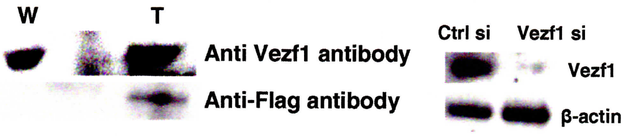
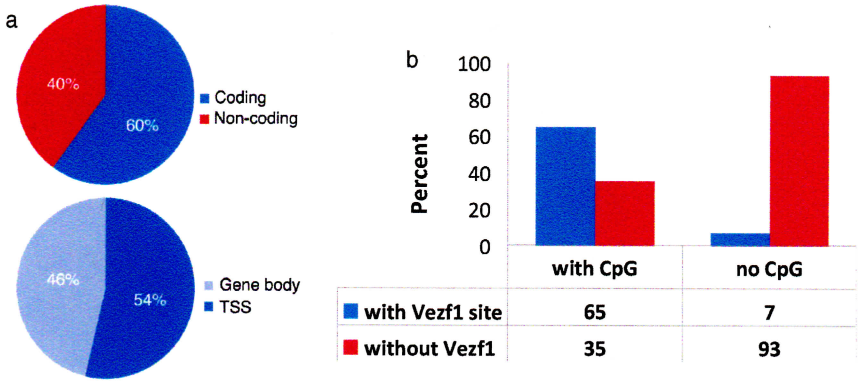
Gowher_ S1

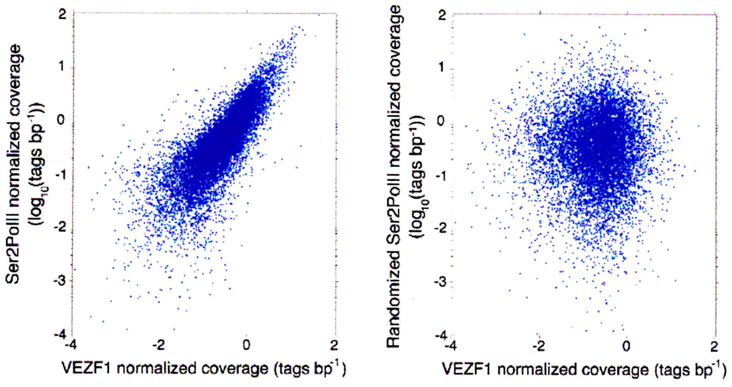


Gowher_S2



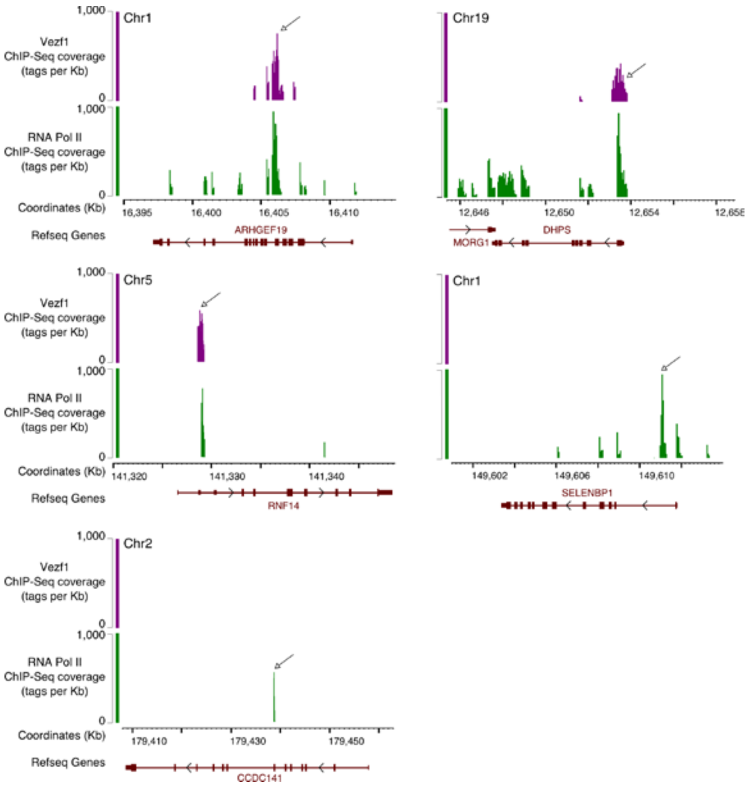
	with CpG	no CpG
🟦 with Vezf1 site	65	7
🟥 without Vezf1	35	93

Gowher_ S3



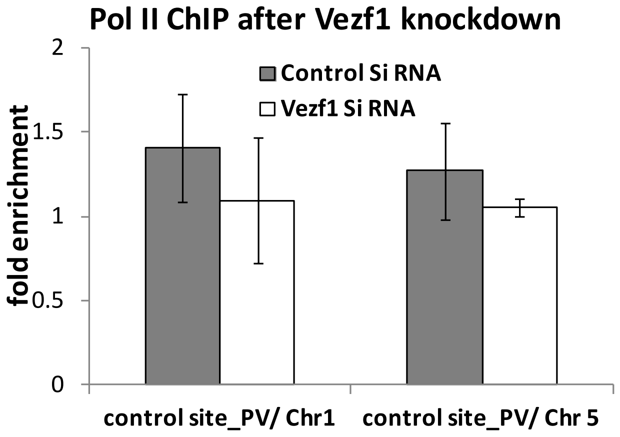
Gowher_S4

a



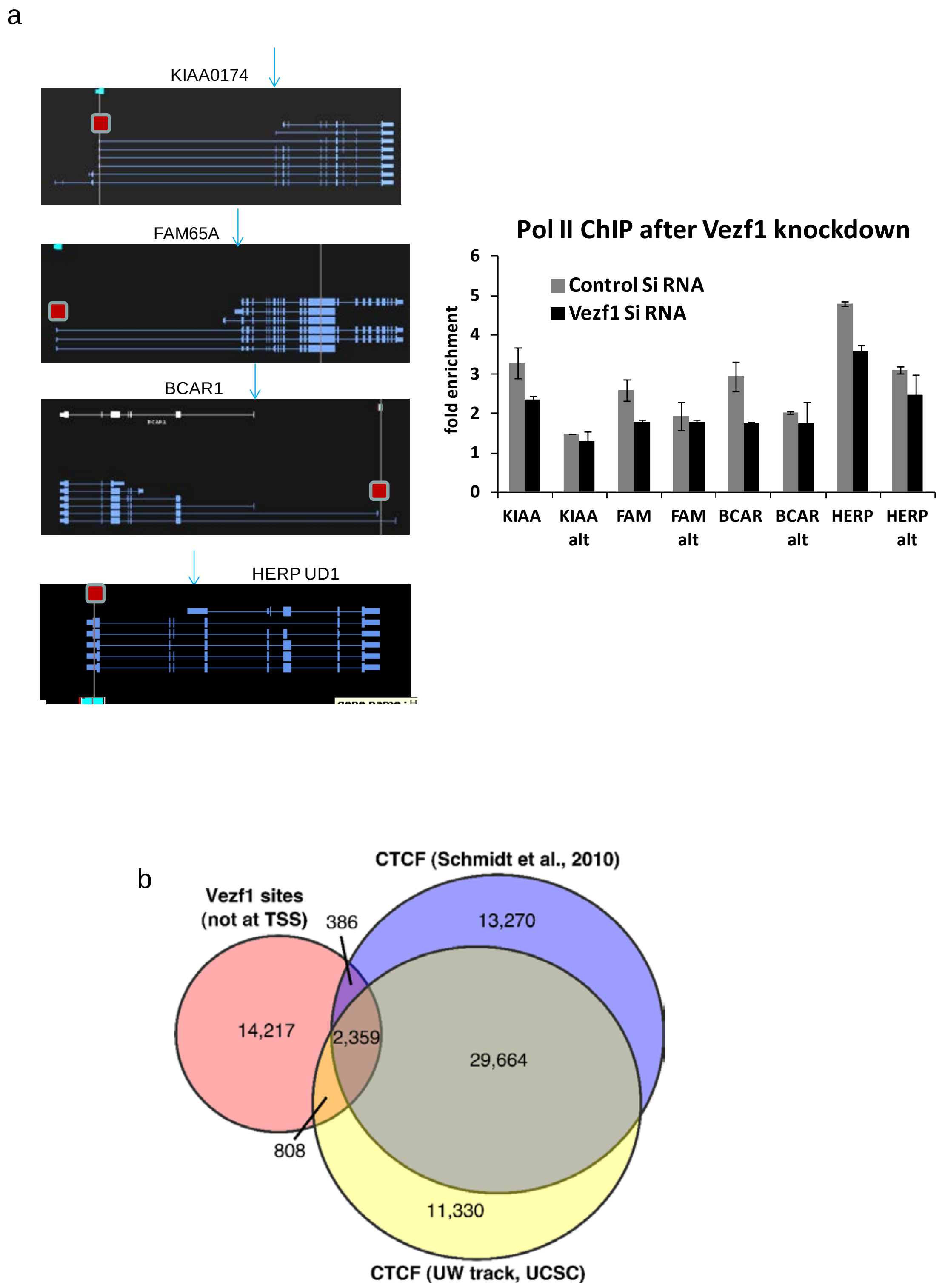
Chr1
Vezf1 ChIP-Seq coverage (tags per Kb)
1,000
0
RNA Pol II ChIP-Seq coverage (tags per Kb)
1,000
0
Coordinates (Kb)
16,395
16,400
16,405
16,410
Refseq Genes
ARHGEF19
Chr19
12,646
12,650
12,654
12,658
MORG1
DHPS
Chr5
141,320
141,330
141,340
RNF14
Chr1
149,602
149,606
149,610
SELENBP1
Chr2
179,410
179,430
179,450
CCDC141

b

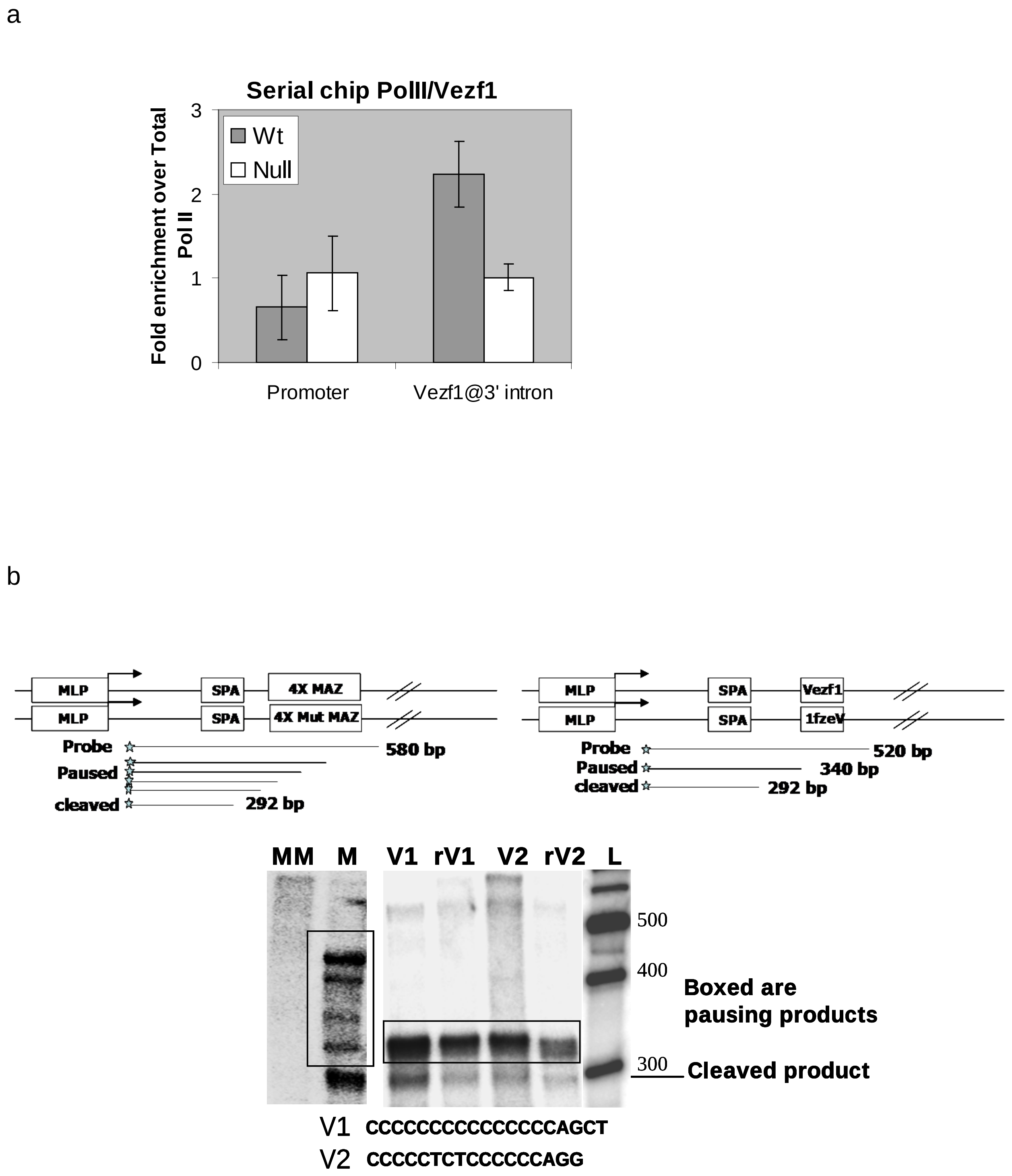


Pol II ChIP after Vezf1 knockdown
2
1.5
1
0.5
0
fold enrichment
■ Control Si RNA
□ Vezf1 Si RNA
control site_PV/ Chr1
control site_PV/ Chr 5

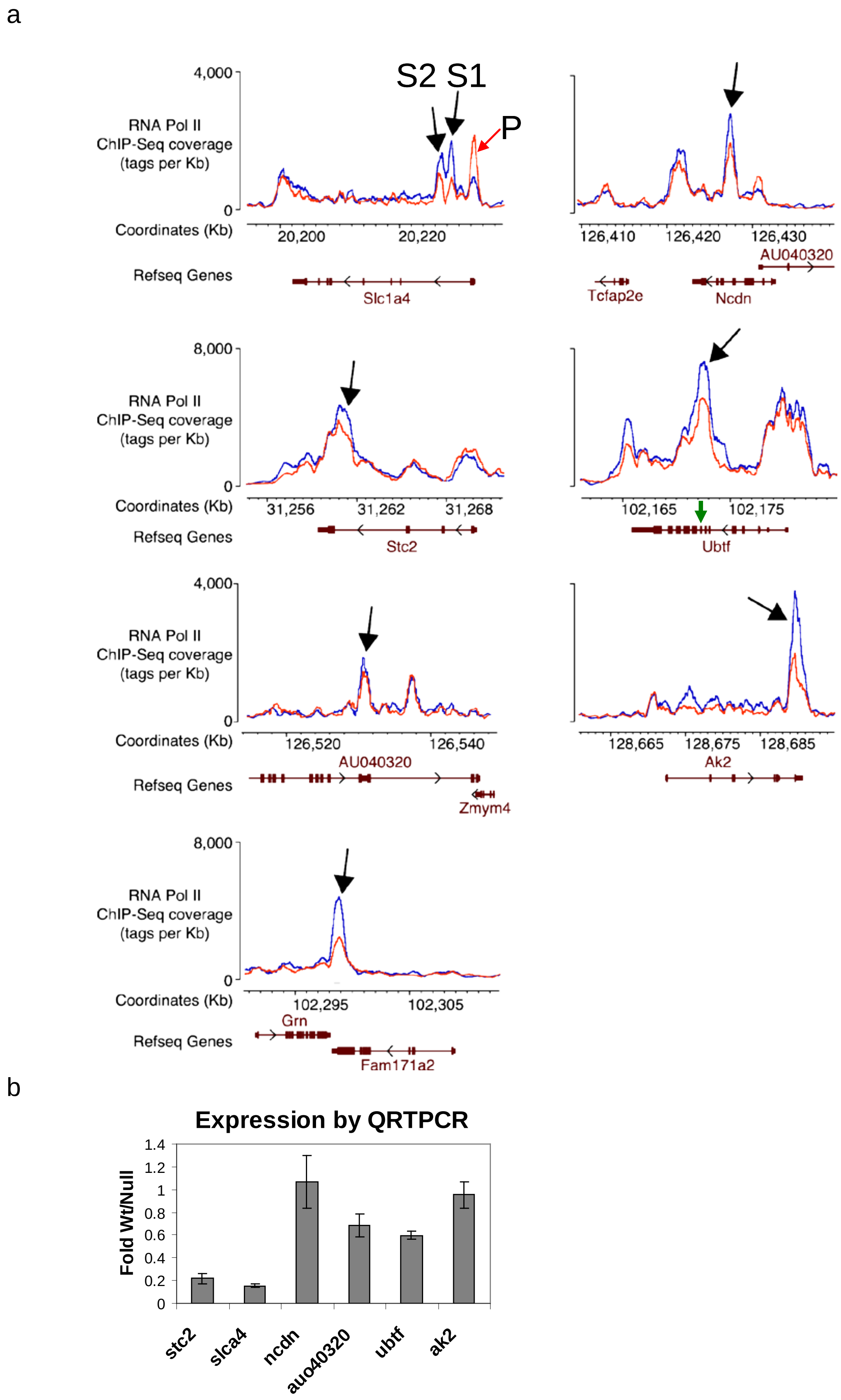
Gowher_S5


a
KIAA0174
FAM65A
BCAR1
HERP UD1
Pol II ChIP after Vezf1 knockdown
■ Control Si RNA
■ Vezf1 Si RNA
fold enrichment
0
1
2
3
4
5
6
KIAA
KIAA alt
FAM
FAM alt
BCAR
BCAR alt
HERP
HERP alt
b
Vezf1 sites (not at TSS)
CTCF (Schmidt et al., 2010)
CTCF (UW track, UCSC)
14,217
386
13,270
2,359
29,664
808
11,330

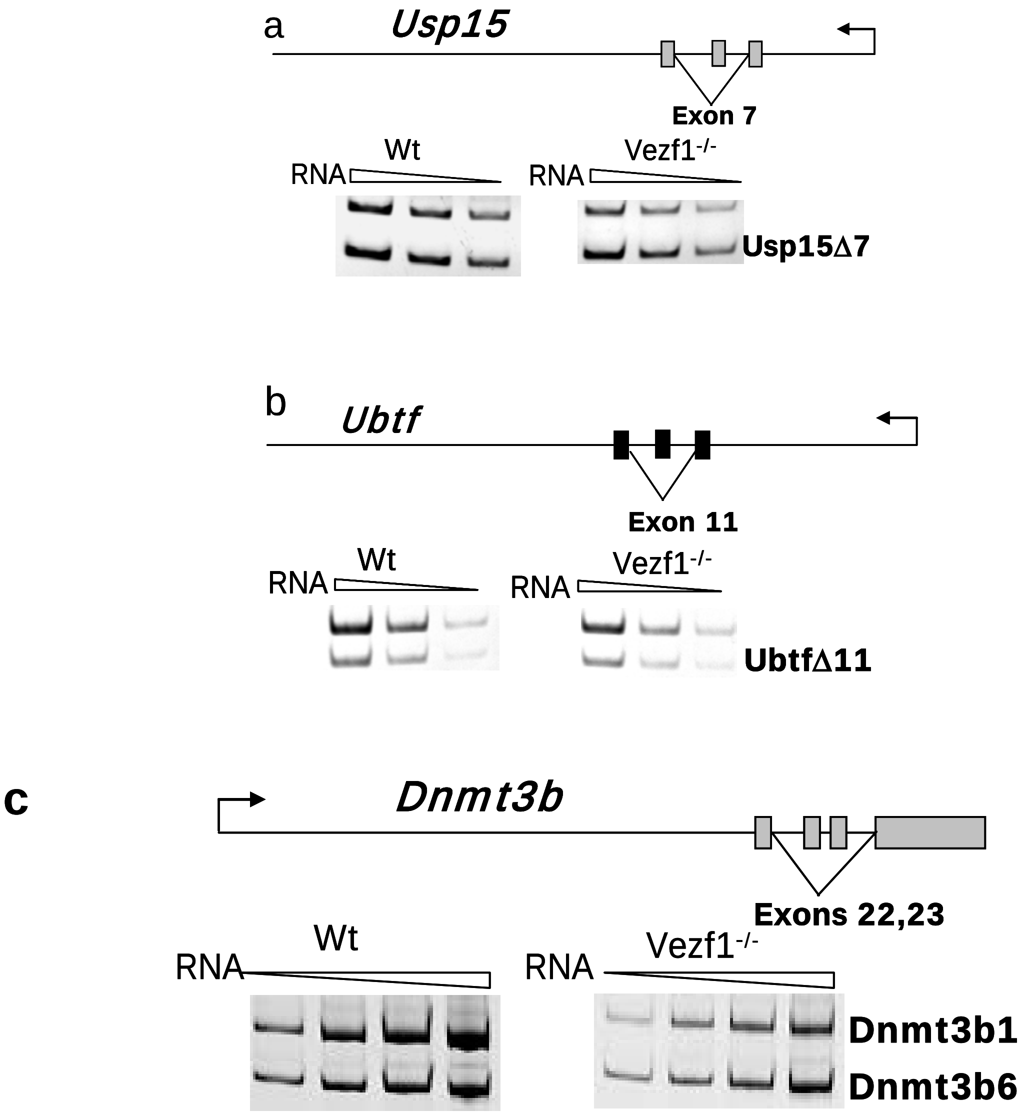
Gowher_S6


a
Serial chip PolII/Vezf1
Fold enrichment over Total Pol II
3
2
1
0
■ Wt
□ Null
Promoter
Vezf1@3' intron
b
MLP
SPA
4X MAZ
4X Mut MAZ
Probe
Paused
cleaved
580 bp
292 bp
MLP
SPA
Vezf1
1fzeV
Probe
Paused
cleaved
520 bp
340 bp
292 bp
MM
M
V1
rV1
V2
rV2
L
500
400
300
Boxed are pausing products
Cleaved product
V1 CCCCCCCCCCCCCCCAGCT
V2 CCCCCTCTCCCCCCAGG

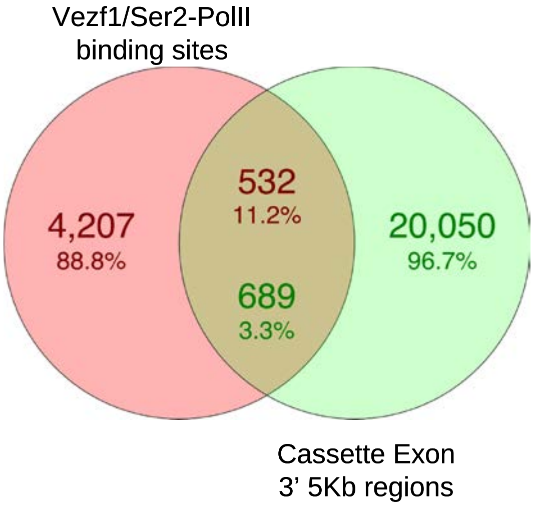
Gowher_S7


a
4,000
RNA Pol II
ChIP-Seq coverage
(tags per Kb)
0
S2 S1
P
Coordinates (Kb)
20,200
20,220
Refseq Genes
Slc1a4
126,410
126,420
126,430
AU040320
Tcfap2e
Ncdn
8,000
RNA Pol II
ChIP-Seq coverage
(tags per Kb)
0
Coordinates (Kb)
31,256
31,262
31,268
Refseq Genes
Stc2
102,165
102,175
Ubtf
4,000
RNA Pol II
ChIP-Seq coverage
(tags per Kb)
0
Coordinates (Kb)
126,520
126,540
AU040320
Refseq Genes
Zmym4
128,665
128,675
128,685
Ak2
8,000
RNA Pol II
ChIP-Seq coverage
(tags per Kb)
0
Coordinates (Kb)
102,295
102,305
Grn
Refseq Genes
Fam171a2
b
Expression by QRTPCR
Fold Wt/Null
1.4
1.2
1
0.8
0.6
0.4
0.2
0
stc2
slca4
ncdn
auo40320
ubtf
ak2

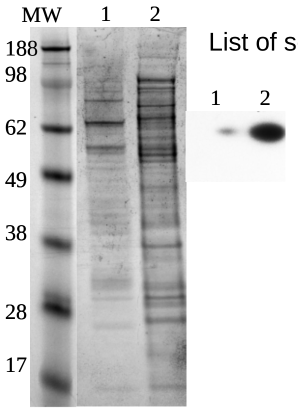
Gowher_S8


a
Usp15
Exon 7
Wt
Vezf1$^{-/-}$
RNA
RNA
Usp15Δ7
b
Ubtf
Exon 11
Wt
Vezf1$^{-/-}$
RNA
RNA
UbtfΔ11
c
Dnmt3b
Exons 22,23
Wt
Vezf1$^{-/-}$
RNA
RNA
Dnmt3b1
Dnmt3b6

Gowher_S9



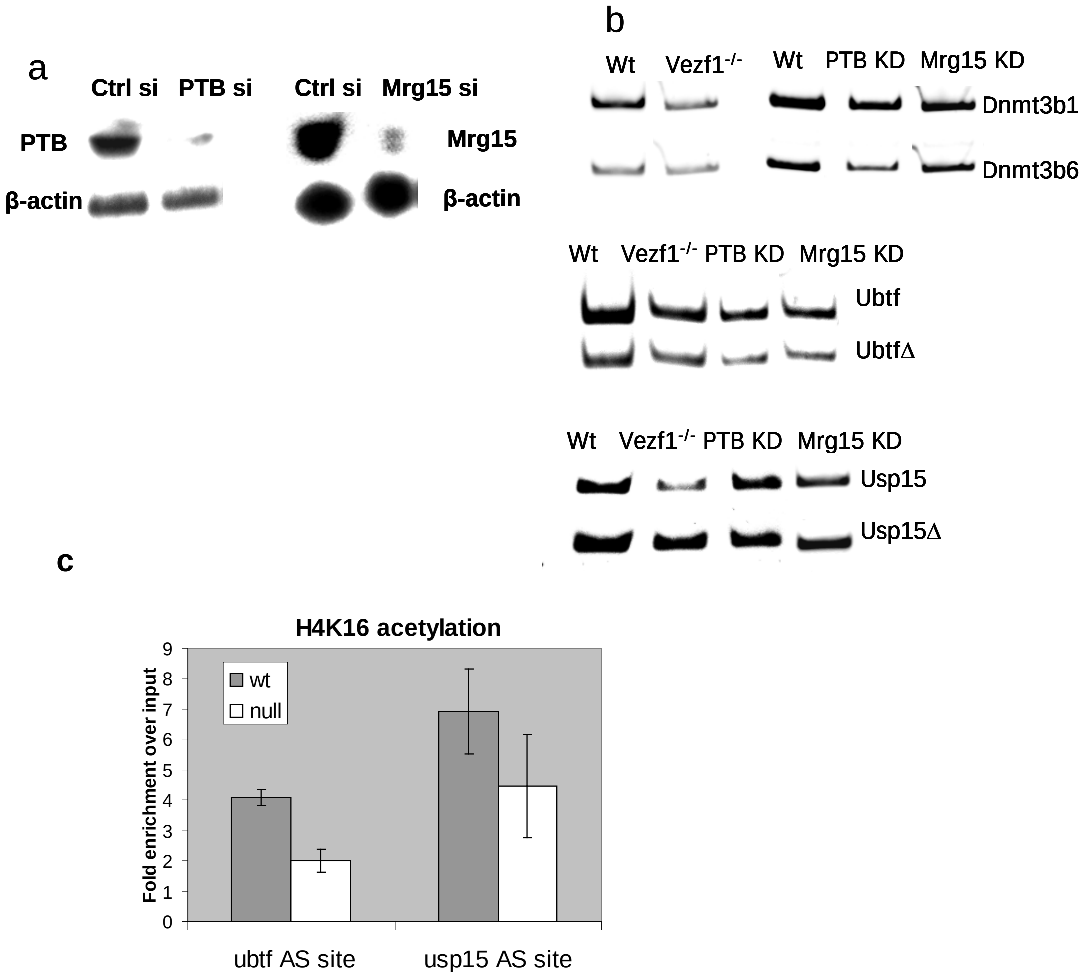
Gowher_S10



List of some Vezf1 interacting proteins

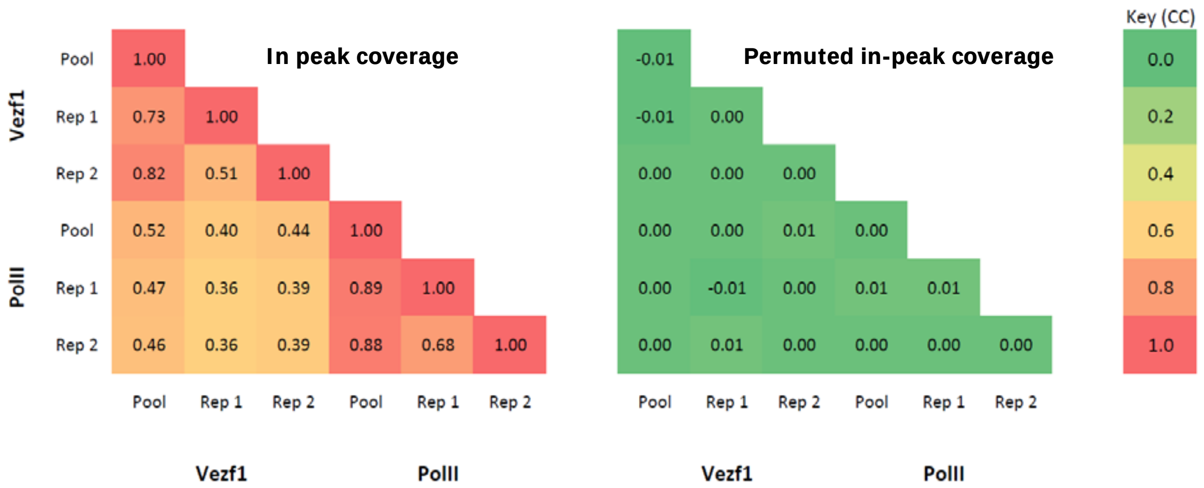
CBX3
CHD4
Mrg15
MRGBP
TIF1B
TOP1
TOP2A

Gowher_S11


a
Ctrl si
PTB si
Ctrl si
Mrg15 si
PTB
Mrg15
β-actin
β-actin
b
Wt
Vezf1-/-
Wt
PTB KD
Mrg15 KD
Dnmt3b1
Dnmt3b6
Wt
Vezf1-/-
PTB KD
Mrg15 KD
Ubtf
UbtfΔ
Wt
Vezf1-/-
PTB KD
Mrg15 KD
Usp15
Usp15Δ
c
H4K16 acetylation
Fold enrichment over input
0
1
2
3
4
5
6
7
8
9
wt
null
ubtf AS site
usp15 AS site

Gowher_S12

Mean per chromosome Spearman CCs for all HELA Vezf1/ Ser2-Pol II pools and replicates



In peak coverage

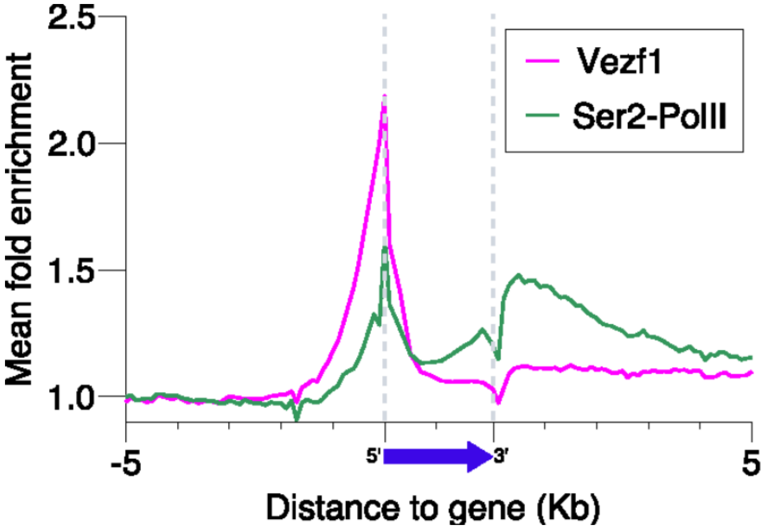
		Vezf1 Pool	Vezf1 Rep 1	Vezf1 Rep 2	PolII Pool	PolII Rep 1	PolII Rep 2
Vezf1	Pool	1.00					
	Rep 1	0.73	1.00				
	Rep 2	0.82	0.51	1.00			
PolII	Pool	0.52	0.40	0.44	1.00		
	Rep 1	0.47	0.36	0.39	0.89	1.00	
	Rep 2	0.46	0.36	0.39	0.88	0.68	1.00

Permuted in-peak coverage

		Vezf1 Pool	Vezf1 Rep 1	Vezf1 Rep 2	PolII Pool	PolII Rep 1	PolII Rep 2
Vezf1	Pool	-0.01					
	Rep 1	-0.01	0.00				
	Rep 2	0.00	0.00	0.00			
PolII	Pool	0.00	0.00	0.01	0.00		
	Rep 1	0.00	-0.01	0.00	0.01	0.01	
	Rep 2	0.00	0.01	0.00	0.00	0.00	0.00

Key (CC)
0.0
0.2
0.4
0.6
0.8
1.0

Gowher_S13



Supplementary Figure legends

S1. Left panel: Western blot showing the expression of Vezf1 and Flag-HA Vezf1 in the untransfected (W) and transfected (T) Hela S3 cells. The same blot was serially probed with anti Vezf1 antibody and anti Flag antibody. Right panel: Gel shows the western blot of the total cell extract from HelaS3 treated with either Control siRNA or Vezf1siRNA for 42 hrs.

S2. a) The genome-wide localization of Vezf1 binding sites. Coding regions are defined as the DNA between all annotated start and stop codons. Transcription start sites (TSSs) are defined as start codons ± 500bp. Intergenic and coding regions which overlap TSS regions are classified as TSS and thus excluded from their respective datasets. b) Vezf1/Ser2-Pol II co-binding sites are preferentially found at CpG island containing promoters, defined as those promoters with an annotated CpG island (from UCSC CpG islands table) within 500nt of the transcription start site.

S3. In-gene Vezf1 coverage is correlated with in-gene Ser2-Pol II coverage. The background subtracted Vezf1 and Ser2-Pol II tag counts were compared in all in-gene non overlapping 50-nt windows. The distributions were highly correlated (Spearman's R = 0.81). Randomization of Ser2-Pol II values abolished all correlation (Spearman's R = 0.01) though this is an excessively punitive comparison which does not take into account local biases in the genomic sequence.

S4. a) Schematic of sites selected for study, showing distribution before knockdown of bound Vezf1 and Ser2P-Pol II. ChIP-Seq tag density is plotted in non-overlapping 50 nt windows. Each ChIP-Seq tag was extended 150nt in the 3' direction from the tag start. Coverage is only shown for statistically enriched regions (see Methods). b) HeLa S3 cells were treated with siRNA to knock down the expression of endogenous Vezf1. Crosslinked chromatin was immunoprecipitated using anti Ser2Phos Pol II antibody followed by quantitative

PCR using Taqman probes. In the bar diagram, control site_PV indicates the region upstream of where Pol II peaks overlap with Vezf1 binding sites.

S5. a) Role of Vezf1 in alternate promoter usage. Left panels are schematic of sites selected for study, showing alternative transcripts and start sites of the genes. Red square shows the position of Vezf1/Ser2Pos-Pol II peaks at the promoters and the blue arrows point to the alternate promoter site and primers for PCR were designed in these regions. Vezf1 was knocked down and Ser2Phos-Pol II occupancy was assayed by ChIP/QPCR. b) CTCF sites were obtained from (1) or from the UW Histone modifications track of the UCSC genome browser. Only Vezf1 sites outside of TSS regions were considered. Each feature (Vezf1 / CTCF sites) was expanded to 1 Kb around the centre point then overlaps between features were counted.

S6. a) . Chromatin immunoprecipitation followed by quantitative PCR using Taqman probes. Crosslinked chromatin from Wt and Vezf1$^{-/-}$ (Null) ES cells was serially immunoprecipitated first with this anti Pol II antibody and then with anti Vezf1 antibody. b) S1 nuclease protection assay. A template containing both, a downstream synthetic polyadenylation signal and, farther 3', a binding site for (in this case) Vezf1, was transcribed in Hela nuclear extract. The probes used could hybridize to the transcript beyond the Vezf1 sites, allowing detection of pausing. When Vezf1 sites were introduced the major transcript terminated at those sites. A shorter transcript was also seen, since pausing leads to cleavage coupled to activation of polyadenylation. This effect was observed with the Vezf1 sites in either orientation; the band intensity indicates a modest preference for C in the top strand. Lanes V1, rV1, V2 and rV2 show a single strong band representing Pol II pausing at the Vezf1 binding site. The positive control (lane M) confirmed the earlier published *in vitro* experiments[13] in which tandem binding sites for the protein factor MAZ were found to cause RNA Pol II pausing. No pausing is observed in the control lacking MAZ or Vezf1 sites (lane MM).

S7. a) Ser2P-Pol II signal is amplified at some sites in Vezf1$^{-/-}$ ES cells. Ser2-Pol II ChIP-Seq tag density is plotted in 1Kb windows with a step size of 100nt. Each

ChIP-Seq tag was extended 150nt in the 3' direction. Blue lines represent Ser2-Pol II coverage in WT cells, red lines in Vezf1$^{-/-}$ cells. Black arrows indicate the sites chosen for subsequent quantitative measurement of Vezf1 and Pol II binding by ChIP-QPCR. These sites have significantly higher Ser2-Pol II ChIP-Seq coverage in Vezf1$^{-/-}$ cells than in WT. The red arrow on the slc1a4 gene points to the promoter and the green arrow on ubtf points to the alternatively spliced exon. b) One step RT-PCR was performed with total RNA from Wt and Vezf1$^{-/-}$ cells; β-actin RNA was the control for the amount of RNA template used as input.

S8. Illustrations of Usp15, Ubtf and Dnmt3b genes showing cassette exons. Gels show two alternatively spliced isofoms by RT PCR analysis using RNA from Wt or Vezf1$^{-/-}$ ES cells at the highest concentration of 1ng followed by two fold dilutions. The primers flank the alternatively spliced exons.

S9. Incidence of Vezf1/Ser2-Pol II co-binding sites downstream of cassette exons. The 3' regions of cassette exons and Vezf1/Ser2-Pol II co-binding sites within ±1Kb of TSS of were not considered.

S10. SDS PAGE showing immunoprecipitated nuclear extracts from the un-transfected (lane1) and transfected (lane 2) mouse ES cells using anti-V5 antibody. The small right panel is the western blot with anti-V5 antibody. The lanes 1 and 2 were excised from the gel and the proteins were identified by Mass spectrometry (MS). The right panel is the list of some chromatin binding proteins that were identified by MS.

S11. a) Western blot using anti Mrg15, PTB and β-actin antibodies after siRNA knockdown of indicated proteins in mouse ES cells. b) RTPCR analysis showing the affect of siRNA knockdown of indicated factors on the transcription/alternative splicing of dnmt3b, ubtf and usp15 genes. c) ChIP showing the H3k16 acteylation levels at the alternative spliced exons of the ubtf and usp15 genes in Wt and Vezf1-/- ES cells.

S12. Vezf1 and Ser2P-Pol II ChIP-Seq profiles are highly reproducible. The ChIP-Seq tag coverage in non-overlapping 1Kb bins was calculated. The heat map displays the mean per chromosome Spearman's R for each pair of ChIP-Seq samples.

S13**.** Vezf1 and Ser2P-Pol II ChIP-Seq coverage across protein coding regions. Data are plotted in non-overlapping 100nt windows. For each sample, coverage was normalized by the sample-specific background (mean signal in 50-100 Kb region upstream of TSS) before calculating fold enrichment over input DNA signal (also normalized). In-gene coverage was calculated in 10 evenly spaced bins for each gene and genes were aligned and oriented at the TSS (genes illustrated as blue arrow). Thick, dashed grey lines represent the start (TSS) and stop (3') codons respectively.

Supplementary methods

Sequencing:

Sequencing was performed on an Illumina GAII. Quality filtered reads were aligned to version hg18 of the human genome using the Illumina GA Pipeline version 1.6. Samples were pooled and 42 million tags were randomly selected from each to create equal sized datasets.

In-gene coverage

Given that the majority of Vezf1 binding sites were located in and around coding regions, we focused on their relationship with Ser2P-Pol II in the neighborhood of genes. We calculated the average tag density across all genes for Ser2P-Pol II and Vezf1 .We calculated the average tag density across all genes for Ser2P-Pol II and Vezf1 (Fig. S.13). As reported earlier (2) we see a gradual increase in the accumulation of Ser2P-Pol II signal towards the 3' end of the genes. This characteristic feature is absent in the average Vezf1 binding profile which shows a sharp peak at the TSS The entire set of hg18 transcripts was obtained from the hg18 RefSeq annotation table in UCSC. A non-redundant gene set was generated using the most extreme 5' and 3' ends of all alternative forms of each gene. The background subtracted coverage of Vezf1 and Ser2P-Pol II was calculated across each gene and the resultant value was normalized by gene length. The Spearman correlation coefficient was calculated between the normalized strength vectors for Vezf1 and Ser2P-Pol II. To assess the significance of the correlation, the strength values for Ser2P-Pol II were randomly permuted.

Putative transcription start sites were defined as the 5' ends of RefSeq transcripts ± 500bp. We observed peaks in both Ser2P-Pol II and Vezf1 coverage at the 5' end of genes. Such a striking pattern of Ser2P-Pol II coverage at the 5' end of genes was unsurprising given recent data(3). Ser2P-Pol II coverage also increased at the 3' end of genes, as expected for the elongating form of this enzyme.

Repeat associations:

Repeats in the human genome were obtained from the repeatmasker database. Randomized peaks were generated by re-distributing peak centers uniformly in non-gap regions of the genome. The overlap of each repeat class in peaks was compared to the expected overlap, as derived from 1,000 sets of randomized peaks. The significance of enrichment was calculated using a one sided binomial test for each repeat class. Bonferroni correction was applied to account for multiple testing.

Enrichment in alternative transcription events:

A list of annotated alternative transcription events was obtained from the alt Events table in UCSC for the human genome version hg18. The expected overlap was generated by permuting the genomic features in question. For cassette exon 3' 5Kb regions, the 3' 5Kb region was identified for all non-cassette exons in the genome. Those regions overlapping TSS were not considered. 10,000 subsets were generated, each with a matching number of randomly selected exon 3' 5Kb regions. The mean overlap with these randomly attributed cassette exon 3' 5Kb regions was used as the expected overlap percentage. For alternative promoter associations at the TSS, the same permutation-based approach was used, albeit for those TSS which did not contain a Vezf1/ Ser2-Pol II peak.

S1 Nuclease Assay

HeLa S3 nuclear extracts were purified as described (4) . MLPIII DNA templates were a kind gift from Dr Nick Proudfoot (4). The MAZ sites in pMLPIII were replaced by *dnmt3b* Vezf1 binding sites in forward and reverse orientation. In vitro transcription using 10μl of HeLa S3 nuclear extract and 200ng of linearized template at 30C for 1 hr and the transcripts were essentially purified as described (5). Specific probes were PCR amplified from different pMLPIII constructs, purified and cleaved with Sap I, end-filled using radiolabelled dGTP thus allowing only the template strand to be radiolabelled. The probe was mixed with transcribed RNA in 80% formamide buffer and annealed at 52°C overnight. The mixture was treated with SI nuclease to cleave single stranded DNA or RNA; DNA/RNA hybrids are protected. Ethanol precipitated RNA/DNA hybrid was resolved on 6% urea/acrylamide gels.

Chromatin Immunoprecipitation and antibodies

Crosslinking for all the cell types used was carried out for 3 mins in 1% formaldehyde. The protocol for chromatin preparation was followed as described (6). For Flag-HA-Vezf1 ChIP, Flag M2 Agarose (Sigma) was used for immunoprecipitation. For ChIP-Seq, crosslinked chromatin from 10^7 cells and 60μg of antibody was used for each immunoprecipitation. Purified DNA was subjected to end-repair and adapter ligation using a sample preparation kit from Illumina according to the manufacturer's protocol. Antibodies used for Ser2 Phosphorylated RNA Pol II; (ab5095), Mrg15 (ab-84520-50), and for histone modifications were purchased from ABCAM, and PTB antibody (32-4800) from Invitrogen.

RT-PCR and Quantitative PCR analysis

RNA was purified using RNeasy Mini Kit (Qiagen), treated with DNase (Roche) and was reverse transcribed using the 1 Step RT-PCR kit (Invitrogen). The primers for the semi-quantitative PCR were chosen to target the exon-intron boundaries such that it allows the amplification of alternatively spliced isoforms. All the RT-PCR products were in the size range of 200-500 bp. For the gene expression analysis, the quantitative RT-PCR procedures used TaqMan® One-Step RT-PCR Master Mix and TaqMan® Gene Expression Assays in ABI PRISM 7900 Sequence Detection System (Applied Biosystems). Quantitative PCR for ChIP studies was carried out on the same instrument with primer-probe sets designed using the software Primer Express. The analysis was done as described (7).

SiRNA Knockdown

Downregulation of PTB and Mrg15 in mouse ES cells was performed using ON-TARGET plus SMARTpool siRNA oligonucleotides. $2x10^6$ ES cells in 100μls of

Amaxa Mouse ES cell Nucleofactor solution (Lonza) were transfected with 5μM PTB or Mrg15 siRNA according to the manufacturer's protocol and were harvested after 48 hrs. For Vezf1 knockdown, $1x10^6$ HeLa S3 cells in 100μls of Amaxa Nucleofactor Kit L solution were transfected with 3μM of Vezf1 ON-TARGET plus SMARTpool siRNA oligonucleotides or control siRNA and harvested after 48 hrs. For chromatin preparation, 10^7, HeLa S3 cells were directly crosslinked by treatment with 1% formaldehyde followed by the ChIP protocol as described above.

Nuclear Run-on assay

Nuclei were isolated and nuclear run-on reactions were performed as described (8). RNA was purified by TRIzol according to the manufacturers' instructions. 80-120 bp amplicons at the 5' end, in the gene body and 3' end of *dnmt3b* gene were used as probes. 1μg of DNA probe was denatured in 100μl of 0.5M NaOH and slot-blotted on GeneScreen Plus Nylon membrane. Filters were neutralized with 100mM Tris-HCl, pH8, UV crosslinked, prehybridized and hybridized in 50% formamide, 5XSSC, 5X Denhardt's solution,1%SDS and 100μg/ml boiled herring sperm DNA.

Immunoprecipitation (IP) and Mass spectrometry.

Vezf1 was cloned in pCDNA Topo V-5 His and expression of the protein was verified by WB using anti V-5 ab. Total nuclear extracts from the transfected and untransfected ES cells was prepared by using a nuclear extract /Co IP kit from Active Motif. V5-Vezf1 protein complex was co-immunoprecipitated from nuclear extracts on V-5 beads (sigma) using buffers and instructions from Active Motif. Control IP was performed with nuclear extracts from Wt cells. The immunoprecipitated proteins were separated by SDS gel electrophoresis. Both the IP and control gel lanes were cut into 4 pieces each and used for peptide analysis by mass spectrometry at Taplin Biological Mass Spectrometry Facility (http://gygi.med.harvard.edu/). The list of proteins from the control lane was subtracted from the list of interacting proteins identified from the IP lane.

1. Schmidt D*, et al.* (A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* 20(5):578-588.

2. Rahl PB*, et al.* (c-Myc regulates transcriptional pause release. *Cell* 141(3):432-445.

3. Gilchrist DA*, et al.* (Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* 143(4):540-551.

4. Ashfield R*, et al.* (1994) MAZ-dependent termination between closely spaced human complement genes. *Embo J* 13(23):5656-5667.

5. Yonaha M & Proudfoot NJ (1999) Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* 3(5):593-600.

6. Gowher H, Stuhlmann H, & Felsenfeld G (2008) Vezf1 regulates genomic DNA methylation through its effects on expression of DNA methyltransferase Dnmt3b. *Genes Dev* 22(15):2075-2084.

7. Recillas-Targa F*, et al.* (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* 99(10):6883-6888.

8. Cuello P, Boyd DC, Dye MJ, Proudfoot NJ, & Murphy S (1999) Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *Embo J* 18(10):2867-2877.

Gowher_S.table1

.

Sample statistics for ChIP-Seq experiments

Antibody	# lanes	# tags	# bound sites
VEZF1	4	42 Mn	34,204
Ser2-PolII	3	42 Mn	114,443
Input	3	42 Mn	n/a

Gowher_S.table2

		Enrichment			
Repeat	**Type**	**(fold)**	**OBS**	**EXP**	**-log10(P)**
Low complexity	Family	1.95	4106	1390	< 300
Simple repeats	Family	1.34	3929	1681	< 300
GC-rich	Low complexity	52.73	3159	59	< 300
G-rich	Low complexity	16.87	430	24	< 300
C-rich	Low complexity	15.64	401	24	< 300
(CGG)n	Simple repeats	60.48	383	6	< 300
(CCG)n	Simple repeats	56.99	370	6	< 300
(CGGGG)n	Simple repeats	47.64	250	5	< 300
(CCCCG)n	Simple repeats	50.20	234	5	299.05
(TG)n	Simple repeats	1.98	666	223	123.34
(CA)n	Simple repeats	1.81	631	225	105.96
(CGGG)n	Simple repeats	43.59	70	2	83.85
(CCGGG)n	Simple repeats	60.33	46	1	60.60
(CCCCCG)n	Simple repeats	44.71	48	1	57.32
(CTG)n	Simple repeats	15.59	70	4	54.92
(CCCG)n	Simple repeats	30.41	49	2	50.78
(GGA)n	Simple repeats	11.68	71	6	48.02
(CGGGGG)n	Simple repeats	35.61	41	1	44.77
(CCCGG)n	Simple repeats	51.24	35	1	43.18
(CGTG)n	Simple repeats	25.01	45	2	42.90
(TCC)n	Simple repeats	10.07	68	6	42.23
(CAG)n	Simple repeats	12.84	53	4	37.15
(CACG)n	Simple repeats	22.68	36	2	32.46
MER52A	LTR	3.31	109	25	30.97
tRNA	tRNA	6.58	65	9	30.71
(CGCGG)n	Simple repeats	42.86	25	1	28.32
(CCGCG)n	Simple repeats	46.06	24	1	27.81
(CG)n	Simple repeats	32.80	24	1	24.45
(CTGGGG)n	Simple repeats	9.98	37	3	21.83
MER52D	LTR	5.44	51	8	20.52
(CCCCAG)n	Simple repeats	9.97	34	3	19.86
(TGG)n	Simple repeats	5.43	42	7	16.48
(GTGTG)n	Simple repeats	20.51	20	1	16.19
(CGAG)n	Simple repeats	53.17	13	0	14.73
(CCGAG)n	Simple repeats	90.67	11	0	14.56
(GGCTG)n	Simple repeats	13.18	19	1	12.01
(CACCAT)n	Simple repeats	6.56	27	4	11.39
(CCCTG)n	Simple repeats	8.95	22	2	11.18
(CAGC)n	Simple repeats	18.48	15	1	10.92
(CACAC)n	Simple repeats	17.99	15	1	10.76
(CAGGG)n	Simple repeats	8.48	20	2	9.56
(CTCG)n	Simple repeats	51.94	9	0	9.34
(ATGGTG)n	Simple repeats	5.76	25	4	9.31
(CGAGG)n	Simple repeats	49.00	9	0	9.12
CT-rich	Low complexity	0.72	204	118	9.07
(CCA)n	Simple repeats	3.92	30	6	8.21
(TCCG)n	Simple repeats	26.27	9	0	6.81
(GAATG)n	Simple repeats	4.18	23	4	6.15
(CGGA)n	Simple repeats	28.63	8	0	6.04
(CGGAG)n	Simple repeats	59.00	6	0	5.68
(TCCCC)n	Simple repeats	4.21	21	4	5.45
(CATG)n	Simple repeats	6.43	15	2	5.14
GA-rich	Low complexity	0.57	187	119	5.03
(CCCCCT)n	Simple repeats	9.68	11	1	4.65
(TCG)n	Simple repeats	49.00	5	0	3.90
(CTCGG)n	Simple repeats	44.45	5	0	3.70
(TCCC)n	Simple repeats	2.31	24	7	2.92
(CAGCC)n	Simple repeats	7.00	10	1	2.87
(CCTCG)n	Simple repeats	28.41	5	0	2.77
(GGGGA)n	Simple repeats	3.26	17	4	2.75
(C)n	Simple repeats	12.73	7	1	2.73
(CACCC)n	Simple repeats	4.51	13	2	2.68
U1	snRNA	7.33	9	1	2.46
(CACGC)n	Simple repeats	35.36	4	0	2.04
(CCTG)n	Simple repeats	6.09	9	1	1.90
(CATTC)n	Simple repeats	4.45	11	2	1.83
(GCCTG)n	Simple repeats	15.13	5	0	1.52
(AGGGGG)n	Simple repeats	6.34	8	1	1.51
(CGGTG)n	Simple repeats	59.00	3	0	1.48
(CGA)n	Simple repeats	59.00	3	0	1.48
(AGGTG)n	Simple repeats	14.63	5	0	1.45
(GGGA)n	Simple repeats	2.01	20	7	1.45

Vezf1 binding sites are enriched for many GC-rich repeat elements. Only over-represented repeats are shown.