Supplementary Information

Chromatin Poises miRNA- and Protein-coding Genes for Expression

Artem Barski[1,*], Raja Jothi[1,2,*], Suresh Cuddapah[1,*], Kairong Cui[1], Tae-Young Roh[1,3], Dustin E. Schones[1] and Keji Zhao[1,#]

[1] *Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA*
[2] *Current Address: Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA*
[3] *Current Address: Dept. of Life Science, Pohang University of Science and Technology (POSTECH) San 31 Hyojadong, Pohang 790-784 Rep. of Korea*
[*] *These authors contributed equally to this work.*
[#] *Corresponding author*

Contact Information

Keji Zhao
LMI, NHLBI, NIH
Bldg 10; Room 7B05
9000 Rockville Pike
Bethesda, MD 20892
Phone: 301 496 2098
Fax: 301 480 0961
Email: zhaok@nhlbi.nih.gov

Contents

Supplementary Tables

Table S1. Summary of ChIP-Seq datasets.

Cell state	Protein	Residue	Modification	Mapped reads	Start[a]	End[a]	Threshold[b]	Antibody
Resting	H2A.Z			7.5E+06	-500	0	8	ab4174
	H2A	K5	ac	6.8E+06				ab1764
	H2A	K9	ac	2.1E+06				Up07-289
	H2B	K5	ac	3.3E+06				ab40886
	H2B	K5	me1	8.9E+06				ab12929
	H2B	K12	ac	3.6E+06				ab40883
	H2B	K20	ac	4.1E+06				Up07-347
	H2B	K12 0	ac	3.4E+06				Up07-564
	H3	K4	ac	3.5E+06				Up07-539
	H3	K4	me1	1.1E+07	500	1500		ab8895
	H3	K4	me2	5.4E+06				ab7766
	H3	K4	me3	1.7E+07	0	500	11	ab8580
	H3	K9	ac	4.0E+06				ab4441
	H3	K9	me1	9.3E+06	0	1000		ab8896
	H3	K14	ac	6.5E+06				Up07-353
	H3	K18	ac	4.2E+06				ab1191
	H3	K23	ac	6.5E+06				Up07-355
	H3	K27	ac	3.4E+06				ab4729
	H3	K27	me1	1.0E+07	0	TES		Up07-448
	H3	K27	me2	9.1E+06				ab24684
	H3	K27	me3	1.8E+07	0	TES		Up07-449
	H3	K36	ac	4.4E+06				Up07-540
	H3	K36	me3	1.4E+07	0	TES		ab9050
	H3	K79	me2	4.7E+06	0	TES		ab3594
	H4	K5	ac	4.1E+06				Up07-327
	H4	K8	ac	4.3E+06				Up07-328
	H4	K12	ac	5.5E+06				Up07-595
	H4	K16	ac	7.1E+06				sc-8662R
	H4	K20	me1	1.1E+07	0	TES		ab9051
	H4	K91	ac	3.2E+06				ab4627
	Pol II			1.1E+07	-250	250	9	ab5408
Activated	H2A.Z			3.8E+06	-500	0	6	see above
	H3	K4	me1	4.0E+06	500	1500		see above
	H3	K4	me3	3.4E+06	0	500	5	see above
	H3	K9	me1	7.1E+06	0	1000		see above
	H3	K27	me1	4.3E+06	0	TES		see above
	H3	K27	me3	7.4E+06	0	TES		see above
	H3	K36	me3	4.5E+06	0	TES		see above
	H3	K79	me2	5.6E+06	0	TES		see above
	H4	K20	me1	3.7E+06	0	TES		see above
	Pol II			8.6E+06	-250	250	8	see above

[a] Start and end values represent the start and end positions of region of interest with maximal difference in the enrichment of modification between the constitutively expressed (E→E) and constitutively silent (S→S) genes. Negative and positive numbers represent the number of base pairs up and downstream of transcription start site (TSS; represented as 0). TES denotes the transcription end site. Number of tags within the region of interest for each gene was used for generating the box-plots in Figs. 2 and S3 as well as the bar-plots in Figure 3A-C

[b] The minimum number of tags required within a gene's region of interest in order to classify that gene as containing statistically significant levels of a given chromatin modification ($p < 10^{-3}$). This was determined using Poisson probabilities, and was used to count the fraction of genes (within each set) containing a certain modification (as shown in Figure 3A-C).

Table S2. Gene ontology analysis of 167 genes induced by TCR signaling (S→E subset). Analysis for level 2 Biological Process terms was performed as described in Supplementary Methods- GO analysis. Total 9399 interogated genes were used as a background.

Term	Genes	%	P-Value	Benjamini
immune response	18	10.8	1.10E-05	1.60E-03
response to stress	21	12.6	4.20E-04	3.10E-02
response to chemical stimulus	15	9	6.10E-04	3.00E-02
cell proliferation	18	10.8	1.60E-03	5.60E-02
response to external stimulus	15	9	2.80E-03	8.00E-02
response to endogenous stimulus	10	6	3.70E-03	8.70E-02
behavior	10	6	7.00E-03	1.40E-01
cell cycle process	13	7.8	1.50E-02	2.50E-01
cell cycle	14	8.4	2.10E-02	3.00E-01
somatic diversification of immune receptors	3	1.8	2.40E-02	3.00E-01
cell division	6	3.6	2.40E-02	2.80E-01
immune system development	6	3.6	3.00E-02	3.20E-01
defense response	11	6.6	4.40E-02	4.00E-01
regulation of metabolic process	34	20.4	5.80E-02	4.70E-01
macromolecule metabolic process	67	40.1	6.30E-02	4.80E-01
regulation of cellular process	46	27.5	6.90E-02	4.90E-01
leukocyte activation	5	3	7.60E-02	5.00E-01
cytokine production	4	2.4	8.90E-02	5.40E-01

Table S3. miRNA expression and promoter identification.

Table shows number of sequencing tags mapped to **miR** or **miR*** and predicted promoters. Table is provided as a separate Excel file. See key and summary sheets within the Excel document for details.

.

Table S4. Statistical analysis of tag density in gene subsets. Table shows p-values obtained using two-tailed Wilcoxon rank-sum test comparing pairwise median tag numbers of two gene sets. Table is provided as a separate Excel file.

Table S5. Gene ontology analysis of Pol II poised genes. 1,143 silent genes possessing RNA polymerase II at their promoters were analyzed. Analysis for level 2 GO biological process terms was performed as described in Supplementary Methods- GO analysis. Total of 5484 silent genes were used as a background.

Term	Genes	%	P-Value	Benjamini
primary metabolic process	389	35	3.60E-23	5.20E-21
cellular metabolic process	381	34.2	2.40E-21	1.70E-19
macromolecule metabolic process	338	30.4	3.10E-21	1.50E-19
regulation of cellular process	226	20.3	7.80E-08	2.80E-06
transcription	148	13.3	1.30E-07	3.80E-06
regulation of metabolic process	163	14.6	2.20E-07	5.20E-06
cellular component organization and biogenesis	144	12.9	3.00E-07	6.20E-06
regulation of gene expression	146	13.1	2.70E-06	4.80E-05
regulation of biological process	236	21.2	3.00E-06	4.90E-05
cell cycle	44	4	3.10E-04	4.50E-03
macromolecule localization	37	3.3	8.70E-04	1.10E-02
response to endogenous stimulus	23	2.1	1.20E-03	1.40E-02
cell division	13	1.2	1.40E-03	1.60E-02
cell cycle process	38	3.4	2.00E-03	2.00E-02
establishment of protein localization	32	2.9	3.30E-03	3.10E-02
establishment of cellular localization	38	3.4	2.60E-02	2.10E-01
cellular localization	39	3.5	3.40E-02	2.50E-01
ensheathment of neurons	5	0.4	4.70E-02	3.20E-01
response to abiotic stimulus	16	1.4	6.40E-02	3.90E-01
RNA processing	12	1.1	7.70E-02	4.40E-01

Table S6. Primers used in 5'RACE experiment.

miRNA	RT primer	Nested primer
hsa-mir-146b	AGCCTATGGAATTCAGTTCTCAGTG	GGAGGAGAAAGAGTTCCCTGAAGCACA
hsa-mir-132	AGTAACAATCGAAAGCCACGGTTGC	AACAATCGAAAGCCACGGTTGCCCT
hsa-mir-24-2		
hsa-mir-23a		
hsa-mir-27a	TGCTCACAAGCAGCTAAGCCCTGCT	CCAGAGGCTGGCACCTGGAGGGGAGAA
hsa-mir-150	CACTGGTACAAGGGTTGGGAGACAG	RT primer was used in PCR
hsa-mir-125a	TCACAGGTTAAAGGGTCTCAGGGACCT	GGGGGTGGGGTGGGGGTGGTTTGAGAA
hsa-mir-99b		
hsa-let-7e		
hsa-mir-148a	AGTCGGAGTGTCTCAGAACTTTGCC	CCGCTCCGCTCCCTTCCATCTTGACTT
hsa-mir-363	AAATTGCATCGTGATCCACCCGACA	TTTCGCCCTTGCCAGGCGCCCTCTTCA
hsa-mir-20b		
hsa-mir-19b-2		
hsa-mir-106a		
hsa-mir-92a-2		
hsa-mir-18b		

Table S7. Primers used for cloning miRNA promoters into pGL3 Enhancer vector.

3'Bgl-con1	CTGCAAGATCTTATGCATGTTGGGGAAGACA
3'Bgl-con2	CTGCAAGATCTACTCCCAGGCCCTACGTAAT
3'Bgl-mir-122	CTGCAAGATCTAATGTCCCCTACCTGCACAC
3'Bgl-mir-187	CTGCAAGATCTCACTGGGTCCCCCTCCTG
3'Bgl-mir-199a-2	CTGCAAGATCTTTGGCATGTGGCATTTACTT
3'Bgl-mir-422a	CTGCAAGATCTCTCAGTCTTTGGAGCCAGGT
3'Bgl-mir-647	CTGCAAGATCTTCTGTCTTTGAAGGTGGCATC
3'Bgl-mir-760	CTGCAAGATCTATCCTCCGAGAGCTCGTCTT
5'Kpn-con1	CACGTGGTACCCACCTGTGTGCCATGTATCC
5'Kpn-con2	CACGTGGTACCTCATATGCCACCCATCAAGA
5'Kpn-mir-122	CACGTGGTACCACTCCTGGGTAGGGCCTAAG
5'Kpn-mir-187	CACGTGGTACCGGAGCCCAGCCAGATTGTAT
5'Kpn-mir-199a-2	CACGTGGTACCGGTAGTGTAATCTTTCAGGGAATTG
5'Kpn-mir-422a	CACGTGGTACCGTTGCACTCGGAATGTGGTT
5'Kpn-mir-647	CACGTGGTACCCTTGCAGGCCCTTTTGTTTA
5'Kpn-mir-760	CACGTGGTACCCACTGGACTCTTCCCAGGAC

Supplementary Figure Legends

Figure S1. Purification and activation of $CD4^+$ T cells. (A) CD4+ T cells were purified by negative selection to 95-98% purity. (B) T cells were incubated with anti-CD3/28 beads for 18 hours, which resulted in 85-95% activation as indicated by the percentage of CD25-positive cells. (C) Box plot shows the distribution of gene expression values (log_2 scale) in resting and activated T cells for each of the four gene sets. Expression was defined as probe-set intensity divided by sample median intensity. The data points for each gene set were divided into quartiles, and the interquartile range (IQR) is calculated as the difference between the first and the third quartiles. The filled box denotes the middle 50% of the data points, with the horizontal line in between and the notch representing the median and confidence intervals, respectively. Data points more than 1.5 times IQR lower or higher than first or third quartiles, respectively represent outliers and are shown as dots. The horizontal line that is connected by vertical dashed lines, above and below the filled box (whiskers), represents the largest and the smallest non-outlier data points.

Figure S2. **Examples of genes whose chromatin status does not change upon induction.** Leukemia Inhibitory Factor (*LIF*) (A) and Cytokine Induced SH2-containing Protein (*CISH*) (B) genes are not expressed in resting T cells, but are induced upon T cell activation. Active chromatin modifications are present at the promoter and in the gene body even when these genes are silent in resting cells.

Figure S3. ChIP-Seq tag density profiles for expression-based gene sets in resting and activated T cells. Gene sets: S→E (Silent in resting and Expressed in activated T cells), red broken line; E→S, blue broken line; E→E, red line; S→S, blue line. (A) H2A.Z, (B) H3K4me1, (C) H3K9me1, (D) H3K27me1, (E) H3K36me3, (F) H4K20me1, (G) Pol II. Box-plot summarizes the distribution of the number of tags in the region of interest (highlighted in yellow) for each gene set. The average tag density values for resting and activated cells are not directly comparable as they have not been normalized across samples. Box-plot captures the median, the middle 50% of the data points, and the outliers. The data points for each gene set is divided into quartiles, and the interquartile range (IQR) is calculated as the difference between the first and the third quartiles. The filled box denotes the middle 50% of the data points, with the horizontal line in between and the notch representing the median and confidence intervals, respectively. Data points more than 1.5 times IQR lower or higher than first or third quartiles, respectively represent outliers and are shown as dots. The horizontal line that is connected by vertical dashed lines above and below the filled box (whiskers) represents the largest and the smallest nonoutlier data points. The cluster of horizontal red and black lines below each box-plot signifies whether or not the difference between the medians of two gene sets are statistically significant, respectively ($p<0.01$; two-tailed Wilcoxon rank-sum test).

Figure S4. ChIP-Seq tag density profiles for expression-based gene sets in resting T cells. Gene sets: S→E (Silent in resting and Expressed in activated T cells), red broken line; E→S, blue broken line; E→E, red line; S→S, blue line. (A) H2BK5ac, (B) H2BK12ac, (C) H2BK20ac, (D) H2BK120ac, (E) H3K4ac, (F) H3K4me2, (G) H3K9ac,

(H) H3K18ac, (I) H3K23ac, (J) H3K27ac, (K) H3K36ac, (L) H4K5ac, (M) H4K8ac, (N) H4K12ac, (O) H4K91ac, (P) H2AK9ac, (R) H2BK5me1, (S) H4K16ac.

Figure S5. Comparison of PolII levels in the gene body.

Pol II tag density profile is shown on the left. Box-plot (center) summarizes the distribution of the number of tags in the region of interest (highlighted in yellow) for each gene set. See Fig. S3 legend for box-plot description. The cluster of horizontal red and black lines below each box-plot signifies whether or not the difference between the medians of two gene sets are statistically significant, respectively ($p<0.01$; two-tailed Wilcoxon rank-sum test). Table on the right indicates actual p-values. Significant values are shown in bold.

Figure S6. Accuracy of promoters predicted using chromatin modification patterns.

The promoter prediction algorithm (See SI Methods) was used to predict transcription start sites for 1000 highly expressed genes in resting T cells. Frequency distribution (A) and cumulative distribution (B) of predicted transcription start sites relative to the annotated start site is shown.

Figure S7. 5'RACE experiment.

(A) Experimental approach. Pri-miRNA was reverse transcribed using RT primer. After reaching the end of template MMLV Reverse Transcriptase adds several C nucleotides and switches template to Smart oligo (Clontech). Resulting cDNA is amplified by PCR with Universal primer and either RT primer or nested primer. (B) Table shows seven

clusters of miRNAs (16 miRNAs) located within 5 kb of their putative promoter. miRNAs transcribed from the same TSSs are shown by color. Columns show distance from miRNA to predicted TSS, expected and obtained PCR amplicon size. (C) Agarose gel analysis of PCR fragments obtained in 5'RACE experiment. Colors were inverted and contrast adjusted for the whole picture for printing.

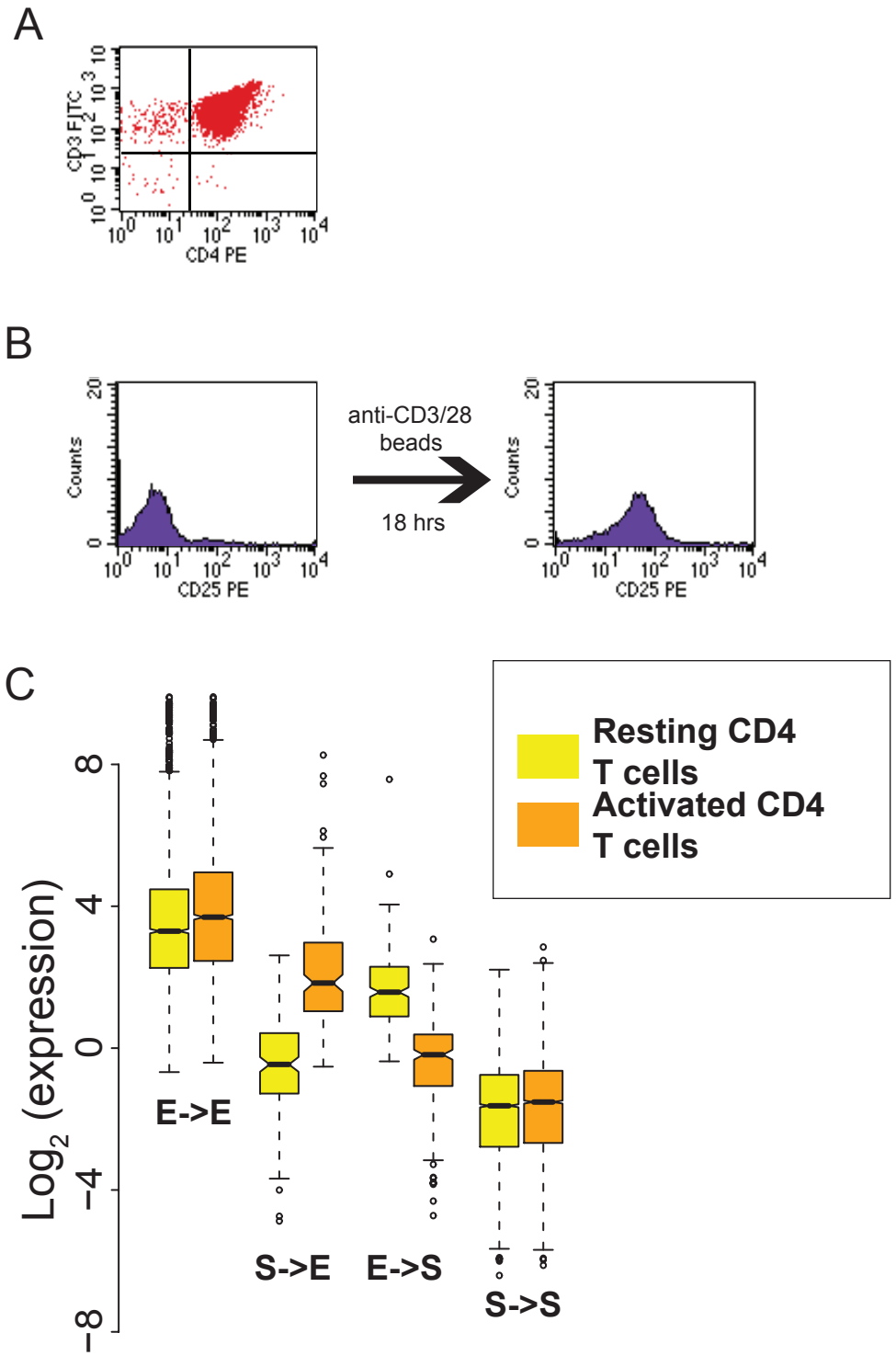
Figure S8. A majority of intragenic miRNAs share promoters with their host genes. Chromatin modification patterns in the region surrounding the intragenic *MIR185* (A), and *MIR101-2* (B) are shown. Potential promoter regions are marked by H3K4me3, H2A.Z and Pol II peaks. The green bars extend from the predicted transcription start site to the 3' end of the pre-miRNA.

Figure S9. ChIP-Seq tag density profiles near predicted TSSs of miRNA genes. Modification patterns near all, intra- and intergenic miRNA promoters are shown in resting T cells. (A) H2A.Z, (B) H3K4me3, (C) Pol II, (D) H3K4me1, (E) H3K4me2, (F) H4K9ac, (G) H3K9me1, (H) H3K14ac, (I) H3K27ac, (J) H4K20me1, (K) H3K36me3. While H3K4me3, H2A.Z and Pol II were used for promoter prediction and thus one can expect to see their peaks at the promoters, the rest of modifications were enriched at promoters independently of the prediction algorithm. Due to miRNA TSS being determined imprecisely, we are not able to resolve two peaks that we see surrounding TSS of "usual" genes in the case of miRNA genes.

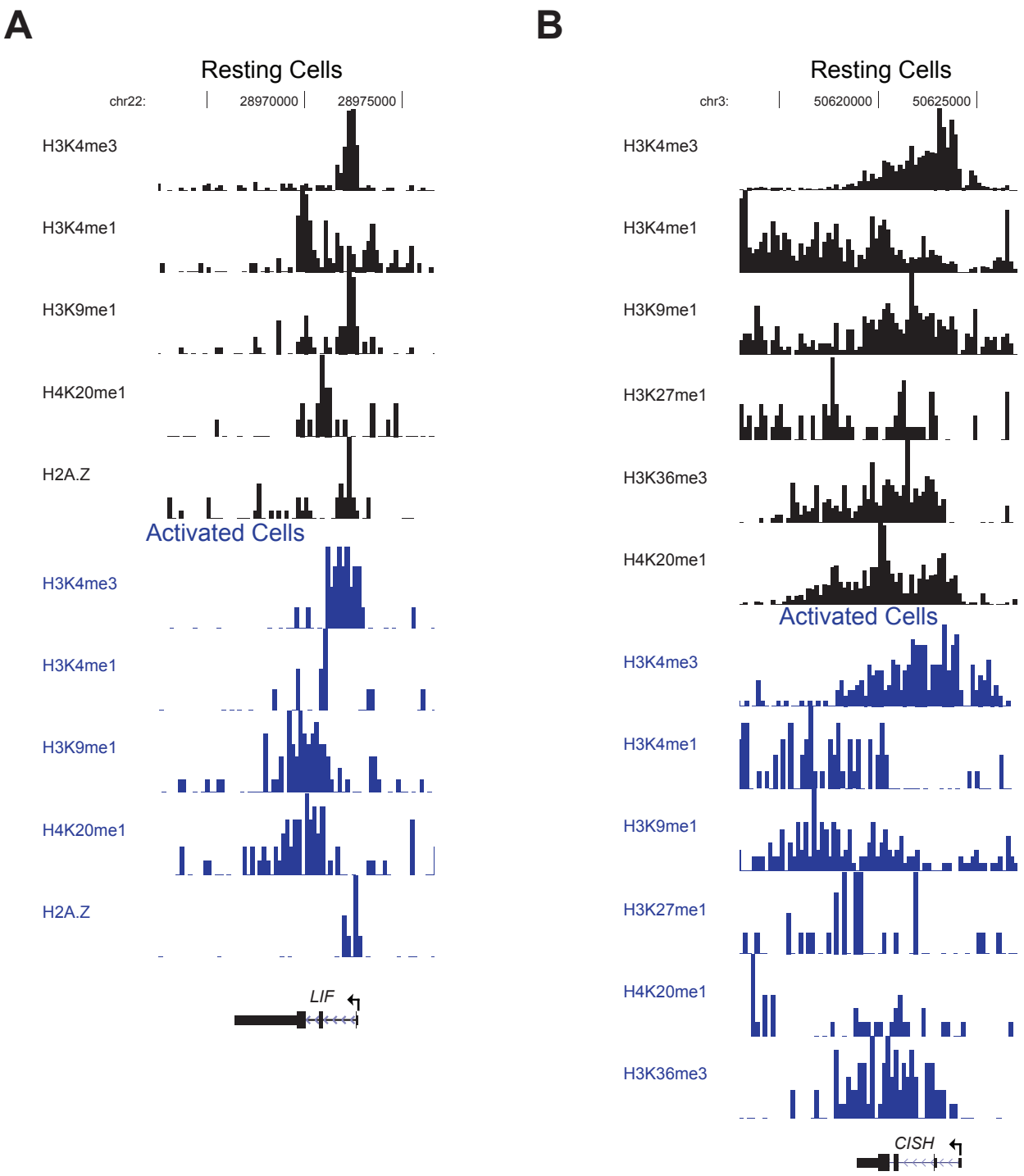
Figure S10. Examples of inducible miRNA genes that are poised for expression. (A) *MIR147B* expression increased from 1 read in resting T cells to 10 in activated. (B) *MIR7-1* expression increased from 169 reads in resting T cells to 2206 in activated.

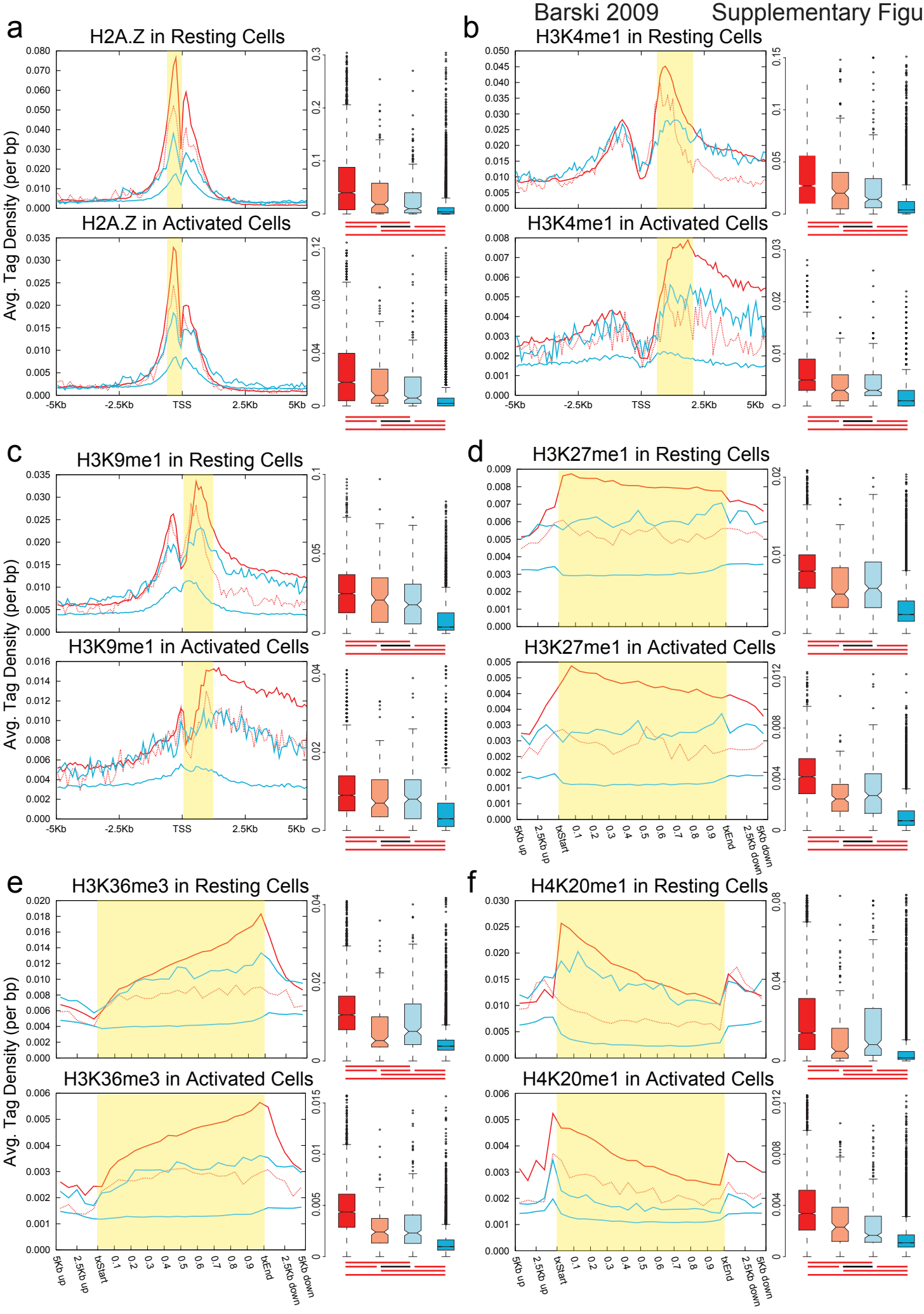
Figure S11. Use of more stringent gene-set definition does not influence results. Alternative gene sets were defined as described in Methods. Panels show tag density profiles in resting (top) and activated (bottom) cells for (A) H2A.Z, (B) H3K4me3, (C) Pol II, (D) H3K36me3, (E) H3K79me2, (F) H3K27me3. Gene sets: S→E (Silent in resting and Expressed in activated T cells), red broken line; E→S, blue broken line; E→E, red line; S→S, blue line.

Supplementary Figures

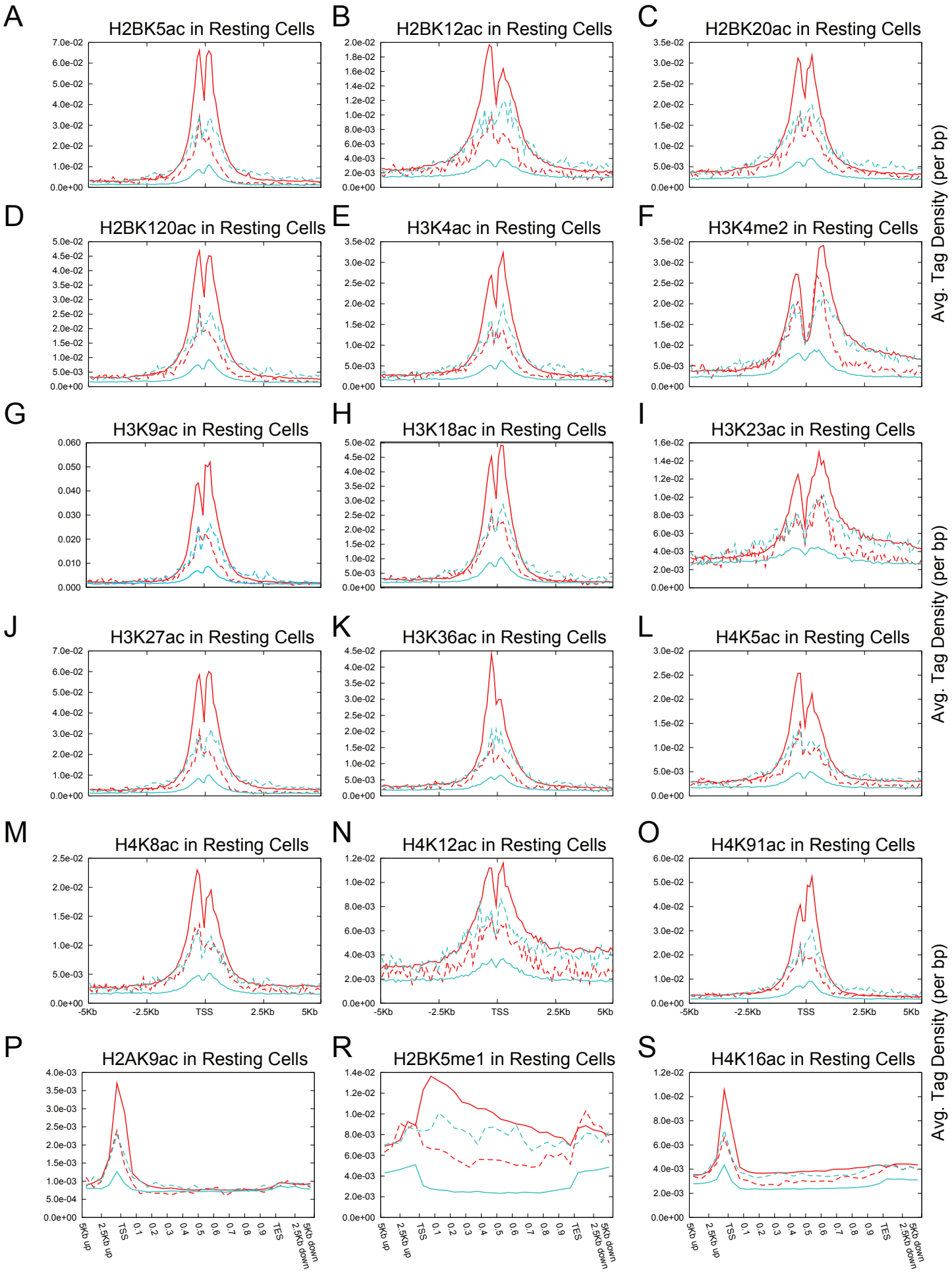

A
CD3 FITC
CD4 PE
B
Counts
CD25 PE
anti-CD3/28
beads
18 hrs
C
Resting CD4
T cells
Activated CD4
T cells
Log2 (expression)
E->E
S->E
E->S
S->S

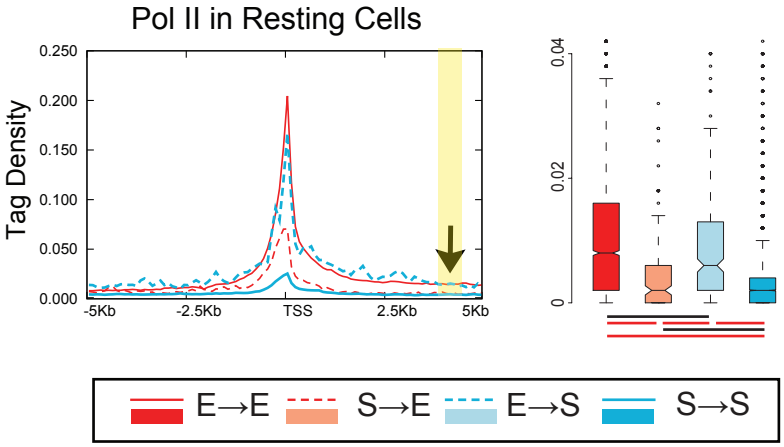
Barski 2009 Supplementary Figure 2




a
H2A.Z in Resting Cells
H2A.Z in Activated Cells
b
H3K4me1 in Resting Cells
H3K4me1 in Activated Cells
c
H3K9me1 in Resting Cells
H3K9me1 in Activated Cells
d
H3K27me1 in Resting Cells
H3K27me1 in Activated Cells
e
H3K36me3 in Resting Cells
H3K36me3 in Activated Cells
f
H4K20me1 in Resting Cells
H4K20me1 in Activated Cells
Avg. Tag Density (per bp)
-5Kb
-2.5Kb
TSS
2.5Kb
5Kb
5Kb up
2.5Kb up
txStart
txEnd
2.5Kb down
5Kb down

Barski 2009 Supplementary Figure 4

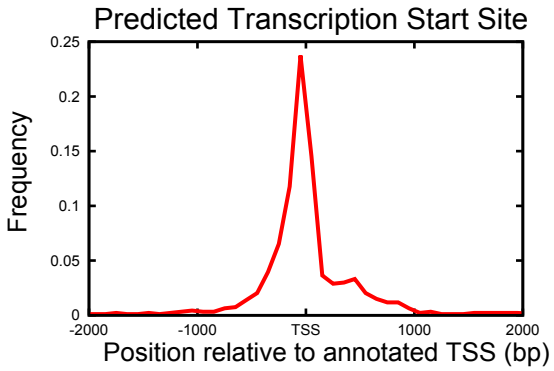




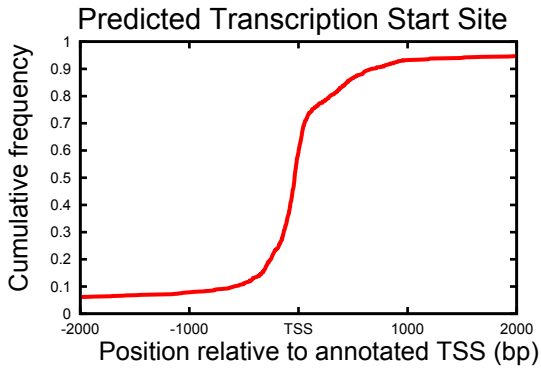
Wilcoxon (2-tailed) test	PolII
Region Start	4 kb down TSS
Region End	4.5 kb down TSS
E→E vs S→E	1.91E-16
E→E vs E→S	0.04
S→E vs E→S	8.05E-09
S→E vs S→S	0.25
E→S vs S→S	5.43E-24
E→E vs S→S	0.00E+00

Barski 2009 Supplementary Figure 6

A

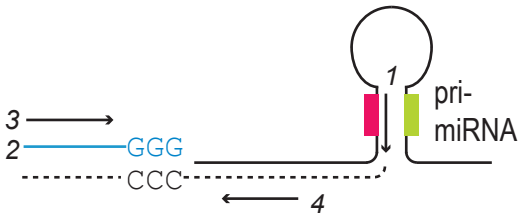


B



Barski 2009 Supplementary Figure 7

A



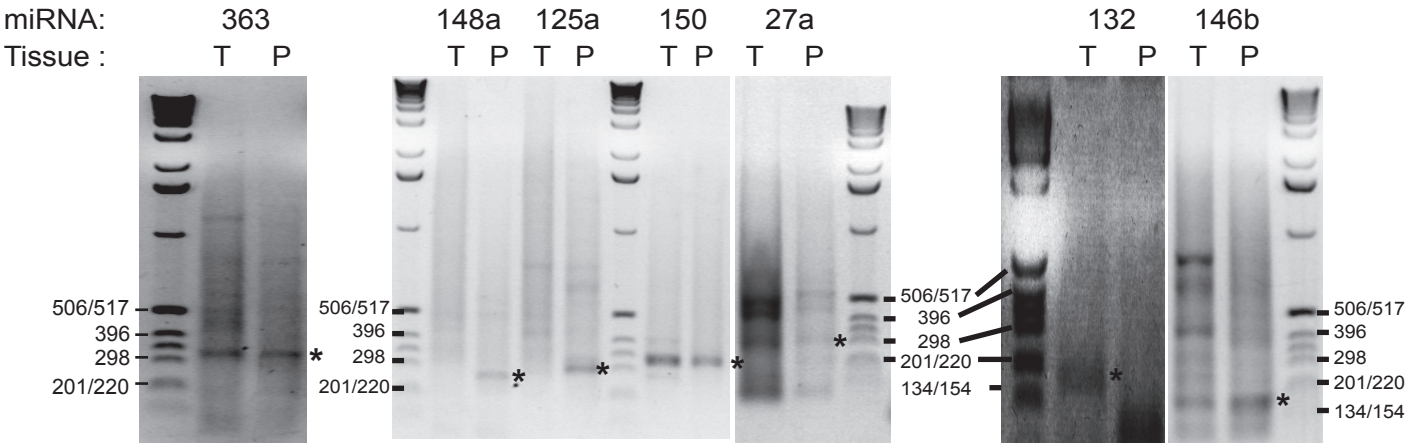
1 Reverse transcription primer
2 Smart Oligo
3 Universal primer
4 Nested primer

B

miRNA	Distance*	Amplicon size	
		Expected	Obtained**
hsa-mir-363	3746	226	320
hsa-mir-20b	3321		
hsa-mir-19b-2	3432		
hsa-mir-106a	2920		
hsa-mir-92a-2	3586		
hsa-mir-18b	3087		
hsa-mir-148a	1354	252	270
hsa-mir-125a	2336	278	300
hsa-mir-99b	1694		
hsa-let-7e	1868		
hsa-mir-150	160	242	330
hsa-mir-27a	386	256	300, 450,500
hsa-mir-23a	244		
hsa-mir-24-2	544		
hsa-mir-132	148	245	190
hsa-mir-146b	695	222	180

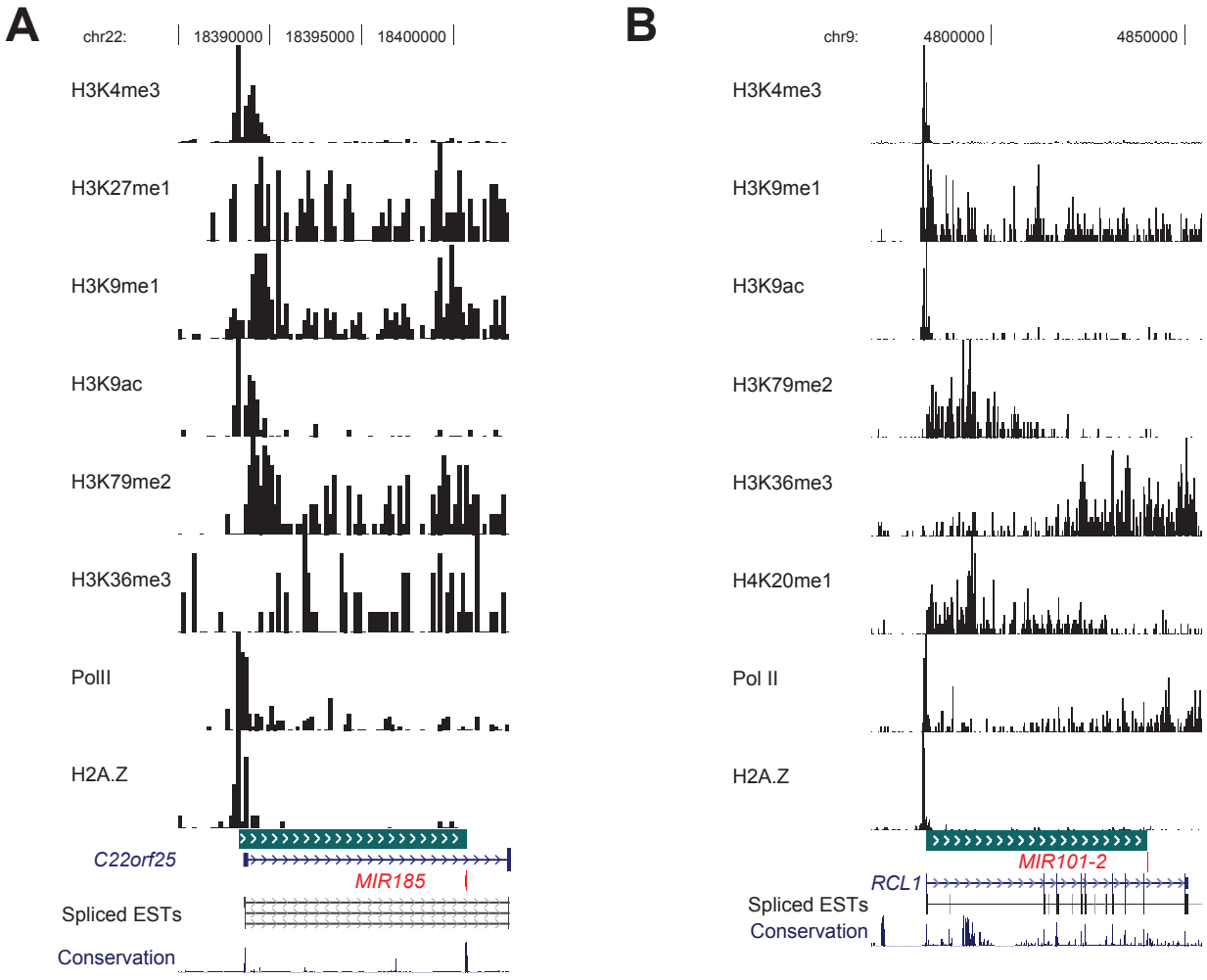
* Distance between predicted TSS and pre-miRNA.
** Estimate

C

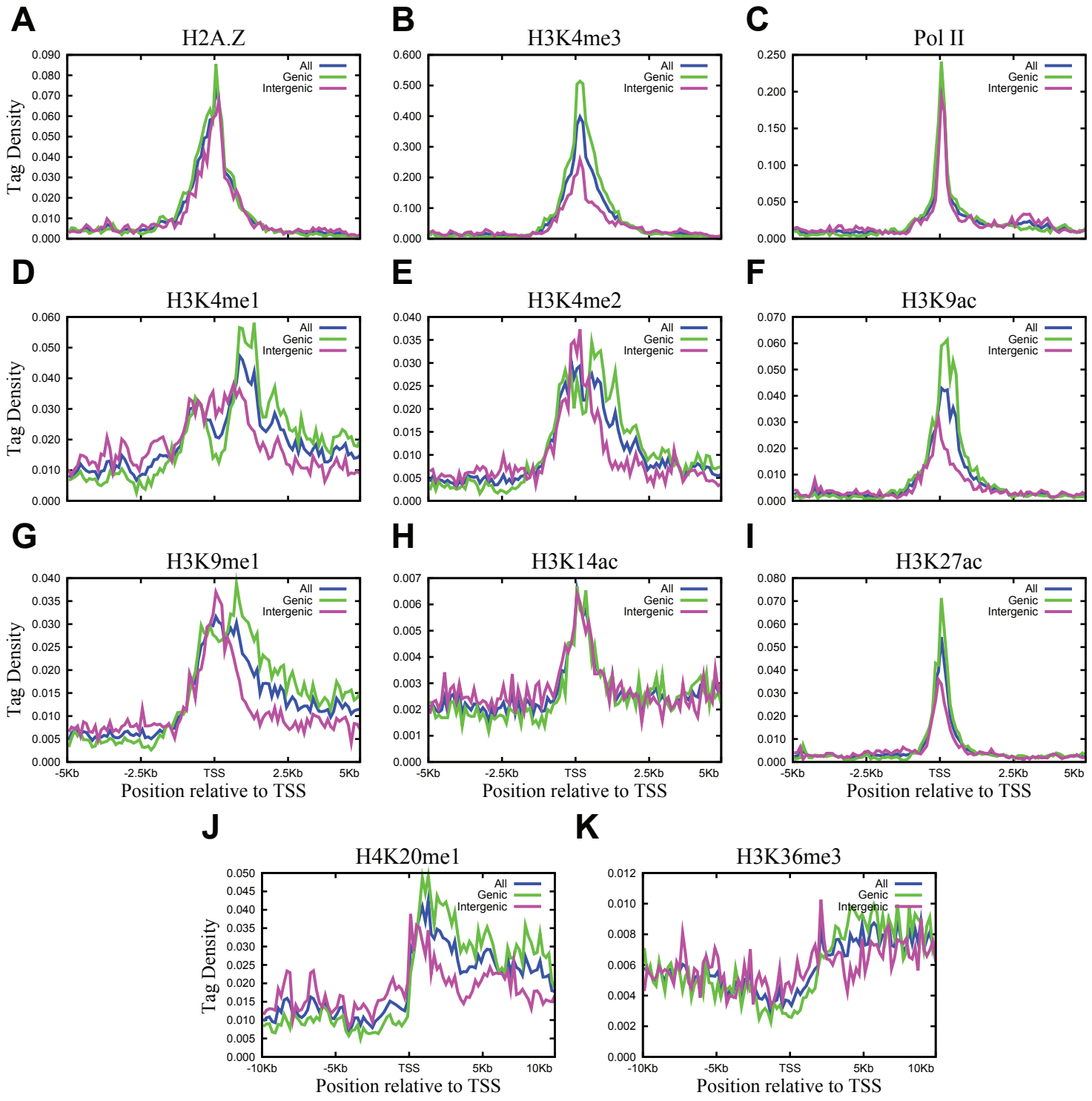


P= placenta, T = CD4 T cells (resting and activated)

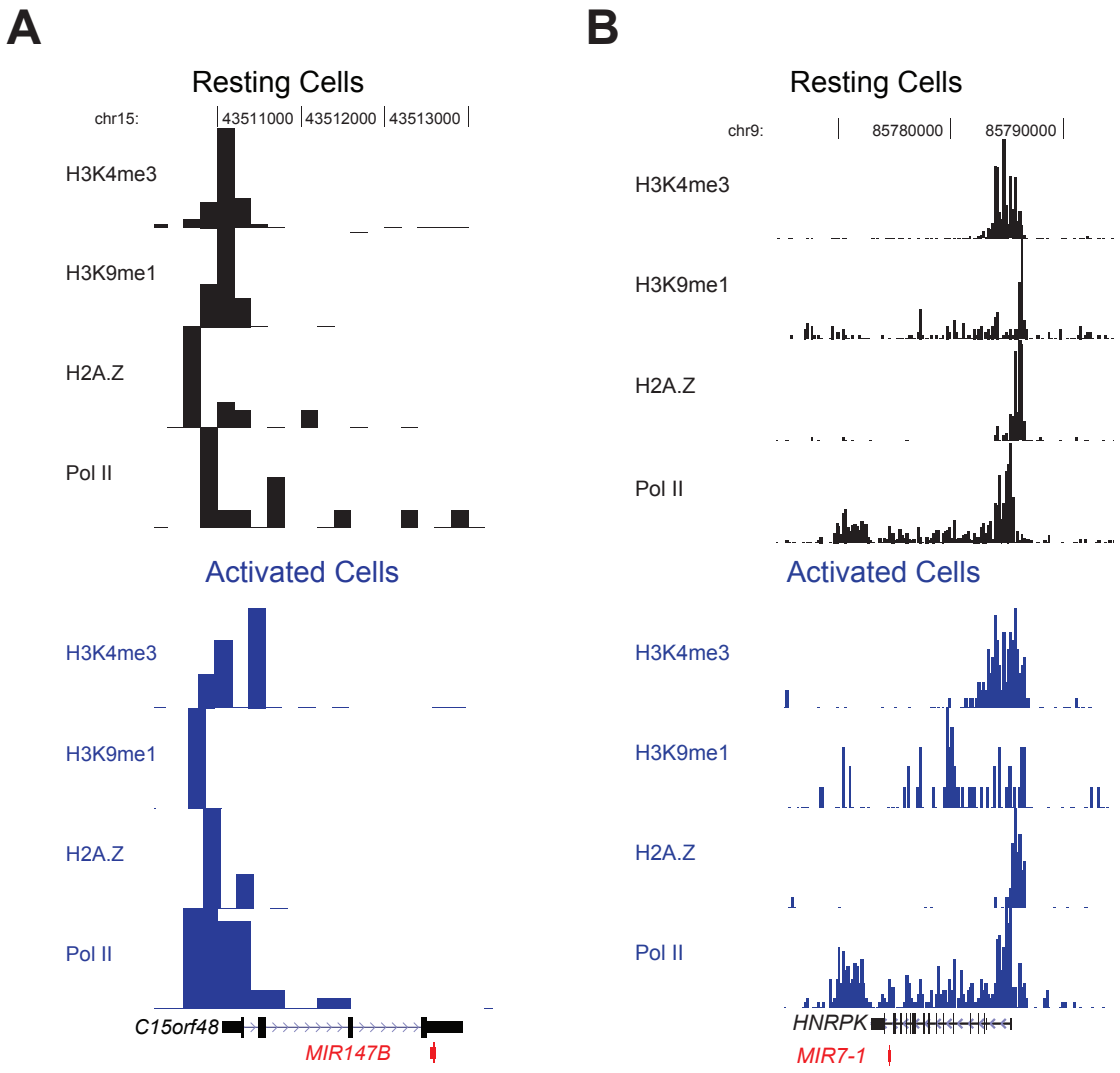
Barski 2009 Supplementary Figure 8



Barski 2009 Supplementary Figure 9



Barski 2009 Supplementary Figure 10



Barski 2009 Supplementary Figure 11