show: 1) Annotations in the RefSeq database (black), 2) Reconstructed transcripts using total RNA-Seq data (blue), promoters called by our annotation pipeline (gray box), the arrow points to a novel or DC specific promoter for Ncoa6. Tracks 3,4) show RNA-Seq read density plots obtained from DCs before LPS stimulation and four hours post stimulation (blue). 5,6,7) HT-

ChIP read density plots for "compressed" data for H3K4me1, H3K4m3 and K4k27Ac (gray). **B.** The Lhx2 locus. In addition to the tracks displayed in panel *a*. we also included our enhancer annotations and HT-ChIP binding data for Stat1 and PU.1. Black arrows indicate promoters associated with genes through RNA-Seq reconstructions, red arrows indicate promoters associated with unannotated transcripts reconstructed from RNA-Seq. **C-I.** Scatter plots show peaks that scored above an enrichment cutoff of 5 for chromatin marks and above a cutoff of 10 for TFs in at least one of the libraries. Red lines indicate the x=y line and the 2 fold threshold for the chromatin plots and the 3 fold threshold for the TF plots. Figure **S2F** also includes scatter plots for Stat1 biological replicates at 2 hours post stimulation. The scatter plots shown in **S2F** compares two biological replicates against the library used in the main analysis. **J.** A summary of all motifs found by applying MEME to the set of high scoring peaks for each factor. Motifs shown have an E-value < 0.01.

Figure S3. Runx1 3' end binding, corresponds to Figure 2

A. Overview of the Cxcl2 inflammatory gene loci showing time course RNA-Seq and ChIP-Seq data for selected factors, including Runx1. **B.** Comparison of the enrichment of promoter and 3' end bound Runx1 targets in inflammatory (dark gray) vs anti-viral genes (light gray). **C.** Cumulative distribution plots of Runx1 peak scores at promoters (red) and 3' end (blue). **D.** Cumulative distribution plots of the expression (in RPKM) of genes bound by Runx1 at the promoter (red) and 3' end (blue). **E.** Median score of Runx1 peaks at the promoter (red), 3' end (blue) and enhancer (green) across the LPS response time course. **F.** Expression fold change in Runx1 knockdown DC 6hr post stimulation compared to non- targeting shRNA (Amit et al., 2009) of significantly down and up regulated genes. Blue and red starts indicate genes that are bound by Runx1 at the 3' end and promoter respectively.

Figure S4. Distribution of TF binding in regulatory regions and robustness of dynamic behaviour to different thresholds, corresponds to Figure 4

A. The average TF complexity (#bound TF) of bound regions in the original data (red bars) and in the randomized data (blue bars, See Supplemental text). **B**. For each complexity value (x-axis), the figure depicts the expected ratio of regions with this complexity (y-axis). Plots are shown for the original data (red), and the randomized data (blue). **C.** The percentage of bound regions that have a complexity of 1 - observed (red asterisk) and expected from the randomized

data (blue bars). **D.** The percentage of dynamic changes in binding (gain or loss) with different fold cutoffs. For each TF *f* and each cutoff level *x* the figure depicts the percentage of regions (promoters or enhancers) bound by *f* that have more than *x*-fold change (up/down) in their binding enrichment score, in at least one time point, compared to the basal state.

Figure S5. Combined analysis of gene expression and TF binding data, corresponds to Figure 5

A. Fold enrichment of TF binding at induced genes during the un-stimulated state. **B.** The percentage of induced genes bound during the un-stimulated state by each factor. **C.** Fold enrichment of TF binding gain at induced genes. **D.** The percentage of induced genes associated with gain of binding by each factor. For panels c and d we only consider cases where the change in binding co-occurs or precedes the change in expression (Methods). Only significant results ($p<10^{-3}$) are presented.. **E.** table showing profiles of differentially expressed genes clustered by both expression and binding. Each row shows three plots for each cluster. First, mean expression (RNA-Seq, black) and Pol-II occupancy (red) for genes in the cluster. Second, percent of genes in the cluster bound by the factor (in the x-axis) each bar is broken by static (black), gain (red) or loss (red) of binding. Third, the mean binding enrichment scores of genes in the cluster for each TF we ChIPed. TF whose binding is significantly enriched in the cluster ($p<10^{-3}$; Methods) are labeled in blue**. F.** IGV browser tracks for the Histone gene cluster for the indicated sequencing libraries (RNA-Seq, Pol-II, Chromatin marks, and various TFs as indicated) and time points.

Figure S6. TF knockdown effects on bound genes, corresponds to Figure 6

A. Top. Binary matrix indicating TF binding at the signature immune genes described in (Amit et al., 2009). Bottom. Heat map showing expression in DCs infected with TF targeting shRNAs (labels of the TF are shown below) compared to DC infected with control shRNA 6 hours post stimulation ((Amit et al., 2009) Supplemental text). **B.** Top: Percent of functional binding: for each TF we present the number of bound genes that are affected by its knockdown divided by the total number of bound genes. Bottom: percent of indirect effect. For each TF we present the number of non-bound genes that are affected by its knockdown divided by the total number of affected genes The analysis is limited to the signature set of genes (Amit et al., 2009)

Supplemental Tables

Table S1. Design of ChIP experiments

Sheet #1 (TF list): List of all TF expresses in DC (using RNA-Seq). Sheet #2 (Antibody list): All purchased antibodies for TF expressed in DC . Sheet #3 (ChIP-String): Data collected for antibody validation using ChIP-String. Sheet #4 (ChiPed factors): list of the ChIPed factors and example of known validated targets.

Table S2. Peaks overlap with regions

Sheet #1 (Protein binding): the overlap between pairs of ChIP assays. Each cell correspond to a pair of assays, say *x* (column) and *y* (row). The displayed values are:
(i) p-value of overlap (computed as a binomial score), (ii) the number of x peaks that fall close (500bp) to y peaks, (iii) the percentage of x peaks that fall close (500bp) to y peaks, (iv) fold enrichment. Sheet #2 (Genomic regions): the overlap between ChIP assays and annotated genomic regions. The displayed values are: (i) p-value of overlap (computed as a binomial score), (ii) the number of x peaks that fall inside the region (taking a 50bp margin), (iii) the percentage of x peaks that inside the region, (iv) fold enrichment.

Table S3. Clusters memberships and functional enrichment

Sheet #1 (Expression& binding clusters): list of the genes in each of the 19 Expression& binding clusters; Sheet #2 (Expression& binding - enrichment): - functional enrichment analysis using the MsigDB database (Methods). Sheet #3 (Binding data only - clusters): list of the genes in each of the 8 clusters computed based on binding data only; Sheet #4 (Binding data only-enrichment): - functional enrichment analysis. Sheet #5 (HOT regions - enrichment): functional enrichment analysis of genes associated with HOT regions. Sheet #6 (Genes with many or few bound TF): functional enrichment analysis of genes associated with few (<5) or many (>15) bound TF.

Table S4. Enrichment Gene Tfs

Sheet #1 (Functional enrichment): Functional enrichment analysis of genes that are bound by specific TFs. We perform three separate analyzes for each TF: (i) enrichment of genes bound in

promoters, (ii) enrichment of genes bound in enhancers, (iii) enrichment of genes bound in either promoter or enhancer. Sheet #2 (Bacterial or viral genes): Enrichment of TF binding at inflammatory or anti-viral genes (Amit et al.). Only significant results ($p<10^{-3}$; Methods) are presented.

Table S5. TF hierarchy

For each pair of TF X (row) and Y(column), the table presents the percentage of regions (promoter or enhancer) bound by X that are also bound by Y.

Table S6. RNA-Seq expression

Gene expression using RNA-Seq (from polyA selected total RNA and from 4SU labeled RNA) with RPKM values computed on a constituent exon set when genes have multiple isoforms.

Experimental Procedures

HT-ChIP

ChIP module

20 million dendritic cells were used for each ChIP experiment. Cells were fixed for 10 min with 1% formaldehyde, quenched with glycine and washed with ice-cold PBS and pellets where flash frozen in liquid nitrogen. Cross-linked DC where thawed on ice and resuspended in RIPA lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 14 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC) supplemented with protease inhibitor (Roche, 04693159001). Cells were lysed for 10 min on ice and the chromatin was sheared according to the calibrated conditions for DC cells using a Branson Sonifier (model S-450D) with a custom sample cooling system (sample holder-Mecour; #99-401, CB-LS00-60/24, Chiller-Thermo; RTE-7 D1). 96 well magnets were used (Invitrogen) in all further steps for seperation. The sonicated cell lysate (Whole Cell Extract) was cleared by centrifugation and mixed in 96 well plates with 75 ul of protein G magnetic dynabeads (Invitrogen) coupled to target antibody and incubated over night at 4 degrees. For the coupling, beads were washed once (200ul) in a binding/blocking buffer (PBS, 0.5% Tween 20, 0.5% BSA), incubated in 96 wells with 10 ug of antibody in binding/blocking buffer for 1 hour at room temperature, and then washed to remove excess antibody. Using the 96 well magnets, cell lysate was removed and samples was washed 5 times with cold RIPA (200ul per wash), twice with RIPA buffer supplemented with 500 mM NaCl (200ul per wash), twice with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA), and then eluted in 50 ul of 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0. The eluate was reverse crosslinked at 65C for 4 hours and then treated sequentially with 2ul of RNaseA (Roche, 11119915001) for 30 min and 2.5 ul of Proteinase K (NEB, P8102S) for two hours.

Library construction module

Solid-phase reversible immobilization (SPRI) cleanup steps were performed using the Bravo liquid handling platform (Agilent) using a modified version of (Fisher et al., 2011). 120ul SPRI AMPure XP beads (Agencourt) were added to the reverse-crosslinked samples, pipette-mixed 15 times and incubated for 2 minutes. Supernatant were separated from the beads using a 96-well magnet for 4 minutes. Beads were washed on the magnet with 70% ethanol and then air dried for 4 minutes. The DNA was eluted in 40 ul EB buffer (10 mM Tris-HCl pH 8.0) by pipette mixing 25 times. For the remainder of the library construction process (DNA end-repair, A-base addition, adaptor ligation and enrichment) a general SPRI cleanup involves addition of buffer containing 20% PEG and 2.5 M NaCl to the DNA reaction products (without moving them from their original well position). After thorough mixing and a 2-minute incubation at room temperature, plates are transferred to a magnet plate, incubated for 4 minutes and supernatant removed. Beads are then washed twice on the magnet with 150ul 70% ethanol and then air dried for 4 minutes. The DNA is eluted with 40ul of EB buffer by pipette mixing 25 times. All enzymatic steps are carried out using enzymes from New England Biolabs. The DNA end-repair was performed by adding 5 ul T4 PNK enzyme, 5 µl T4 polymerase (3 units) to each well, incubated at 12C for 15 min, 25C for 15 min.. Following SPRI bead clean up A-base addition was performed by adding 3 µl Klenow (3'->5' exonuclease) to each well and incubated at 37C for 30 min. SPRI bead clean up method was used to purify the product. Adaptor ligation was performed by adding 5 µl DNA ligase and 5 µl PE Indexed oligo adaptors (0.75 uM) samples were incubated 25C for 15 min. SPRI bead clean up with size selection was used to purify the ligated products (0.7X SPRI/reaction). Finally, enrichment PCR was performed by adding 10 µl of a master mix (2 µl Forward/Reverse Index Primer, 0.5 µl dNTP mix, 5 µl 10x Pfu Ultra Buffer, 1 µl Pfu Ultra-II Fusion, 1.5 µl Nuclease free water) to each well and amplified at 95C for 2 min, 14 cycles of: 95C for 30 sec, 55C for 30 sec, 72C for 60 sec, and 72C for 10 min. The final SPRI clean up coupled to size selection was performed (0.7X SPRI/reaction). For a detailed Automated HT-ChIP setup procedure and the Bravo liquid handling platform see Supplemental text and http://www.weizmann.ac.il/immunology/AmitLab/data-and-method/HT-ChIP/.

Automated HT-ChIP Library construction module

This part of the pipeline uses directly from the 96 well ChIP plate, 40 ul of the ChIPed material immediately after reverse cross-linking RNAse and Proteinase K treatment.

End Repair Mastermix Preparation

1| Thaw the reagents on ice.

2| Once the reagents have thawed, prepare the appropriate amount of mastermix for the samples plus an additional 10 samples worth of reagent to account for dead volume as detailed in Table 1.

Table 1: Reagents Used for End repair Mastermix Preparation

Reagent Name	1 RXN (µL)	30 RXN (µL)	115 RXN (µL)
NEB Buffer 2	5	150	575
10 mg/ml BSA	0.5	15	57.5
25 mM ATP	2	60	230
100 mM dNTP Mix	0.2	6	23
Nuclease Free Water	9.3	279	1069.5
T4 Polynucleotide Kinase	5	150	575
T4 DNA Polyermase	5	150	575
Total Volume	***27***	***810***	***3105***

3| Once the reagents have been combined, gently mix the mastermix, then place back on ice.

End Repair Automated Mastermix Dispense Protocol

1| Set head mode to 1 column: 12.

2| Pick up clean 70µl ST V11 Tips from quadrant 1 in column one from a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck. (Tips only need to be present in quadrants 1 and 2 of each position in which a sample is located).

3| Aspirate 27 µl of End Repair Mastermix from the 5ml Deerac disposable reservoir located in column 3 of the low volume insert holder located at position 7 on the Agilent Bravo deck.

4| Dispense the 27µl of End Repair Mastermix into samples located in Column one of the sample plate located at position 8 of the Agilent Bravo.

5| Tips are knocked off for disposal at quadrant 1 in column one of an empty 70 µl ST V11.

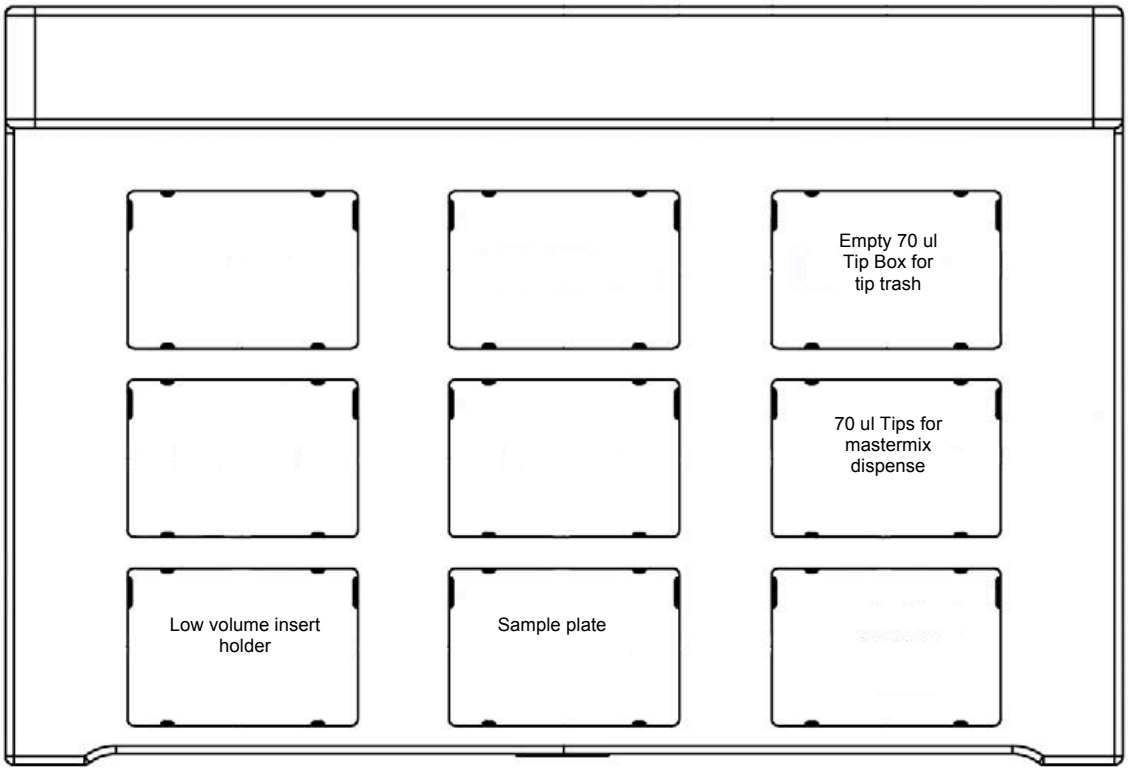
6| Repeat steps 1 through 3 for all subsequent columns on the sample which contain samples. Clean tips should be used each time mastermix is aliquotted into a new column on the sample plate. For column 1, put the tips on in column 1 of the tip box and off in column 1 of the tip trash. For column 2 of the sample plate, put the tips on in column 3 of the tip box and off in column 3 of the tip trash. For column 3 of the sample plate, put the tips on in column 5 of the tip box and off in column 5 of the tip trash. And so on.

7| Pick up clean 70 µl ST V11 Tips from quadrant 2 of a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck.

8| Perform a Dual Height Mix on the wells containing sample and mastermix. Aspirate 40 µl at a height of 1 mm from the bottom of the well and dispensing 40 ml from the bottom of the well and dispensing at a height of 5 mm. Mix approximately 15 times.

9| Knock off tips for disposal into each quadrant 2 of an empty 70µl St V11.

10| Once the protocol is complete, seal wells containing sample with ABI optical caps and place the sample plate on thermocycler. (Thermoprofile consist solely of an initial incubation of 12°C for 15 minutes followed by 12°C for 15 minutes then a held at 4°C indefinitely).



Set up of the Agilent Bravo with LT head for End Repair 2.2X Cleanup

(Figure 3A displays deck layout)

Process Steps automated on the Bravo

1| Put on 180 µl tips from tip box # 1.

2| Aspirate 147.4 µl of 20% PEG 2.5M NaCl from the 20%PEG 2.5M NaCl source plate, and dispense into sample plate.

3| Perform a Dual Height Mix to ensure the AMPure XP beads are properly resuspended in 20% PEG 2.5 M NaCl buffer. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 13 mm from the bottom of the well. Mix approximately 130 µl 12 times.

4| Allow the sample plate to sit for 2 minutes, after which time place the sample plate, on the Dynal MPC – 96 S plate magnet for 4 minutes to allow the AMPure XP beads to separate from the solution.

5| Remove and discard the supernatant into the 20% PEG 2.5M NaCl source plate.

6| Discard the used 180 µl tips into tip box # 1.

7| Put on 180 µl tips from tip box # 2.

8| Leaving the sample plate on the Dynal MPC-96S magnet plate, aspirate 100 µl of 70% EtOH and dispense into the sample plate. DO NOT MIX.

9| Allow the AMPure XP beads and sample sit in the 70% EtOH for 30 seconds, then remove the EtOH and discard into the 70% EtOH source plate.

10| Discard tips into 180 µl Tip Box # 2.

11| Move the sample plate off of the Dynal MPC-96s magnet plate and allow the sample-AMPure bead complex to air dry for approximately 4 minutes at room temperature.

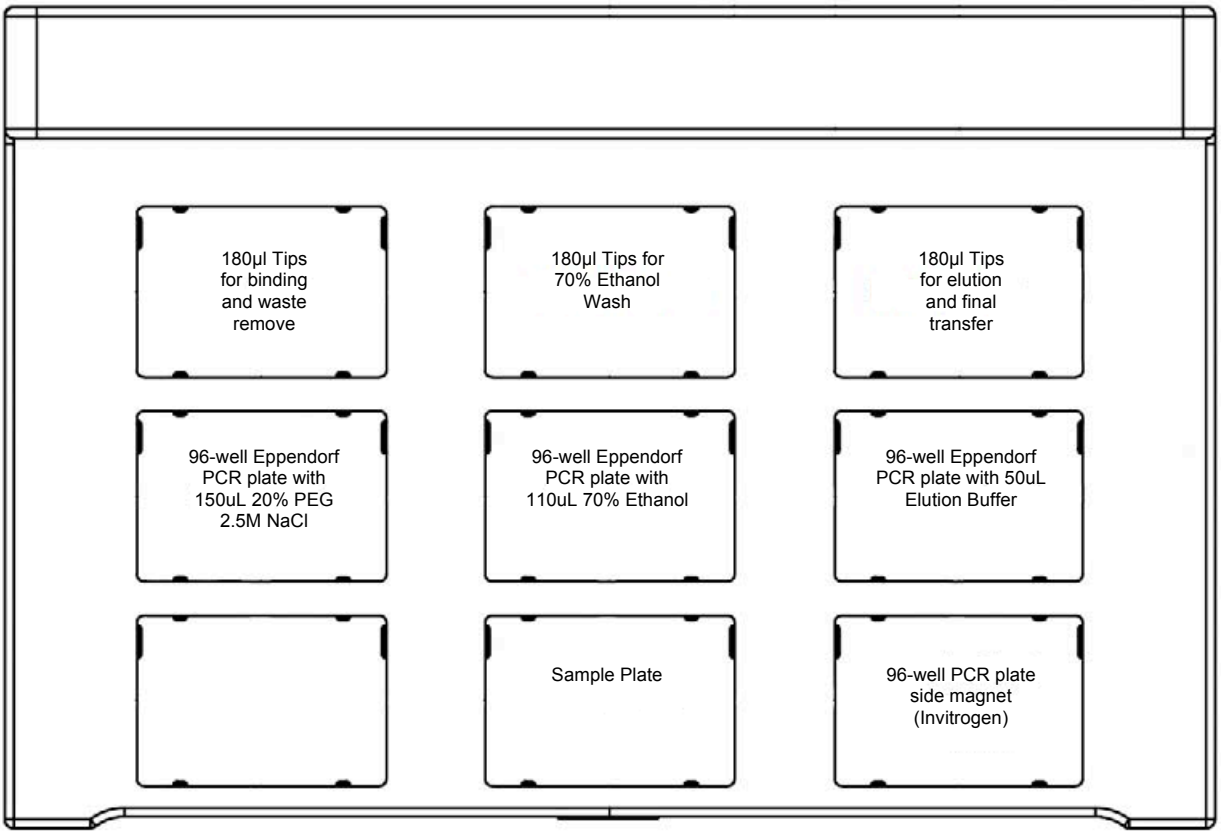
12| Put on 180 µl tips from tip box # 3.

13| Aspirate 40µl of Tris-HCl pH 8.0 and dispense into sample plate.

14| Perform a Dual height mix. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 6 mm from the bottom of the well. Mix approximately 40 µl 15 times.

15| Discard used 180 µl tips into tip box # 3.

16| Using ABI optical caps, seal Eppendorf plate containing samples.



A Base Addition Mastermix Preparation

1| Thaw the reagents on ice.

2| Once the reagents have thawed, prepare the appropriate amount of mastermix for the samples plus an additional 10 samples to account for dead volume as detailed in Table 2.

Table 2: Reagents Used for A Base Mastermix Preparation

Reagent Name	1 RXN (µL)	30 RXN (µL)	115 RXN (µL)
NEB Buffer 2	11.9	357	1368.5
100 mM dATP	5	150	575
Nuclease Free Water	0.1	3	11.5
Klenow exo-	3	90	345
Total Volume	**20**	**600**	**2300**

3| Once the reagents have been combined, gently mix the mastermix, then place back on ice.

Automated A Base Addition Mastermix Dispense Protocol

1| Set head mode to 1 column: 12.

2| Pick up clean 70µl ST V11 Tips from quadrant 1 in column one from a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck. (Tips only need to be present in quadrants 1 and 2 of each position in which a sample is located).

3| Aspirate 20 µl of A Base Addition Mastermix from the 5ml Deerac disposable reservoir located in column 3 of the low volume insert holder located at position 7 on the Agilent Bravo deck.

4| Dispense the 20 µl of A Base Addition Mastermix into samples located in Column one of the sample plate located at position 8 of the Agilent Bravo.

5| Tips are knocked off for disposal into each quadrant 1 in column one of an empty 70 µl St V11.

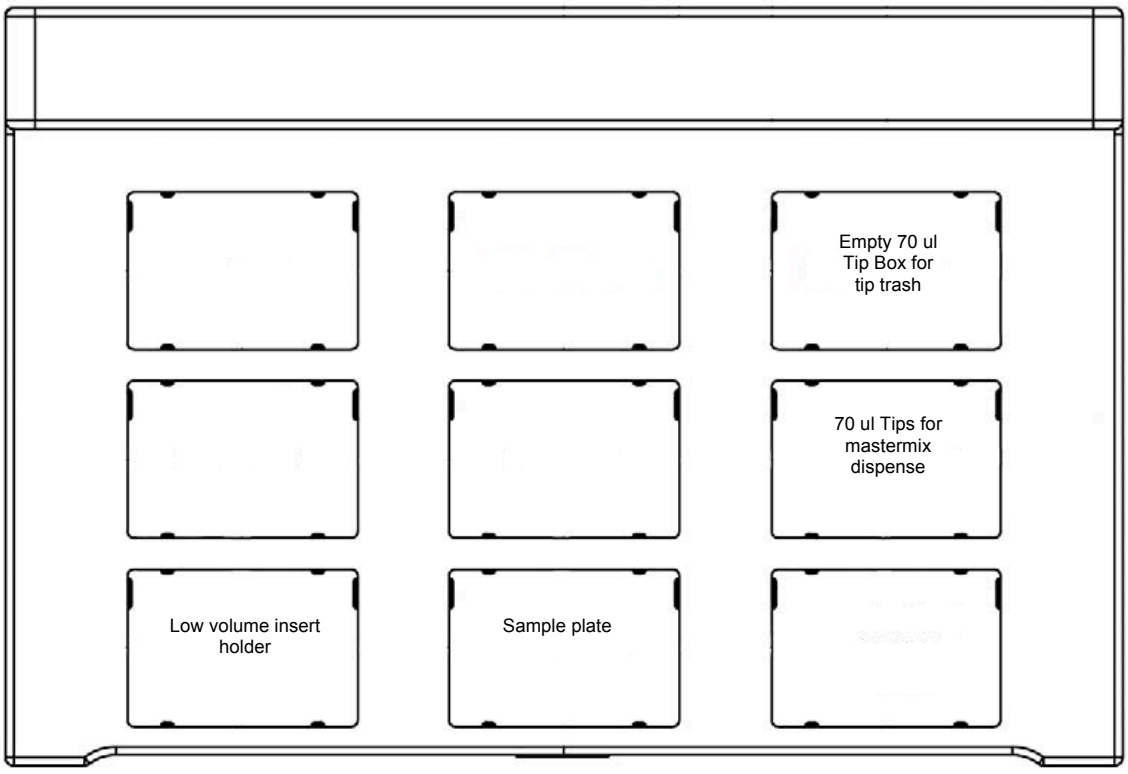
6| Repeat steps 1 through 3 for all subsequent columns on the sample which contain samples. Clean tips should be used each time mastermix is aliquotted into a new column on the sample plate. For column 1, put the tips on in column 1 of the tip box and off in column 1 of the tip trash. For column 2 of the sample plate, put the tips on in column 3 of the tip box and off in column 3 of the tip trash. For column 3 of the sample plate, put the tips on in column 5 of the tip box and off in column 5 of the tip trash. And so on.

7| Pick up clean 70 µl ST V11 Tips from quadrant 2 of a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck.

8| Perform a Dual Height Mix on the wells containing sample and mastermix. Aspirate 40 µl at a height of 1 mm from the bottom of the well and dispensing 40 ml from the bottom of the well and dispensing at a height of 5 mm. Mix approximately 15 times.

9| Knock off tips for disposal into quadrant 2 of an empty 70 µl St V11.

10| Once the protocol is complete, seal well containing sample with ABI optical caps and place the sample plate on thermocycler. (Thermoprofile consists solely of 37°C for 30 minutes then held at 4°C indefinitely).



Set up of the Agilent Bravo with LT head for A Base Addition 2.2X Cleanup

(Figure 3A displays deck layout)

Process Steps automated on the Bravo

1| Put on 180 µl tips from tip box # 1.

2| Aspirate 132 µl of 20% PEG 2.5M NaCl and dispense into sample plate.

3| Perform a Dual Height Mix to ensure the AMPure XP beads are properly resuspended in 20% 2.5 M NaCl buffer. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 13 mm from the bottom of the well. Mix approximately 130 µl 15 times.

4| Allow the sample plate to sit for 2 minutes, after which time place the sample plate, on the Dynal MPC – 96S plate magnet for 4 minutes to allow the AMPure XP beads to separate from the solution.

5| Remove and discard the supernatant into the 20% PEG 2.5M NaCl source plate

6| Discard the used 180 µl tips into tip box # 1.

7| Put on 180 µl tips from tip box # 2.

8| Leaving the sample plate on the Dynal MPC-96S magnet plate, aspirate 100 µl of 70% EtOH and dispense into the sample plate. DO NOT MIX.

9| Allow the AMPure XP beads and sample sit in the 70% EtOH for 30 seconds, then remove the EtOH and discard into the 70% EtOH source plate.

10| Discard tips into 180 µl Tip Box # 2.

11| Move the sample plate off of the Dynal MPC-96S magnet plate and allow the sample-AMPure bead complex to air dry for approximately 4 minutes at room temperature

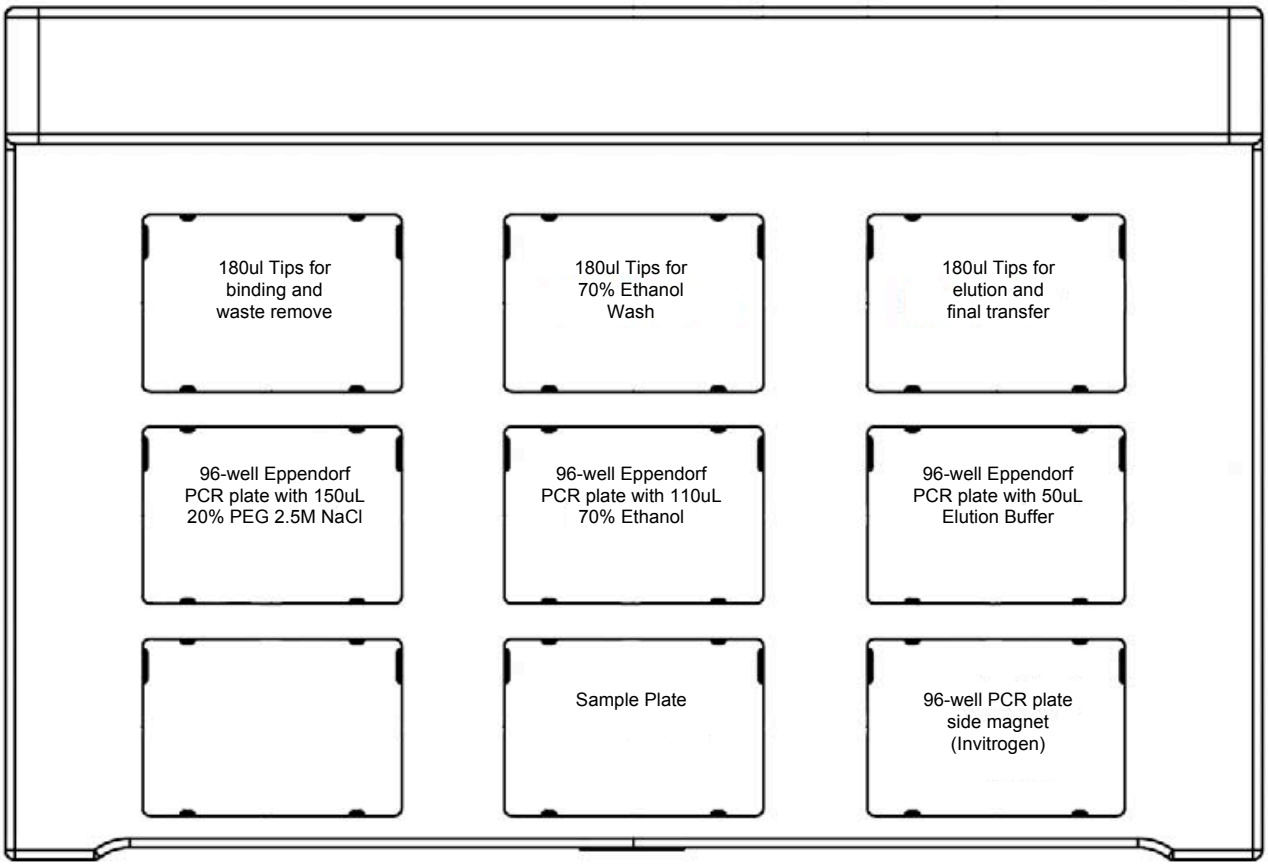
12| Put on 180 µl tips from tip box # 3.

13| Aspirate 40µl of Tris-HCl pH 8.0 and dispense into sample plate.

14| Perform a Dual height mix . Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 6 mm from the bottom of the well. Mix approximately 40 µl 15 times.

15| Discard used 180 µl tips into tip box # 3.

16| Using ABI optical caps, seal Eppendorf plate containing samples.



Adapter Ligation Mastermix Preparation

1| Thaw reagents on ice. Thaw enough for the number of samples being run plus an extra 15 samples for to account for dead volume.

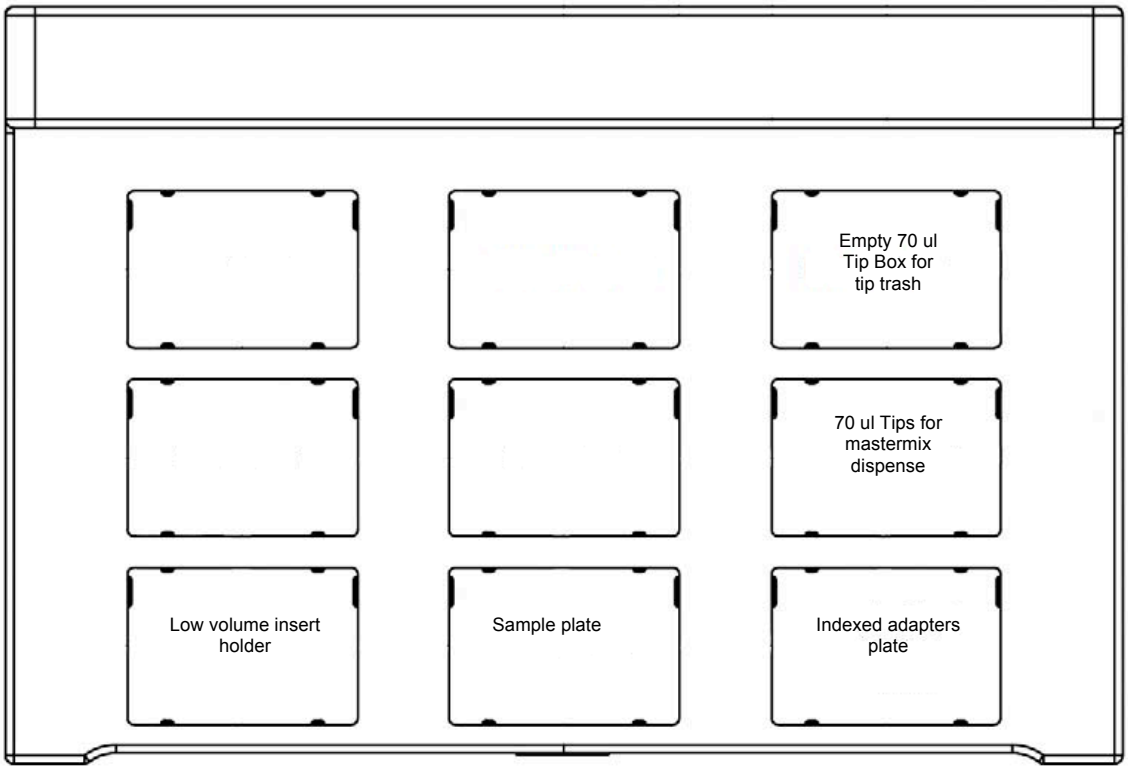
2| Prepare Adapter Ligation Mastermix as described in Table 3.

Table 3: Reagents used for Adapter Ligation mastermix preparation

Reagent Name	1 RXN (µL)	30 RXN (µL)	120 RXN (µL)
2x DNA Ligase Buffer	29	870	3480
DNA Ligase	5	150	600
Total Volume	34	1020	4080

3| Thaw the Indexed adapter plate, with each well containing at least 6 µl of a unique adapter, at room temperature.

4| Once thawed vortex the Indexed adapter plate at a moderate speed followed by a quick spin down.



Automated Adapter Ligation Mastermix Dispense Protocol

1| Set head mode to 1 column:12.

2| Pick up clean 70 µl ST V11 Tips from position 1 of each quadrant in column one from a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck. (Tips only need to be present in quadrants 1 and 2 of each position in which a sample is located)

3| Aspirate 34 µl of Adapter Ligation Mastermix from the 5ml Deerac disposable reservoir located in column 3 of the low volume insert holder located at position 7 on the Agilent Bravo deck.

4| Dispense the 34 µl of Adapter Ligation Mastermix into samples located in Column one of the sample plate located at position 8 of the Agilent Bravo.

5| Knock off tips for disposal into quadrant 1 in column one of an empty 70 µl St V11.

6| Repeat steps 1 through 3 for all subsequent columns on the sample which contain samples. Clean tips should be used each time mastermix is aliquotted into a new column on the sample plate. For column 1, put the tips on in column 1 of the tip box and off in column 1 of the tip trash. For column 2 of the sample plate, put the tips on in column 3 of the tip box and off in column 3 of the tip trash. For column 3 of the sample plate, put the tips on in column 5 of the tip box and off in column 5 of the tip trash. And so on.

7| Pick up clean 70 µl ST V11 Tips from quadrant 2 of a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck.

8| Aspirate 6 µl of adapter from the Indexed adapter plate located at position 9 on the Agilent Bravo.

9| Dispense 6 µl of adapter into the corresponding wells of the sample plate. NOTE: Do not discard tips. They will be used to mix sample.

10| Perform a Dual Height Mix on the wells containing sample and mastermix. Aspirate 40 µl at a height of 1 mm from the bottom of the well and dispensing 40 ul from the bottom of the well and dispensing at a height of 5 mm. Mix approximately 15 times.

11| Knock off tips for disposal into quadrant 2 of an empty 70 µl St V11.

12| Once the protocol is complete, seal wells containing sample with ABI optical caps and place the sample plate on thermocycler. (Thermoprofile consists solely of 25°C for 15 minutes then held at 4°C indefinitely).

Set up of the Agilent Bravo with LT head for Adapter Ligation 0.7X Cleanup

Process Steps automated on the Bravo

1| Put on 180 µl tips from tip box # 1.

2| Aspirate 40.6 µl of 20% PEG 2.5M NaCl and dispense into sample plate.

3| Perform a Dual Height Mix to ensure the AMPure XP beads are properly resuspended in the 20% PEG 2.5 M NaCl buffer. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 7 mm from the bottom of the well. Mix approximately 80 µl 15 times.

4| Allow the sample plate to sit for 2 minutes, after which time place the sample plate on a Dynal MPC – 96 S plate magnet for 4 minutes to allow the AMPure XP beads to separate from the solution.

5| Remove and discard the supernatant into the 20% PEG NaCl 2.5M source plate.

6| Discard the used 180 µl tips into tip box # 1.

7| Put on 180 µl tips from tip box # 2

8| Leaving the sample plate on the Dynal MPC-96S magnet plate, aspirate 100 µl of 70% EtOH and dispense into the sample plate. DO NOT MIX.

9| Allow the AMPure XP beads and sample sit in the 70% EtOH for 30 seconds, then remove the EtOH and discard into the 70% EtOH source plate.

10| Discard tips into 180 µl Tip Box # 2.

11| Move the sample plate off of the Dynal MPC-96S magnet plate and allow the sample-AMPure bead complex to air dry for approximately 4 minutes at room temperature.

12| Put on 180 µl tips from tip box # 3.

13| Aspirate 40µl of Tris-HCl pH 8.0 and dispense into sample plate.

14 Perform a Dual height mix. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 6 mm from the bottom of the well. Mix approximately 40 µl 15 times.

15| Allow the resuspended sample to sit for approximately 2 minutes.

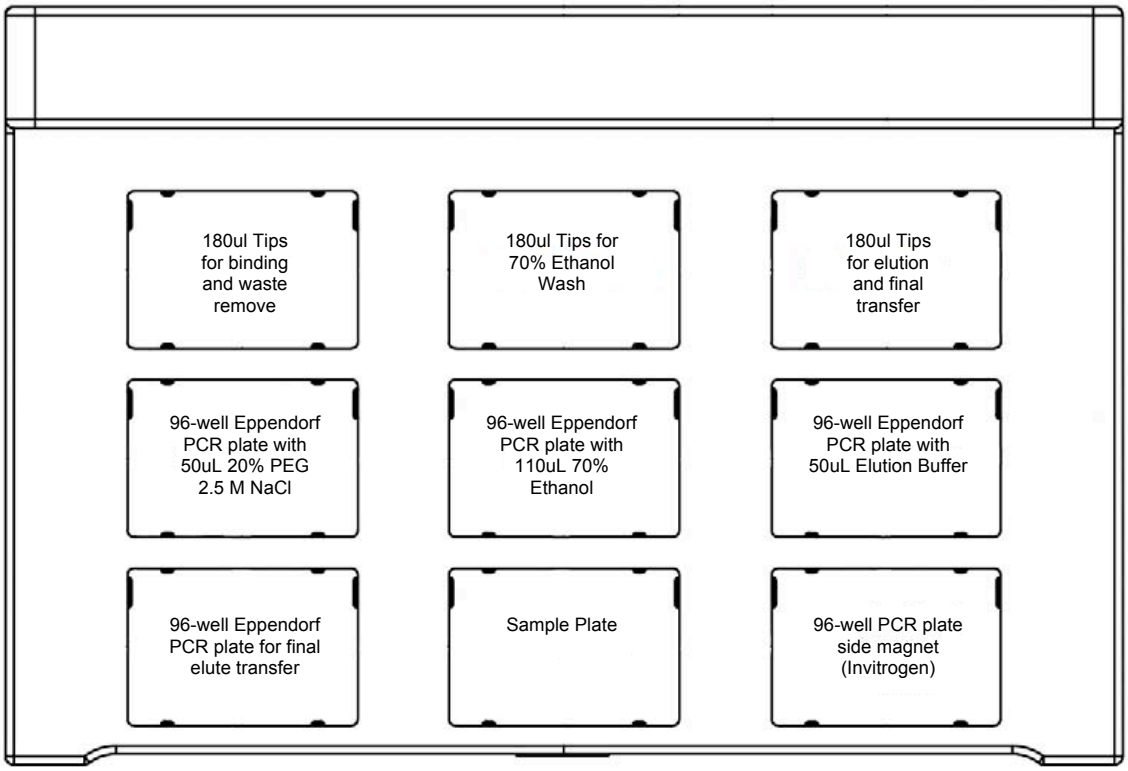
16| Place the sample plate, on a Dynal MPC – 96S plate magnet for 3 minutes to allow the AMPure XP beads to separate from the solution.

17| Aspirate the eluate and dispense into the Eppendorf 96 well twin.tec plate located at position 7.

18| Discard tips into tip box # 3.

19| Using ABI optical caps, seal Eppendorf plate containing samples.

20| Proceed to Automated/Manual Pond Enrichment Master Mix addition.



Enrichment Mastermix Preparation

1| Thaw the reagents on ice.

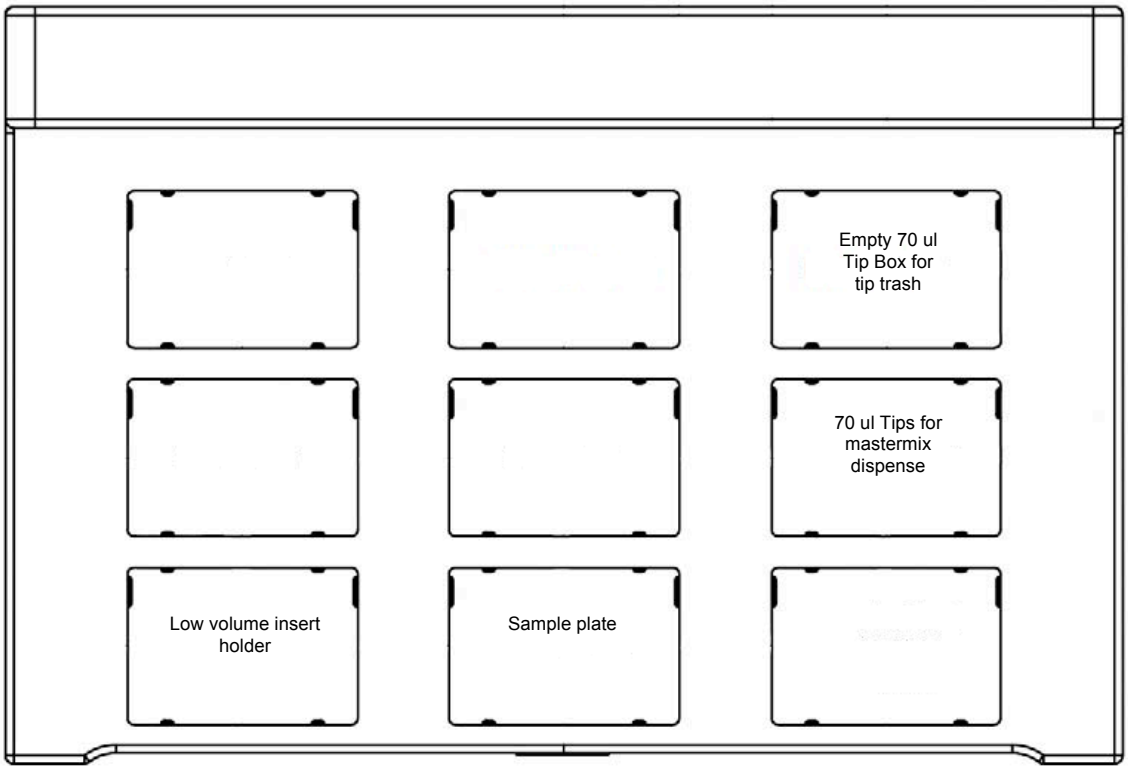
2| When the reagents have thawed, prepare the appropriate amount of mastermix for the samples plus an additional 10 samples to account for dead volume as detailed in Table 4.

Table 4: Reagents Used for Pond Enrichment Mastermix Preparation

Reagent Name	1 RXN (L)	30 RXN (L)	115 RXN (L)
10x Pfu Ultra Buffer	6	180	690
PE Primer 1.0	2	60	230
PE Primer 2.0	2	60	230
100 mM dnTP	1	30	115
Pfu Ultra Fusion HS DNA Polymerase	2	60	230
Nuclease Free Water	7	210	805
Total Volume	**20**	**600**	**2300**

3| Once

the reagents have been combined, gently mix the mastermix, then place back on ice.



Pond Enrichment Automated Mastermix Dispense Protocol

1| Set head mode to 1 column: 12.

2| Pick up clean 70 µl ST V11 Tips from quadrant 1 in column one from a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck. (Tips only need to be present in quadrants 1 and 2 of each position in which a sample is located).

3| Aspirate 20 µl of Pond Enrichment Mastermix from the 5ml Deerac disposable reservoir located in column 3 of the low volume insert holder located at position 7 on the Agilent Bravo deck.

4| Dispense the 20 µl of Pond Enrichment Mastermix into samples located in Column one of the sample plate located at position 8 of the Agilent Bravo.

5| Knock off tips for disposal into quadrant 1 in column one of an empty 70 µl St V11.

6| Repeat steps 1 through 3 for all subsequent columns on the sample which contain samples. Clean tips should be used each time mastermix is aliquotted into a new column on the sample

plate. For column 1, put the tips on in column 1 of the tip box and off in column 1 of the tip trash. For column 2 of the sample plate, put the tips on in column 3 of the tip box and off in column 3 of the tip trash. For column 3 of the sample plate, put the tips on in column 5 of the tip box and off in column 5 of the tip trash. And so on.

7| Pick up clean 70 µl ST V11 Tips from quadrant 2 of a clean 70 µl ST V11 Tips box located at position 3 on the Agilent Bravo deck.

8| Perform a Dual Height Mix on the wells containing sample and mastermix. Aspirate 40 µl at a height of 1 mm from the bottom of the well and dispensing 40 ml from the bottom of the well and dispensing at a height of 5 mm. Mix approximately 15 times.

9| Knock off tips for disposal into quadrant 2 of an empty 70 µl St V11.

10| Once the protocol is complete, seal wells containing sample with ABI optical caps and place the sample plate on thermocycler. (Thermoprofile is diagramed in Table 5).

Table 5: Enrichment Thermoprofile

Step	Temperature (°C)	Time (seconds)	Number of Cycles
Initial Denaturation	95	120	1
Amplification Cycle	95	30	
	65	30	6
	72	60	
Final extension	72	600	1
Hold	4	Infinite	1

Set up of the Agilent Bravo with LT head for Pond Enrichment 1.8 X Cleanup

Pond Enrichment 1.8X Automated Cleanup deck preparation

(Figure 3A displays deck layout)

Process Steps automated on the Bravo

1| Put on 180 µl tips from tip box # 1.

2| Aspirate 108 µl of Agencourt AMPure XP beads and dispense into sample plate. Mix beads prior to aspiration

3| Perform a Dual Height Mix (mixing approximately 15 times), then allow the sample plate to sit for approximately 2 minutes. Be sure to set the aspiration height to 1.5 mm from the bottom

of the well and the dispense height to 7 mm from the bottom of the well. Mix approximately 155 µl 15 times.

4| Place the sample plate on a Dynal MPC – 96 S plate magnet for 4 minutes to allow the AMPure XP beads to separate from the solution.

5| Remove and discard the supernatant into the AMpure XP beads source plate.

6| Discard the used 180 µl tips into tip box # 1.

7| Put on 180 µl tips from tip box # 2.

8| Leaving the sample plate on the Dynal MPC-96S magnet plate, aspirate 100 µl of 70% EtOH and dispense into the sample plate. DO NOT MIX.

9| Allow the AMPure XP beads and sample sit in the 70% EtOH for 30 seconds, then remove the EtOH and discard into the 70% EtOH source plate.

10| Discard tips into 180 µl Tip Box # 2.

11| Move the sample plate off of the Dynal MPC-96S magnet plate and allow the sample-AMPure bead complex to air dry for approximately 4 minutes at room temperature

12| Put on 180 µl tips from tip box # 3.

13| Aspirate 40 ml of Tris-HCl pH 8.0 and dispense into sample plate.

14| Perform a Dual height mix. Be sure to set the aspiration height to 1.5 mm from the bottom of the well and the dispense height to 6 mm from the bottom of the well. Mix approximately 40 µl 10 times.

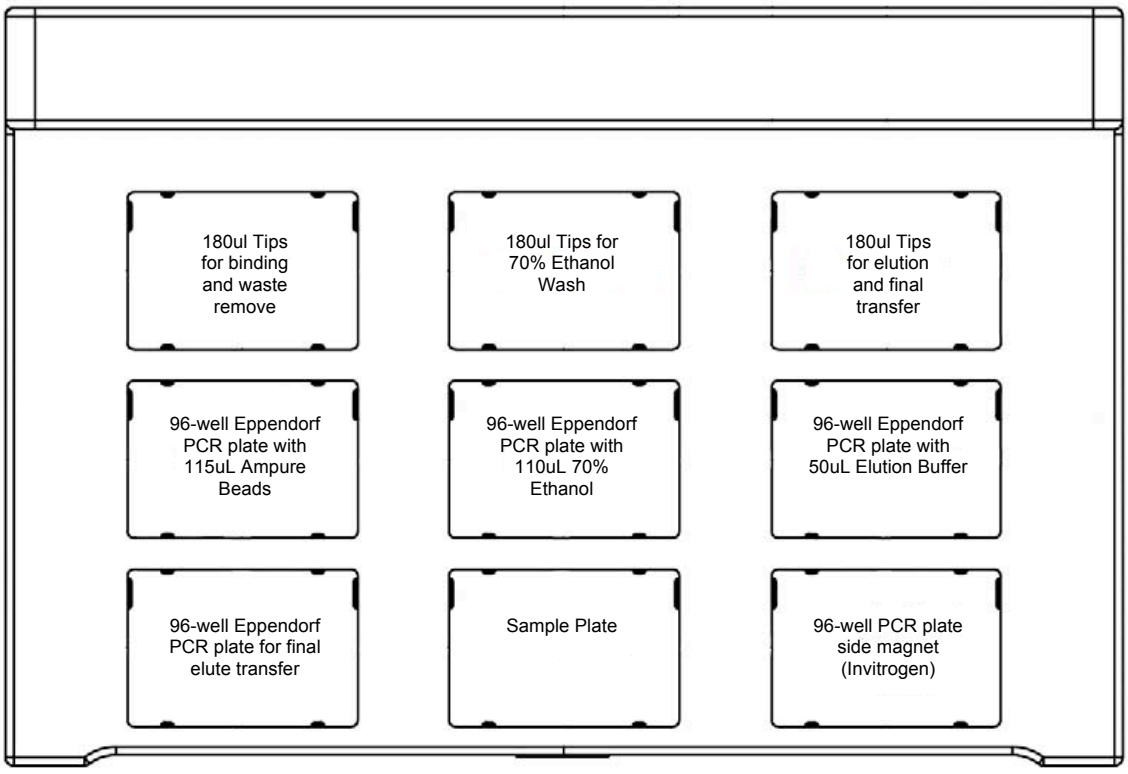
15| Allow the resuspended sample to sit for approximately 2 minutes.

16| Place the sample plate, on a Dynal MPC–96S plate magnet for 3 minutes to allow the AMPure XP beads to separate from the solution.

17| Aspirate the eluate and dispense into the Eppendorf 96 well twin.tec plate located at position 7.

18| Discard tips into tip box # 3.

19| Using ABI optical caps, seal Eppendorf plate containing samples.



ChIP-String

Enrichment validation: ChIP-String, DNA measurement on Nanostring

Details on the nCounter system are presented in full in (Geiss et al., 2008). We used a custom CodeSet constructed to detect a total of 786 probes covering ~200 genes (for detailed design of the code-set see below). 5 µl of HT-ChIP libraries DNA where denatured at 95°C for 5 minutes and immediately cooled on ice. The denatured DNA product was applied directly into the hybridization reaction (5X SSPE, 0.1% Tween-20), and incubated at 65°C for 16 hours in a PCR machine with a heated lid. The samples where loaded onto the nCounter prep station followed by quantification using the nCounter Digital Analyzer 2.

Notably, with recent increase in sequencing yields and substantial decrease in cost, using Nanostring for antibody QC is no longer as cost effective as it was when we optimized this approach.

Antibody quality control and Nanostring probe design

In order to rapidly and efficiently QC antibodies in our system we designed nCounter probe-set that target regulatory regions that are active during immune stimulation. We first selected a list of 185 induced post stimulation in DCs together with a set of 16 control genes that are either not expressed (Cryaa, Pck1, Hbb-b1, Gabrb1, Drd2, Pou5f1, Sox2) or that are expressed but their expression remains unchanged (Gapdh, Mea1, Ndufa7, Ndufs5, Rbm6, Shfm1, Tbca, Tomm7, Ywhaz) upon stimulation with LPS. Since active regulatory regions are enriched in signature chromatin marks (H3K4me3 and H3K4me1) and PolII peaks, we used a dataset that became available at the onset off our project and which was composed of ChIP of PolII and K4me3 in un-stimulated and stimulated macrophages (Ghisletti et al., 2010). Since macrophages have a similar gene expression program to DCs after LPS stimulation, we combined this dataset and the curated list of genes to design probes that target candidate regulatory regions. We first used scripture (see below) to call peaks of H3K4me3 and PolII enrichment. We then targeted annotated transcription start sites of all genes and designed 4 probes (2 probes for control genes) that centered at the TSS and complemented this set with two probes tiling of any significant PolII peak or K4me3 peak that lied within the gene body but did not overlap any of the original probes (Fig. 1b). We finally added 2 probes for any significant K4me3 peak that lied within 30Kb of the TSS of the genes we targeted. The final probeset consisted of 786 probes targeting regulatory regions of ~200 genes.

Sample preparation

Dendritic cell isolation, culture, and LPS stimulation

To obtain sufficient number of cells, we implemented a modified version of the DCs isolation used in (Lutz et al). Briefly, 6-8 week old female C57BL/6J mice were obtained from the Jackson Laboratories. RPMI medium (Invitrogen) supplemented with 10% heat inactivated FBS (Invitrogen), β-mercaptoethanol (50uM, Invitrogen), L-glutamine (2mM, VWR) penicillin/streptomycin (100U/ml, VWR), MEM non-essential amino acids (1X, VWR), HEPES (10mM, VWR), sodium pyruvate (1mM, VWR), and GM-CSF (20 ng/ml; Peprotech, Rocky Hill, NJ) was used throughout the study. At day 0, bone marrow-derived dendritic cells (BMDCs) were collected from femora and tibiae and plated on twenty (per mouse), 100mm non-

tissue culture treated plastic dishes using 10ml medium per plate. At day 2, cells were fed with another 10ml medium per dish. At day 5, cells were harvested from 15ml of the supernatant by spinning at 1400rpm for 5 minutes; pellets were resuspended with 5ml medium and added back to the original dish. Cells were fed with another 5ml medium at day 7. At Day 8, all non-adherent and loosely bound cells were collected and harvested by centrifugation. Cells were then resuspended with medium, plated at a concentration of 15x106 cells in 10ml medium per 100mm dish. At day 9, cells were stimulated for various time points with LPS (100ng/ml, rough, ultra-pure E. coli K12 strain, Invitrogen).

RNA extraction and RNA-Seq library preparation

Total RNA was extracted with QIAzol reagent following the miRNeasy kit's procedure (Qiagen), and sample quality was tested on a 2100 Bioanalyzer (Agilent). We prepared the RNA-A+-Seq libraries using the 'dUTP second strand (strand specific) protocol as described in (Levin et al., 2010). Briefly, extracts were treated with DNase (Ambion 2238). Polyadenylated RNAs were selected using Ambion's MicroPoly(A)Purist kit (AM1919M) and RNA integrity confirmed using Bioanalyzer (Agilent). RNA was fragmented by incubation in RNA fragmentation buffer (Affymetrix) at 80 degrees for 4 minutes. Fragmented RNA was mixed with 3 µg random hexamers (Invitrogen), incubated at 70 °C for 10 min, and placed on ice briefly before starting cDNA synthesis. We synthesized first-strand cDNA with this RNA primer mix by adding 4 µl 5× first-strand buffer, 2 µl 100 mM DTT, 1 µl 10 mM dNTPs, 4 µg of actinomycin D, 200 U SuperScript III and 20 U SUPERase-In, incubating at room temperature for 10 min followed by 1 h at 55 °C. We synthesized second-strand cDNA by adding 4 µl of 5× first-strand buffer, 2 µl of 100 mM DTT, 4 µl of 10 mM dNTPs with dTTP replaced by dUTP (Sigma), 30 µl of 5× second-strand buffer, 40 U of Escherichia coli DNA polymerase, 10 U of E. coli DNA ligase and 2 U of E. coli RNase H, and incubating at 16 °C for 2 h. cDNA was eluted using the Qiagen MiniElute kit with 30 µl of the manufacturer's EB buffer. DNA ends were repaired using dNTPs and T4 polymerase (NEB), followed by purification using the MiniElute kit. Adenine was added to the 3′ end of the DNA fragments using dATP and Klenow exonuclease (NEB; M0212S) to allow adaptor ligation, and fragments were purified using MiniElute. Adaptors were ligated and incubated for 15 min at room temperature (25 °C). Phenol/chloroform/isoamyl alcohol

(Invitrogen 15593-031); extraction followed to remove the DNA ligase. The pellet was then resuspended in 10 µl EB buffer. The sample was run on a 3% agarose gel (Nusieve 3:1 agarose, Lonza) and a 160–380 base pair fragment was cut out and extracted. PCR was performed with Phusion High-Fidelity DNA Polymerase with the manufacturer's GC buffer (New England Biolabs) and 2 M betaine (Sigma). PCR conditions were 30 s at 98 °C; 16 cycles of 10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C; 5 min at 72 °C; forever at 4 °C. Products were run on a polyacrylamide gel for 60 min at 120 V. The PCR products were cleaned up with Agencourt AMPure XP magnetic beads (A63880) to completely remove primers and the product was submitted for Illumina sequencing.

We sequenced all libraries using the Illumina Genome Analyzer (GAII). We sequenced two lanes for each sample, corresponding to 45 million paired-end reads/sample (90 million single reads, 76 bases long) on average.

Western Blot and antibody validation

Nuclear extracts from mouse bone marrow dendritic cells (DC) were prepared by using NE-PER nuclear and cytoplasmic extraction reagents (Thermo scientific, USA), and following the instructions of the manufacturer with minor modifications. Briefly, 10 million cells were harvested by centrifugation and washed with PBS. Cells were transferred to an eppendorf tube and 1ml of CER I buffer was added. Cells were resuspended by vigorous vortexing for 15 sec followed by incubation on ice for 10 min. 55ul of ice-cold CER II buffer was added and vortexed for 5 sec on the highest setting. After one min incubation on ice the tube was vortexed for another 5 sec and the nuclear fraction was separated from cytoplasmic extract (supernatant) by spinning for 5 min at maximum speed. Insoluble pellet was resuspended with 125ul of NER buffer and vortexed for 15 sec. It was then sonicated in Branson digital sonifier for 30 sec. using 45% amplitute and 0.7sec pulse on/ 1.3 sec pulse off setting. The tube was centrifuged for 10 min at 16000xg speed and the supernatant containing nuclear extract was transferred to a clear tube. All steps were carried out at 4C. 20ug of the nuclear proteins were separated by SDS-PAGE and transferred to PVDF membrane (BioRad). Prestained protein molecular mass marker (BioRad) was run to monitor electrophoretic transfer and to determine relative size. Membranes

were probed with antibodies and visualized by the enhanced chemiluminescence (ECL, Amarsham) method according to the instructions of the manufacturer.

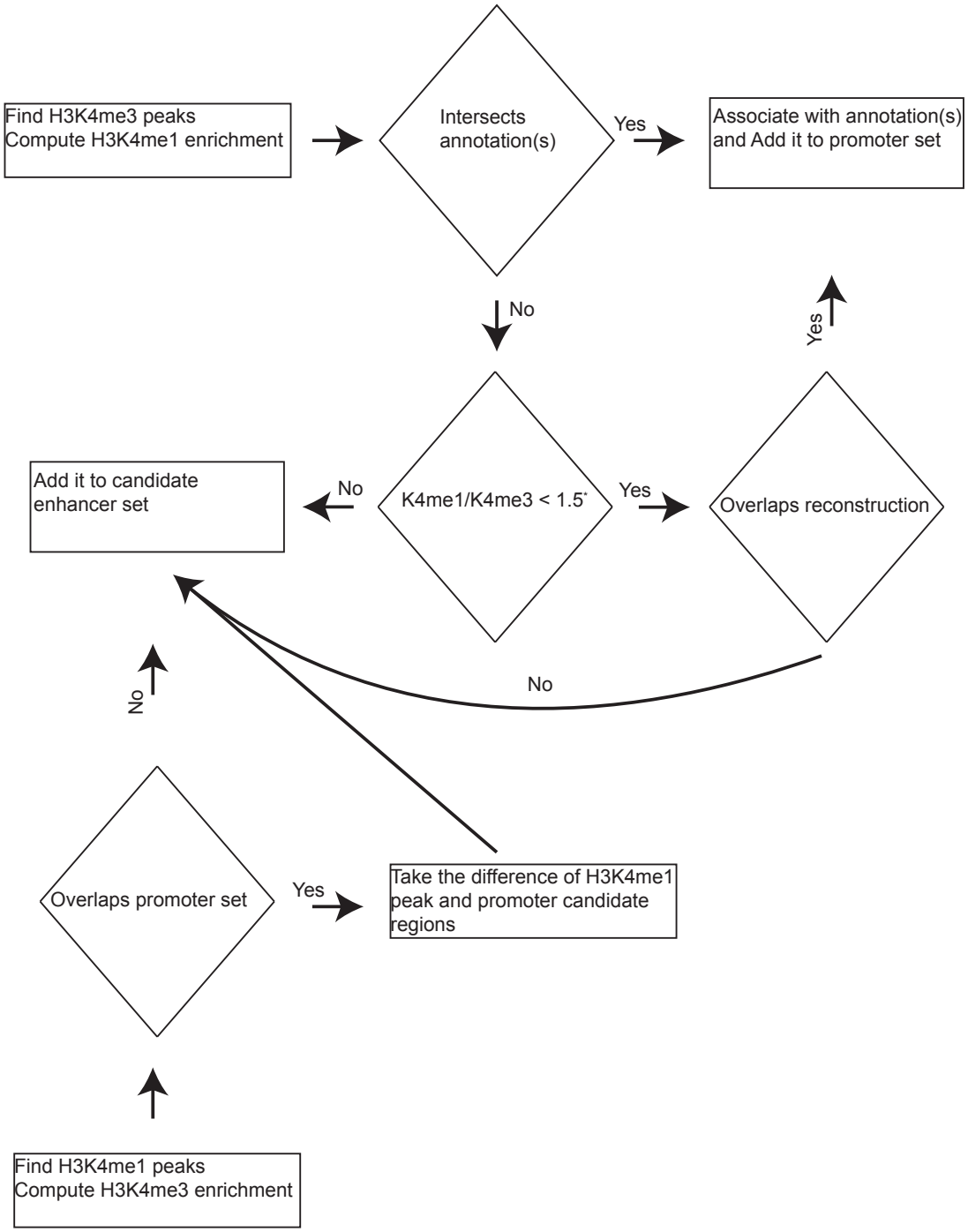
Data analysis

Annotations of promoters and enhancers

We implemented an annotation pipeline (shown above) to find candidate promoter and enhancers. First K4me3 and K4me1 peak calling was done using scripture as described above. We note that we weighted K4me3 marks slightly more than K4me1 marks (enhancer marks required a ratio K4me1/K4me3 >1.5) to preferentially annotate promoter regions and to account for the higher overall score when using the K4me1 library.

To define promoters, we first identified all the 19,057 H3K4me3+ regions as candidates. We then retained those 15,350 regions that overlapped a transcription start site of either a Refseq annotated transcript (Pruitt et al.) or of a transcript reconstructed from our RNA-Seq data (Guttman et al.) (Figure S2a top). Of the retained promoter regions, 11,505 (75%) were bound by at least one of the TF we ChIPed.

Next, we defined candidate enhancers as the 85,061 regions with higher H3K4me1 enrichment. We then retained only the 47,961 (56%) candidate enhancers that were also bound by at least one TF. We associated enhancers and promoters with genes using a new approach that, in the absence of genome wide 3 dimensional conformation data, uses the binding data to find plausible enhancer-promoter relationships. First, we associated each candidate promoter with the gene it overlaps. Next, we associated candidate enhancers to genes based on one of two criteria. (1) Enhancers were associated with genes when they were adjacent (<50bp) to their promoters. (2) Because active enhancers and their target promoters are thought to be in close physical proximity, and formaldehyde cross-linking affects both protein-protein and protein-DNA interactions, we expect some factors to immunoprecipitate with both the promoter and the enhancer with which it interacts (even if they are directly bound only to one of those sequences). We therefore associated each enhancer region with all promoters within 150 Kb which were bound by at least 2 of the factors found at the candidate enhancer region (Figure 4C S2A). We associated 38,439 (80%) of the annotated enhancer regions with 11,265 genes with an average of 4.8 enhancers per-gene and 1.2 genes per enhancer.



Sequencing and read alignments

ChIP Sequencing was done on Illumina HiSeq-2000 at the Broad Institute sequencing center. Pooled libraries were sequenced in ~12 samples per lane at a sequencing depth of ~8 million

aligned reads per sample. Initially we sequenced several libraries using different sets of read lengths with and without paired-end reads to test the impact on the analysis. The optimal read length for both cost and sensitivity was 44 bases (8 bases in this scheme are used for the indexes) with which we sequence all later libraries.

Reads for each index and each lane was aligned to the reference mouse genome NCBI37, using BWA (Li and Durbin, 2009) version 0.5.7 with parameters

`-q 5 -l 32 -k 2 -t 4 -o 1 –f` for the `aln` command and
`-P -a 600 –f` for the `sampe` command.

After combining reads from different lanes corresponding to the same factor at the same time point we aligned an average of 11,249,898 (7,370,024 sd) reads per time point for transcription factors and an average of 22,678,617 (10,176,016 sd) for chromatin and PolII libraries.
In addition we created a "compressed" alignment for each factor, obtained by merging the alignments for each of the timepoints for a given library.
RNA sequencing was done for samples obtained from DCs pre-stimulation, 1, 2, 4 and 6 hours post stimulation and performed on an Illumina GA-II using 2 lanes per sample and a read length of 76 bases. All reads were aligned to the mouse reference genome (NCBI 37, MM9) using the TopHat aligner version 1.1.4 (Trapnell et al., 2009). Briefly, TopHat uses a two-step mapping process, first using Bowtie (Langmead et al., 2009) to align all reads that map directly to the genome (with no gaps), and then mapping all reads that were not aligned in the first step using gapped alignment. TopHat uses canonical and non-canonical splice sites to determine possible locations for gaps in the alignment. We used the EST database, which we downloaded from the UCSC genome browser (Fujita et al., 2011) to improve TopHat sensitivity for splice alignments. Specifically we used the following TopHat parameters:

```
-g 15 -r 250 --library-type fr-firststrand –G spliced.est.gtf -p 4
```

Where the spliced.est.gtf file was downloaded from UCSC. We obtained an average of 73 million uniquely aligned reads, of which and average 55 million aligned in proper pairs and 15

million aligned spanning a putative spliced junction. In addition, we also used our previously described 4SU labeled libraries collected from un-stimulated DCs and every hour for 6 hours after stimulation. All bam files were submitted to GSE36104.

Peak calling

We implemented our contiguous segmentation algorithm, described in (Guttman et al., 2009) as part of the Scripture package (available from http://www.broadinstitute.org/software/scripture/) and used it to call, score and filter peaks for both chromatin and TF libraries. Scripture calls peak using the same statistical methods reported previously (Guttman et al., 2010), with efficiency improvements that are possible thanks to the contiguity of enriched and result in faster runtime compared to rnaSeq transcript reconstruction. The algorithm scans fixed-sized windows across the genome, identifies significant windows, then merges and trims to obtain peaks. Since different ChIP libraries have very different peak characteristics (e.g. H3K27Ac is elongated with moderate enrichment, while transcription factors peaks are small and highly enriched) the window size allows us to find peaks of very different nature. Significance (at the user specified level) is assessed using the scan statistic (Wallenstein and Neff, 1987) with underlying Poisson distribution with mean the average number of reads found for a similar window size in the genome. Since we scan fixed windows, windows at the edges of peaks tend to have lower coverage at the ends. We trim bases at the ends of enriched regions, using a quantile of coverage specified by the user.

For transcription factors we used the following parameters:

```
Java –Xmx3000m –jar scripture.jar -task chip -maskFileDir <Our local file of Mouse masked regions> -trim -windows 200 -fullScores -alpha 0.05 -minMappingQuality 30 -alignment <TF BAM alignment>
```

To scan 200 base windows and define peaks composed of windows whose coverage at 0.05 significance (under the scan statistic).

To allow for longer, possibly less enriched peaks we used -windows 750,1500 for Chromatin marks and PolII but we only considered windows with read coverage at 0.01 significance. For each TF and chromatin modification we used the compressed alignment as the input to Scripture, this allowed us to obtain a single set of regions using the combined power of all time points combined.

Peak filtering and scoring

For a given library *C*, a significant region *R* is scored by the enrichment score: $S_C(R) = \frac{N_R}{l_R} / \frac{N}{L}$, where *N* is the total number or reads, *L* is the length of the alignable genome, N_R the total reads overlapping the region and l_R is the length of the region *R*. Regions of open chromatin tend to generate more reads than regions of less accessible DNA regardless of enrichment for a specific antibody target. In order to control with this and other fluctuations in read coverage not due binding of DNA to the target protein, we used whole cell extract (WCE) libraries as our null set. For every library *C* we further filtered Scripture significant regions by running a fixed window of 150 bases across the region computing a fold score $E_C(R) = \frac{S_C(R)}{S_{WCE}(R)}$ and kept only those significant regions having a fold score $E_C(R) > 3$. The peak score was set to the maximum scoring 150b window across the region. This score makes regions comparable independently of their size. We call this highest scoring 150b window within a peak the peak summit.
Each peak was then scored using the time course alignments by computing the enrichment score of the maximum scoring window for the time point.

Peak consistency across time

For each sequenced library we have generated a time-course data, using four different time points. To some extent, the consistency between time points can provide an estimate for the robustness of the data. To test this, we have re-analyzed the binding data, applying the peak detection algorithm separately at each time point. We then compared the time points to each other. We found that in 86% of the binding sites the binding is called in at least two different time points (Figure S2C-I). The consistency was weak only in cases of drastic changes in TF activity (e.g. at 2 hours for Stat1).

Transcriptome annotation and quantification (RNA-Seq)

Top-Hat alignments were processed by Scripture (Guttman et al., 2010) to obtain significantly expressed transcripts for each time course. Only multi-exonic transcripts were retained.
Scripture was run using the following parameters to find transcripts one chromosome at a time:

```
java -Xmx5000m -jar scripture.jar -minSpliceSupport 2 -trim -maskFileDir <mm9_masked_regions> -windows 0 –alignment <timepoint_rnaseq_alignment.bam> -out chr<chrSymbol>.segments.bed -chr chr<chrSymbol> -chrSequence chr<chrSymbol>.fa
```

We quantified transcript expression of RefSeq (Pruitt et al., 2007) annotations (as of 02/24/2011) downloaded from the UCSC table browser. We used scripture to first find constituent isoforms (Garber et al., 2011; Guttman et al., 2010) for those genes with multiple isoforms. We then computed the Reads Per Kilobase of transcripts per Million reads (RPKM) based on the both the total and labeled RNA-Seq data for further analysis of expression (Table S9).

Systematic selection of transcription factor targets for HT-ChIP

In order to systematically select for potential TFs functional in the DC system we used our RNA-Seq time course data of DC activated with LPS. We compiled a list of 1,885 putative TFs from public databases (Amit et al., 2009) . We then filtered this preliminary list by requiring expression of at least RPKM > 15 at one point in our stimulation time course. We finally manually curated the list to remove any gene that is not a sequence specific TF (e.g. general transcription machinery, chromatin modifiers, etc) to obtain a set of 184 TF targets (Table S1) to screen for potential antibodies in commercial antibody vendors databases.

Motif analysis

We performed both de-novo motif discovery and known motif matching using the MEME software suit. We extracted the sequences of the summit regions in the high scoring peaks (see above definitions) and used those as input for the MEME-ChIP pipeline (http://meme.nbcr.net/meme4_6_1/memechip-intro.html). The pipeline runs MEME (Bailey and

Elkan, 1994) for motif discovery, TOMTOM (Gupta et al., 2007) to search discovered motifs within existing databases and MAST (Bailey and Gribskov, 1998) to search for known TF motifs in the sequence provided. Supplemental Fig. 6 summarizes the motif analysis results. Briefly, at a significance cutoff of 0.01, we found known motifs for 20 of the 25 transcription factors. For two of the factors we found s motif that were not previously associated with the factor: for Ahr we find known motifs associated with a different factor (Nfy), while for Ets2 we find a previously reported motif that was not associated with a transcription factor. The two factors for which we do not find significant motifs within their summits: E2f1, Nfkb1 tend to smear throughout the gene body and do not present clear peaks like all other factors**.**

Generating the global property map

For generating the global property map in Fig. 2b, the peaks were filtered by requiring a minimum score of 26.9 which corresponds to the mean + 0.25*std of all peak enrichment scores. This cutoff also corresponds to the top ~33% scoring peaks. We use this set of peaks throughout the subsequent analysis. Peaks overlapping a promoter region that where closer to a transcriptional start site than to a 3'UTR were considered promoter peaks. Peaks were classified as 3'UTR whenever they were within 1kb of an annotated 3' end and no transcriptional start site was closer. Peaks overlapping enhancer regions were classified as enhancer bound. The remaining peaks were classified as intronic, exonic or "other" whenever they overlapped an annotated intron, exon or neither.

Conservation of binding sites

In order to evaluate sequence conservation of candidate binding motif within TF bound regions we used a slight extension of the Siphy method (Garber et al., 2009) that looks at pattern of substitution rather than decrease in substitution rates and thus can better accommodate more degenerate position within protein-DNA binding sites.

Our motif Siphy module (available with the latest version from `http://www.broadinstitute.org/genome_bio/siphy/`) finds the best element within a given genomic region whose substitution pattern is most consistent with the position weight matrix (PWM) that represents the TF binding preference.

Siphy uses multiple sequence alignment, a position weight matrix (PWM) of size k, a substitution rate matrix Q, a phylogeny and a background stationary nucleotide distribution π_n then scores each k sized element (k-mer) using a log-odds ratio score that indicates whether the k-mer is less or more likely to be under neutral or selection pressures that are consistent with preserving the PWM.

To score a k-mer Siphy takes each base bi within the k-mer (i=1..,k) it computes $\log \frac{P(c_i \mid \pi_i, Q)}{P(c_i \mid \pi_n, Q)}$. The "conservation" score for a k-mer is $S(k\text{-mer}) = \sum_i \log \frac{P(c_i \mid \pi_i, Q)}{P(c_i \mid \pi_n, Q)}$ and the score for a genomic region is the maximum of S(k-mer) for all k-mers in the region. Computation of the likelihood is done using standard techniques as described in Falsenstein's (Felsenstein, 2004).

For each TF factor for which we found a de-novo binding motif we ran Siphy on all its high scoring peaks that had a match to the motif as reported by

```
java -Xmx5g -jar siphy.jar -task maximalPWM -indir <alignment directory with one indexed MAF file per chromsome> -mod <sampled_autosome_combined.out.mod> seedMinScore=-8 -pwm <Motif.pwm> -regions <tomtom motif matches> -out <motif.score.out>
```

Assessing co-binding and overlap of peaks with annotated genomic regions

For each pair of ChIP assays (say x, and y) we used a binomial p-value to assess their overlap in the genome as described in (McLean et al., 2010). The number of hits is set to the number of x peaks that fall within 500bp away from some peak of y. The background probability set to the length of regions associated with y (i.e. taking 500bp margin around each of its peaks) divided by the overall length of genomic regions that are associated with at least one ChIP assay (Table S5).

We performed a similar computation for assessing the overlap of ChIP assay peaks with annotated genomic regions (Table S5). To compute the overlap of assay x with region y, we define the number of hits is set to the number of x peaks that overlap with y. The background probability set to the length of regions associated with y divided by the overall length of the genome. The regions we used include: (i) regulatory features annotations from ensembl (Flicek et al., 2011), (ii) Regulatory features found by the oregano algorithm (Griffith et al., 2008), (iii) Conserved regions annotated by the multiz30way algorithm, here we consider regions with multiz30way score>0.7. Region data was downloaded from the UCSC genome browser.

Computation of the percent of bound motifs

For every de-novo found motif with an E-value of less than 0.01 we scored every peak by evaluating the match to the motif within the peak. To this end, we used the standard log-odds ratio score of the probability of a given k-mer being generated by the inferred position weight matrix and the probability of the k-mer being generated by a neutral model of 40% GC, the mouse genome wide GC content percent.

The 10th percentile was the used in a genome wide scan for available motifs. The percent of bound motifs was the ratio of motifs scoring higher than the cutoff to the number of genome wide matches above the cutoff.

Assessing co-binding and overlap of peaks with annotated genomic regions

For each pair of ChIP assays (say x, and y) we used a binomial p-value to assess their overlap in the genome as described in (McLean et al., 2010). The number of hits is set to the number of x peaks that fall within 500bp away from some peak of y. The background probability set to the length of regions associated with y (i.e. taking 500bp margin around each of its peaks) divided by the overall length of genomic regions that are associated with at least one ChIP assay (Table S2).

We performed a similar computation for assessing the overlap of ChIP assay peaks with annotated genomic regions (Table S2). To compute the overlap of assay x with region y, we define the number of hits is set to the number of x peaks that overlap with y. The background probability set to the length of regions associated with y divided by the overall length of the genome. The regions we used include: (i) regulatory features annotations from ensembl (Flicek et al., 2011), (ii) Regulatory features found by the oregano algorithm (Griffith et al., 2008), (iii) Conserved regions annotated by the multiz30way algorithm, here we consider regions with multiz30way score>0.7. Region data was downloaded from the UCSC genome browser.

Defining TF-region and TF-gene associations

We start by defining a TF-region association matrix, whose columns correspond to TF binding in the four studied time points (altogether 4 columns per TF) and rows correspond to regions (promoters and enhancers). The association value is the sum of enrichment scores over all the peaks of the TF that fall within the given region at the given time point. We also define a categorical TF-region association matrix where each factor is associated with a single column. An entry in the categorical matrix will have a "none" value if there is no peak (in either of the time points) within the respective region that has an enrichment score over the cutoff value. Otherwise, if the respective scores at the TF-region association matrix have substantially increased (3-fold, see next section) over time in comparison to the basal state (at t=0) then the value will be "gain". If the scores have decreased (3-fold), the value will be "loss", and if they remained on the same scale (<3-fold change), the value will be "static".

Turning to the gene level, we define a TF-gene association matrix, whose rows correspond to target genes and whose columns correspond to TF binding in the associated enhancers or promoters during the four studied time points (altogether 4 columns per TF). The association value is the average over all the regions (promoters or enhancer) that are associated with the current gene. As above, we also defined a categorical TF-gene association matrix, where each factor is associated with one column as above. The values in the matrix are determined in accordance to the regions associated with the gene using the categorical TF-region association matrix. If there are no bound enhancers or promoters, the value will be "none". Otherwise, if at

least 50% of the bound enhancers or at least 50% of the bound promoters are associated with gain, then the value will be "gain". If at least 50% of the bound enhancers or at least 50% of the bound promoters are associated with loss, then the value will be "loss" (in 2.5% of the cases where both conditions hold, we mark the entry as "static"). If both conditions do not hold, we mark the entry as "static".

We use the resulting matrices throughout our analysis for defining the TF binding events and the dynamics of these events.

A random model for TF binding across regions

We define a region *TF complexity* as the number of TF that bind to the region during at least one time point, this is also called the region *complexity score* (Roy et al., 2010) Figure 4A presents for each TF a histogram of the complexity values associated with the regions it binds (during at least one time point). We compared this observed distribution to a random model in which the complexity of every region is proportional to its length (in base pairs). To generate the random data, we apply the following procedure:

```
1.FOR EACH transcription factor T
2.   FOR EACH region r
3.     SET complexity(r):=0
4.     IF r is bound by T set complexity(r):=1
5.   END FOR
6.   FOR EACH transcription factor G≠T
7.      Let N be the number of regions bound by G
8.      Let R be the entire set of bound regions
9.      FOR i=1 to N
10.             Choose a region r' at random from R where
            the probability for choosing a region is:
                 length(r')
                ∑ length(r)
                r∈R
11.         SET complexity(r'):= complexity(r')+1
```

```
12.          SET  R := R \ {r'}
13.     END FOR
14.  END FOR
15.    Compute a histogram over the resulting (randomized)
     complexities of the original targets of T
16. END
```

Note that in the random selection stage (lines 9-13) each gene can only be selected once, and so the overall number of binding events in the randomized binding map remains the same. To speed up the process, instead of choosing only one region at a time (line 10) we randomly assign the entire *N* points. If a region is selected more than once then we un-select it. We then update *N* to be *N-N'* (where *N*' is the number of points successfully assigned to regions that were only selected once) and repeat the random selection procedure. In all cases this process converged after no more than 10 iterations.

We repeat the procedure 50 times. Figure 4a shows the average of the resulting randomized histograms for each TF (right).. Notably, we obtain qualitatively similar results (that are even father away from the observed ones) when the selection of regions (line 10) is done in a uniform manner, with no consideration for length (data not shown).

We use this analysis to define the cutoff for HOT regions. To this end, searched for minimal number x of binding events (x>2) such that the frequency of x TF binding in the observed distribution of complexities is higher by at least two fold than the frequency of x in the randomized distribution. We estimate the deviation of the observed average complexity from the randomized ones using a Z-score test. A similar approach is taken to estimate the p-value for the tendency of PU.1 and Cebpb to bind at regions with complexity value of 1.

Evaluating the tendency of TF to co-bind at similar regions

For every pair of TF we evaluate their tendency to bind at the same regions during the same time point using a hypergeometric p-value: $HG(N,B,n,b) = \sum_{m=b}^{\min\{n,B\}} \frac{\binom{B}{m}\binom{N-B}{n-m}}{\binom{N}{n}}$ where N is the overall number of regions bound at that time point, B is the number of regions bound by the first TF at the given time point, n is the number of regions bound by the second TF at the given time point, and b is the number of regions bound by both at the given time point.

We limit the analysis to regions with at least two binding events. Further, to get more specific results, we filter out HOT regions with 8 or more bound TF. We only report TF pairs that had a p-value lower than 10^{-3} during at least one time point.

Determining the timing of changes in binding events

The timing associated with gain of binding (Fig. 4b) is determined as the first time point where the fold increase in binding score reached 50% of its maximum level. The timing of binding-loss events is determined in a similar manner. We only consider cases in which the bound regions had a substantial (>3 fold) change (up/down) in their respective binding score, comparing the basal binding (at t=0).

Computing TF binding- based clusters

To generate the heatmaps in Figure 4D (for genes), we used the *k*-means algorithm with $k=8$ using elements (genes or regions) that have at least three binding events, which in most case translates to an additional TF on top of PU.1 and Cebpb.

Computing expression- based clusters

We clustered genes that transcriptionally responded to LPS stimulation into several prototypical temporal profiles. We start the analysis by filtering non-responsive genes. To this end we only retain genes whose expression levels changed by at least two fold during the response (taking the maximum expression over the minimum expression in time). To avoid variations at very low expression levels we add a constant factor (set to the second quantile over all non zero expression levels), to the numerator and denominator of the fold ratio. We used the *k*-means algorithm with pearson correlation to cluster the genes based on their time course of expression levels. We chose the smallest k (k=5) that provided a within cluster similarity levels (average correlation with centroid) with an average of 0.8 and a minimum of 0.75.

Associating the number of bound TF at basal state with induction potential

We assign a p-value for the relation of number of binding events with induction using the Wilcoxon ranksum test. This test evaluates the difference in the median number of binding events between genes in the induced clusters as compared to non-induced genes (1.38 fold increase; $p<10^{-10}$).

Evaluating enrichment of TF binding in gene clusters

Given a cluster S and a TF f, we wish to test whether f binding in S is stronger than expected by chance. To this end we compute a hypergeometric score assessing the overlap between the set S and the set of genes that are bound by f. We require the hypergeometric p-value to be lower than 10^{-3}.

The overall number of bound TF strongly correlates the strength of expression ($p<10^{-10}$). In order to control for the general tendency of TF to be associated with highly expressed genes, we divided the genes into ten bins of transcription levels (taking percentiles 10, 20, … , 100) and then shuffle the genes inside each bin. We repeat the hypergeometric p-value computation for 50 different random shuffles and require that the original score be no worse than than one random iteration (empirical p-value < 0.05). To get more specific indication on the contribution of f to S we exclude from the analysis gene that are associated with HOT regions (i.e., regions bound by 8 or more TF). We apply this analysis to test which TF preferentially bind at induced genes before

stimulation (Figure S5A,B), and to test which TF preferentially binds at the expression-binding clusters (Figure S5E).

Associating gain or loss of binding with induction or repression

We started this analysis by studying the entire set of induced genes (taking the union of all the induced clusters) to repressed genes (taking the union of all the repressed clusters). We calculate for each gene a gain index, defined as the number of gain events minus the number of loss events (3-fold increase or decrease in the binding score). We applied the Wilcoxon ranksum test to measure the difference in median gain index between genes in the induced clusters and the rest of the genes (4.7 fold increase; $p<10^{-10}$). We also used a hypergeometric p-value to evaluate the overlap between induced genes and genes that have a positive gain index ($p<10^{-10}$). We repeated this analysis to evaluate the overlap between genes in the repressed clusters and genes that have a gain index of zero ($p<10^{-10}$) or negative (loss of binding; $p<10^{-10}$).

Given a cluster of genes S and a TF f, we wish to test whether gain of f binding is significantly associated with the transcriptional changes in S (Figure S5C,D). To this end we compute a hypergeometric score assessing the overlap between the set S and the set of genes that have a gain in f. In a standard hypergeometric score we would consider a gene in S to be a "hit" if it has a gain of f binding. Here we add another constraint and require that the timing of the binding gain is no later than the timing of the transcriptional change of this gene. We require the hypergeometric p-value to be lower than 10^{-3}. As in the previous sections, we also filter out genes associated with HOT regions, and use randomize instances to control for the fact that TF tend to be associated with highly expressed genes. Notably, we did not find any significant hits for the repressed clusters.

Computing the joint TF-binding and expression-based clusters

We clustered the genes that transcriptionally responded to LPS stimulation into groups that are coherent both in the temporal expression profile and in the TF binding profiles. We used a two-level procedure to cluster the responsive genes. We start with the expression-based partition of the genes as defined above. Next, we split every cluster using TF-binding profiles, as given in the categorical TF-gene association matrix. Every gene is represented by a 44-dimensional

vector (2 entries for each TF). A TF will be represented by the pair [0,1] in case of binding loss; [2/sqrt(2), 2/sqrt(2)] in case of static binding; [1,0] in case of gain; and [0,0] otherwise.
Once again, we use the *k*-means algorithm with pearson correlation. We chose, for each cluster separately, the smallest k (k=5) that provided a within cluster similarity levels with an average of 0.6 and a minimum value of 0.55.

We used a randomization analysis in order to evaluate the significance of the overall dependency between the expression and binding data, as captured by the cluster analysis ($p<10^{-10}$). To this end, we compare the distribution of cluster sizes obtained with the original data to that obtained from randomized instances (retaining the expression levels and shuffling the binding data). Statistical significance is evaluated using the Kolmogorov-Smirnov test. The randomization retains the number of binding TF per-gene, in order to control for signals emerging purely from this property.

Functional enrichment in clusters and in TF targets

For each cluster we compute: (i) Overlap with genes associated with immflamatory and anti-viral response (Amit et al., 2009), Figure S5); and (ii) Functional enrichment using annotations from the MsigDB dataset (Table S3) using the "canonical pathway" subset of the curated gene set (c2.cp.v3.0), the "cellular process" subset of the Gene ontology gene set (c5.all.v3.0), and the motif gene set (c3.all.v3.0) http://www.broadinstitute.org/gsea/downloads.jsp).
We evaluate the significance of these overlaps using a hypergeometric p-value (see formula above), where B is the size of the cluster; n is the number of genes that have the investigated property (i.e., a functional group from MsigDB, or genes annotated as immflamatory or anti-viral 4); b is the number of genes that belong to the cluster and to the annotated set; and N is the background set of genes. For the first test, N is the number of genes with at least three binding events (N=11,351); for the second test, N is the number of genes in the MSigDB database (N=14,017). The clusters and the resulting enrichments are provided in Table S6. We report all the results with p-value lower than 10^{-3} .

We also applied the enrichment test to the set of genes associated with HOT regions, and to the target genes of every TF (Tables S6, S7).

nCounter data analysis

We used the following pipeline to re-analyze the knockdown data from (Amit et al., 2009). For each sample, we divide the nanostring count values by the sum of counts that are assigned with a set of control genes that are the least affected by shRNAs and LPS stimulation (10 gene altogether, including Ndufa7, Tbca, and Tomm7; see (Amit et al., 2009). For each condition we compute a fold change ratio comparing to five control samples infected with non targeting shRNA. We then pooled together the results of all pairwise comparisons (i.e. AxB pairs for A repeats of the condition and B control samples): we require a substantial fold change (above a threshold value t) on the same direction (up/down regulation) in more than half of the pairwise comparisons. The threshold is determined as larger of 1.5 and d, where d is the mean + 1.645 times the standard deviation in the fold change shown by the control genes, which corresponds to p=0.05, under assumption of normality. We ignore all pairwise comparisons in which both control and knockdown samples had low counts (<100) before normalization.

TF network redundancy and integration with previous RNAi screen and motif conservation suggests a billboard model for early induced genes

To determine the degree of functional and non-redundant binding in the DC network we integrated the binding data with RNAi perturbation data monitoring mRNA expression changes in a selected set of signature genes six hours post LPS stimulation ((Amit et al., 2009); Figure S6). We find that overall 38% of the binding events involving our signature set were associated with up or down regulation of the target gene. Furthermore, for ~38.8% of gene expression

changes we observe have no direct binding of the factor, suggesting that the regulation may be indirect. We observe various levels of redundancy in the network. While some factors like Irf1 and Nfkb2 affect a relatively small portion of the genes they bind, others are non-redundant and affect a large fraction of the genes they bind. For instance, over 85% of the signature genes included in the late induced, stat-dependent cluster are down regulated in Stat1 knockdown. This lack of redundancy is consistent with a higher sequence conservation of Stat1 binding sites in the late-induced cluster compared with sites of Irf1 and Nfkb2 on their early-induced targets where they affect a smaller percent of bound targets.

The TF hierarchy graph

For every pair of TF <X,Y> we define *coverage(X,Y)* as the percentage of regions (promoters or enhancers) bound by Y that are also bound (at the same or earlier time) by X. The computed coverage scores are provided in Table S5.

We define the TF hierarchy graph by placing a directed edge exists from X to Y if *coverage(X,Y)>0.3*. The resulting network contains 124 edges, many of which link the top of the hierarchy directly to the bottom of the hierarchy. For clarity of presentation we removed direct links between nodes at the top of the hierarchy to nodes at the bottom of the hierarchy. We find all triplets of nodes <X,Y,Z> such that (1) the edges <X,Y>, <X,Z>, and <Y,Z> all exists in the network; (2) the difference between the out degree (number of out going edges) of X and Y is larger than 3 (i.e., X and Y are not at the same "level" of the hierarchy); and (3) the out degree of X is larger than a third of the maximum degree (thus only pruning edges that emanate form the upper tiers). For every found triplet we remove the edge between X and Z.

Each connected component in the network that is rooted at Pu1 or Cebpb represents a unique combination of TF. The combinations derived from the original network cover 89% of the data (i.e. for a given region at a given time point there is a 89% chance that the combination of TF

bound at that region is represented by a directed path in the TF hierarchy graph). The combinations in the trimmed network cover 78% of the data.

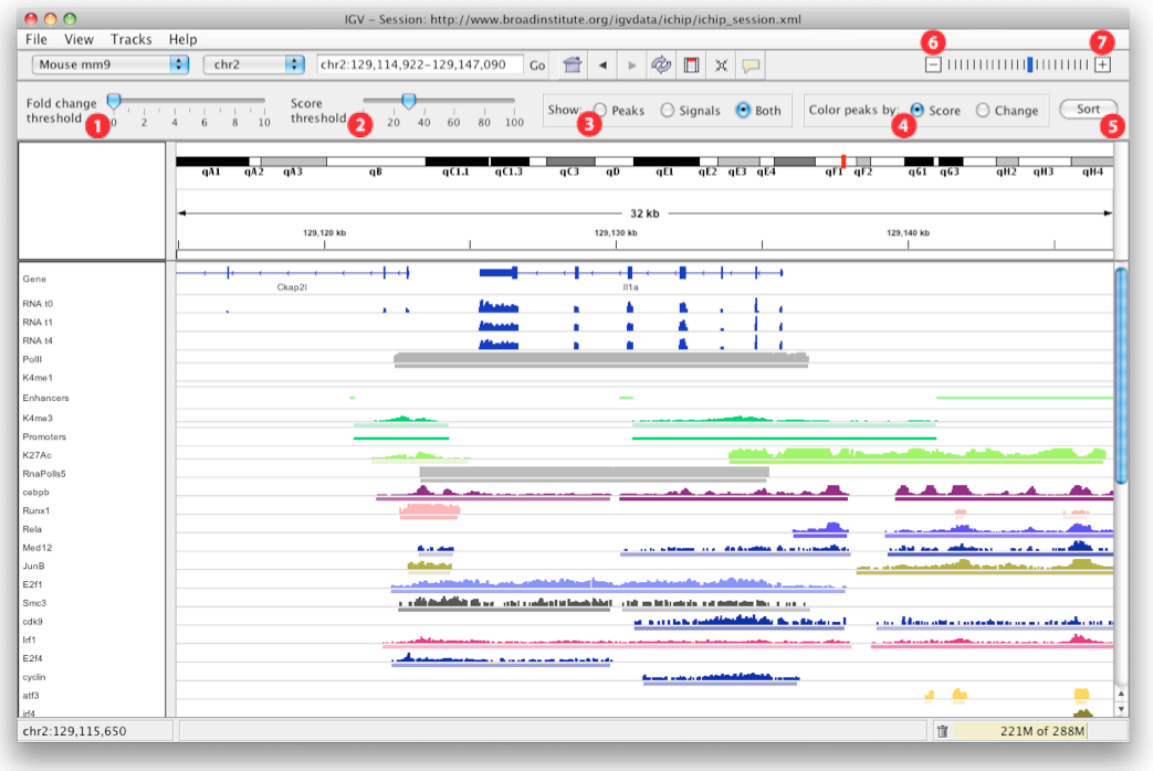
Principal component analysis

The *pcomp* Matlab function was applied to the 23x8 dimensional matrix consistent of all transcription factors (excluding Atf4 for which we had only one time point, and the chromatin insulator Ctcf) and six binding characteristics scored: number of bound regions, percent of dynamic binding events, promoter to enhancer binding ratio, percent of regions bound in isolation, percent of genome wide motifs bound by the TF, conservation of bidning sites, number of out going edges in the TF hierarchy graph, and number of incoming edges in the TF hierarchy graph,. All covariates were standardized (mean zero, STD 1) prior to the analysis. We use the *biplot* Matlab function to present the TFs projections and the loading of the different covariates for the first two principle components. Notably, the first three principal components account for 81.9% of the variance in the data.

Visualizing HT-ChIP with Integrative Genomics Viewer

This section describes the HT-ChIP extensions to Integrative Genomics Viewer (IGV). General documentation is available at http://www.broadinstitute.org/igv.

The HT-ChIP enabled IGV can be launched directly from the following link http://www.broadinstitute.org/igv/ichip. This link will download and install IGV 2.0 with HT-ChIP extensions, and open the HT-ChIP dataset at the il1a locus. The HT-ChIP extensions include a new command bar and popup menu.



1	*Fold change threshold:* Defines a minimum fold-change score, below which peaks and their associated signals below this score are not shown. The fold-change score is defined as $fcs = \log_2(\frac{\max+1}{\min+1})$
2	*Score threshold:* Defines a minimum peak score below which peaks and their associated signals are not shown. Set this value to zero to see all signals, even those with no associated peaks.
3	*Show Peaks|Signals|Both:* Controls what is shown in each factor track. Peaks are represented by rectangles. The associated signals are represented

	as bar-charts.
4	*Color by Score|Change:* Controls color scheme used for peaks. The *score* option colors peaks by factor, with a transparency used to indicate relative score. The *change* option colors peaks by fold change, with red indicating a positive score change and gray a negative. Transparency indicates the relative magnitude of the change.
5	*Sort:* Sorts non-chromatin factor tracks according to the average peak score of the features in the current view. The RNA-seq and chromatin factor tracks are not sorted and remain at the top of the panel.
6	Click on the – block to zoom out one level.
7	Click on the + block to zoom in one level.
	General navigation options: To zoom in double-click anywhere in the main data panel. To pan left and right click and drag in the data panel. To jump to another location enter either a gene symbol or a location string in the search box at the top of the screen and hit "Go".

Track menu

Right clicking on a track pops up a menu with track-specific options, including the following HT-ChIP -specific items.

Compressed: Display the compressed data, with a single value for each time series.

Time Series: Display the time series data.

Open Trend Plot: Opens an X-Y trend plot of peak score versus time at the locus clicked. The peak closest to the location clicked is selected for each track.

Supplemental References

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O.*, et al.* (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257-263.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Bailey, T.L., and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. Bioinformatics *14*, 48-54.

Felsenstein, J. (2004). Inferring Phylogenies (Massachusetts: Sinauer Associates).

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L.*, et al.* (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome biology *12*, R1.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S.*, et al.* (2011). Ensembl 2011. Nucleic acids research *39*, D800-806.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A.*, et al.* (2011). The UCSC Genome Browser database: update 2011. Nucleic acids research *39*, D876-882.

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods *8*, 469-477.

Garber, M., Guttman, M., Clamp, M., Zody, M.C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics *25*, i54-62.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T.*, et al.* (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol *26*, 317-325.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L.*, et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

Griffith, O.L., Montgomery, S.B., Bernier, B., Chu, B., Kasaian, K., Aerts, S., Mahony, S., Sleumer, M.C., Bilenky, M., Haeussler, M.*, et al.* (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic acids research *36*, D107-113.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome biology *8*, R24.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C.*, et al.* (2010). Ab initio reconstruction of cell type-specific

transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol *28*, 503-510.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.
Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods *7*, 709-715.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.
Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research *35*, D61-65.
Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F.*, et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Wallenstein, S., and Neff, N. (1987). An approximation for the distribution of the scan statistic. Statistics in medicine *6*, 197-207.