Supporting Information

Portal et al. 10.1073/pnas.1317608110

SI Materials and Methods

ChIP Assays and Sequencing. B95.8-derived lymphoblastoid cell line (LCL) cells were cross-linked in 1% (vol/vol) formaldehyde, lysed in Lysis Buffer [50 mM Tris·HCl, pH 8.1, 10 mM EDTA, 1% SDS, and protease inhibitors (Roche)] and sonicated for 20 × 20-s cycles on ice. Lysate was diluted 10-fold with Dilution Buffer (16.7 mM Tris·HCl, pH 8.1, 1.2 mM EDTA, 167 mM NaCl, 1.1% Triton X-100, 0.01% SDS, and protease inhibitors) and monoclonal Epstein–Barr virus nuclear antigen (EBNA) leader protein (EBNALP) antibodies (JF186) (1) were added to immune precipitate EBNALP–DNA complexes. EBNALP–DNA complexes were captured with the addition of protein A-agarose/salmon sperm DNA and then washed one time for 15 min at 4 °C with rotation, with Low Salt Wash Buffer (20 mM Tris·HCl, pH 8.1, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 1% SDS), one time with High-Salt Wash Buffer (20 mM Tris·HCl, pH 8.1, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 1% SDS), one time with LiCl Wash Buffer (10 mM Tris·HCl, pH 8.1, 1 mM EDTA, 0.25 M LiCl, 1% Nonidet P-40, 1% deoxycholate acid), and two times with TE buffer. EBNALP–DNA complexes were eluted from the protein A-beads by rotating at room temperature two times for 15 min in 250 μL of Elution Buffer (100 mM $NaHCO_3$, 1% SDS) to a final volume of 500 μL. DNA–protein cross-link was reversed by the addition of 20 μL of 5 M NaCl and incubation at 65 °C for 4 h. DNA was purified using a PCR purification column (Qiagen). ChIP-sequencing (ChIP-seq) DNA libraries were prepared and sequenced, using Genome Analyzer II (Illumina).

ChIP-Seq Peak Calling, Motif Searching, Functional Enrichment, and Overlap Detection. ChIP-seq reads were mapped to hg19 using Bowtie (version 0.12.9) allowing 1-missmatch in the alignment seed-region (parameters –m1). HOMER (2) was used to call peaks using default criteria of false discovery rate < 0.001, a fourfold enrichment over input, and other default filters (3). HOMER was used to conduct de novo motif discovery, identify enrichment of known motifs (200- and 500-bp windows showed similar results), and annotate peak calls using the Gene Ontology (4) and genomic location. The MergePeak function in HOMER was used to identify the overlap between binding sites. In order to define the sites as "overlapping," the two transcription factor binding sites' peak centers must be at a distance equal or less than 200 bp.

K-Means Clustering and Expression Analysis. Transcription factor (TF) binding sites were annotated to genes using HOMER and gene-occupancy clusters were obtained by running the K-means clustering module (5) in the GenePattern (6) Public Server specifying 10 clusters. Cluster gene expression was tested using the Wilcoxon rank sum test, corrected for multiple testing using the Benjamini–Yekutieli procedure (7), and visualized using the R programming language.

Anchor plots. All anchor plots were generated using the function "annotatePeaks" (8) in the HOMER analysis suite. In addition to the default normalization to 10 million mapped tags, the script further normalizes read depth to the units of "per base pair per peak." This is analogous to an RPKM correction and can be interpreted as units of 10 million per base pair per peak. The transcription factor ChIP-seq coverage (depth per peak per base pair) has been normalized by subtracting the Input coverage.

Datasets. ChIP-seq data for factors other than EBNALP and EBNA2 were downloaded from the ENCODE (9) at the University of California, Santa Cruz, genome database or the National Center for Biotechnology Information Short Read Archive (10).

Notes. The transcription factors and histone marks ChIP-seq data are downloaded from ENCODE, listed below:

wgEncodeHaibTfbsGm12878BatfRawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878Corestsc30189IggmusRawDataRep2.fastq.tgz

wgEncodeSydhTfbsGm12878Ctcfsc15914c20StdRawDataRep1.fastq.gz

wgEncodeSydhTfbsGm12878Ebf1sc137065StdRawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878Ets1Pcr1xRawDataRep2.fastq.gz

wgEncodeBroadHistoneGm12878H2azStdRawDataRep2.fastq.gz

wgEncodeBroadHistoneGm12878H3k27acStdRawDataRep2.fastq.gz

wgEncodeBroadHistoneGm12878H3k4me1StdRawDataRep1.fastq.gz

wgEncodeBroadHistoneGm12878H3k4me3StdRawDataRep2.fastq.gz

wgEncodeBroadHistoneGm12878H3k9acStdRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878Irf4sc6059RawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878NfkbTnfaIggrabRawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878NrsfPcr2xRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878Pax5c20RawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878Pol2Pcr2xRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878Pu1Pcr1xRawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878Rad21V0416101RawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878Smc3ab9263IggmusRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878Sp1Pcr1xRawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878Tblr1ab24550IggmusRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878Yy1sc281Pcr1xRawDataRep1.fastq.gz

wgEncodeSydhTfbsGm12878Znf143166181apStdRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr1xRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr1xRawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr1xRawDataRep3.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr1xRawDataRep4.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr1xRawDataRep5.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr2xRawDataRep1.fastq.gz

wgEncodeHaibTfbsGm12878RxlchPcr2xRawDataRep2.fastq.gz

wgEncodeHaibTfbsGm12878RxlchV0416101RawDataRep1.fastq.gz

wgEncodeSydhTfbsGm12878InputTnfaIggrabRawDataRep1.fastq.gz

wgEncodeSydhTfbsGm12878InputTnfaIggrabRawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878IggmusRawDataRep1.fastq.gz

wgEncodeSydhTfbsGm12878IggmusRawDataRep2.fastq.gz

wgEncodeSydhTfbsGm12878InputRawData.fastq.gz

wgEncodeBroadHistoneGm12878ControlStdRawDataRep1.fastq.gz

wgEncodeBroadHistoneGm12878ControlStdRawDataRep2.fastq.gz.

EBNALP, EBNA2, and RBPJ datasets:
The datasets have been submitted to GEO database and can be accessed at the following:

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29498

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49338.

1. 1Finke J, et al. (1987) Monoclonal and polyclonal antibodies against Epstein-Barr virus nuclear antigen 5 (EBNA-5) detect multiple protein species in Burkitt's lymphoma and lymphoblastoid cell lines. *J Virol* 61(12):3870–3878.
2. Benner C, et al. (2010) Homer Software and Data Download. Available at http://biowhat.ucsd.edu/homer/index.html. Accessed January 23, 2013.
3. Benner C, et al. (2010) Homer Software and Data Download. Available at http://biowhat.ucsd.edu/homer/ngs/peaks.html. Accessed January 23, 2013.
4. Ashburner M, et al. (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
5. MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Univ of California, Berkeley), Vol 1, pp 281–297.
6. Reich M, et al. (2006) GenePattern 2.0. *Nat Genet* 38(5):500–501.
7. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188.
8. Benner C, et al. (2010) Homer Software and Data Download. Available at http://biowhat.ucsd.edu/homer/ngs/quantification.html. Accessed January 23, 2013.
9. Benner C, et al. (2010) Homer Software and Data Download. Available at http://genome.ucsc.edu/ENCODE/. Accessed January 23, 2013.
10. Benner C, et al. (2010) Homer Software and Data Download. Available at www.ncbi.nlm.nih.gov/Traces/sra/. Accessed January 23, 2013.

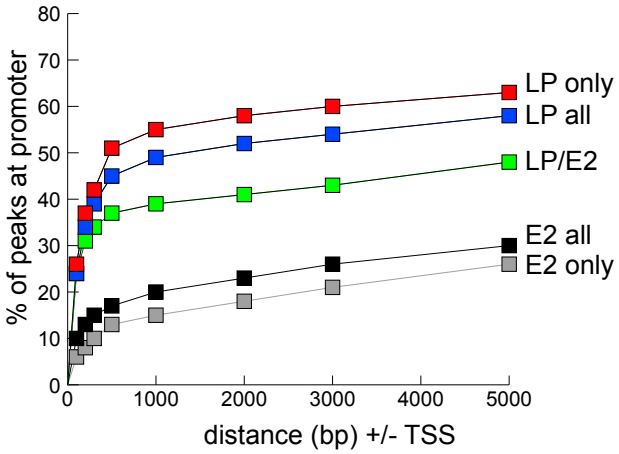

Fig. S1. Percentage of LP or E2 sites within increasingly larger windows ±1 kb of transcriptional start sites (TSSs) indicated on the *x* axis. As windows increase to ±3 kb of TSS, LP sites reach a plateau of ∼60% inclusion and E2 at ∼28%. LP had a stronger effect on E2 than E2 on LP as seen in the LP/E2 curve.

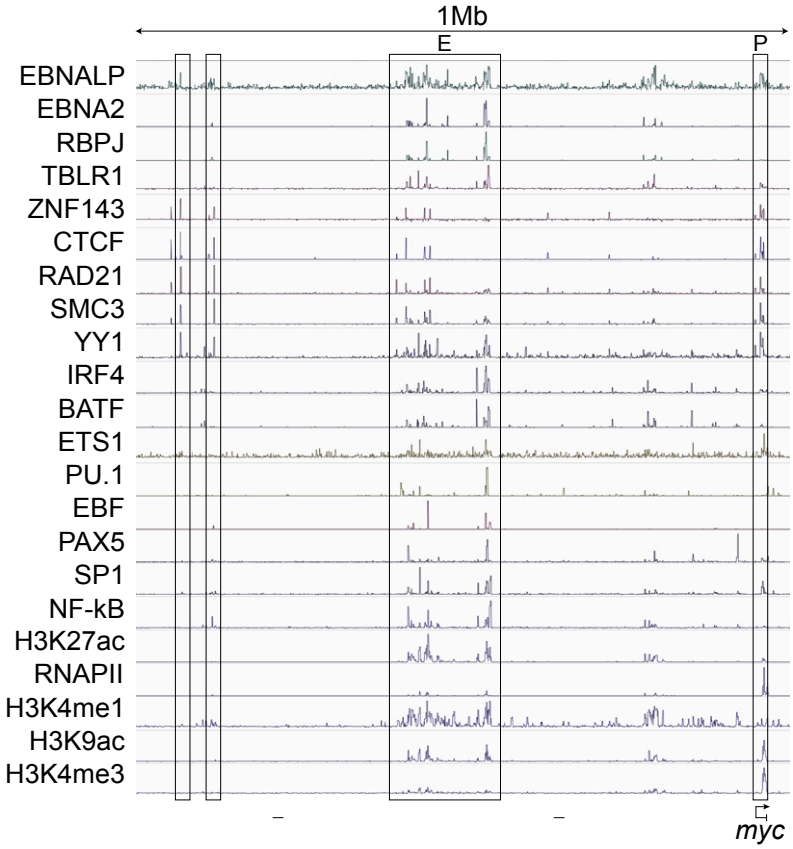

Fig. S2. ChIP-seq tracks for TFs and epigenetic modifications at the 1-Mb *myc* locus. Schematic of the ChIP-seq peaks at 1 Mb upstream of the *myc* locus. LP sites are at the promoter (P) and enhancers (E), overlapping with looping factors CTCF, RAD21, and SMC3. Two other LP peaks overlap with CTCF, RAD21, SMC3, and YY1 upstream of the *myc* enhancer (E). The *myc* ORF is indicated at *Bottom*, and the TSS is indicated by an arrow. The two horizontal lines upstream of myc indicate two ORFs (*pou5f1b* and *pcat1)*, not expressed in LCLs.

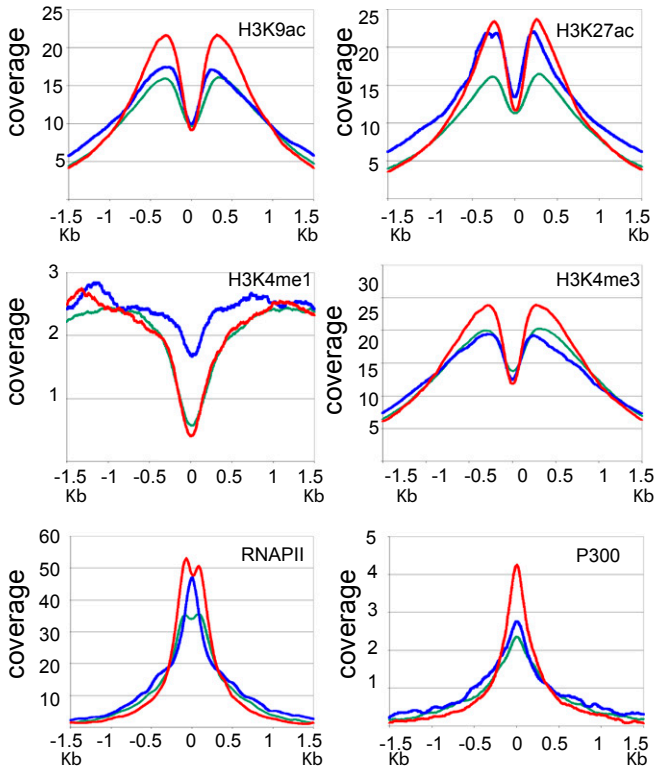

Fig. S3. Epigenetic and basal cell TF anchor plots at LP, E2, and LP/E2 promoter sites as opposed to all sites. Anchor plots show coverage of the indicated TFs or epigenetic marks (±1.5 kb) of LP (green), E2 (blue), and LP/E2 (red) sites. Coverages are similar, only when promoter peaks are considered for LP, LP/E2, and E2, consistent with the box plots in Fig. 3*B*.

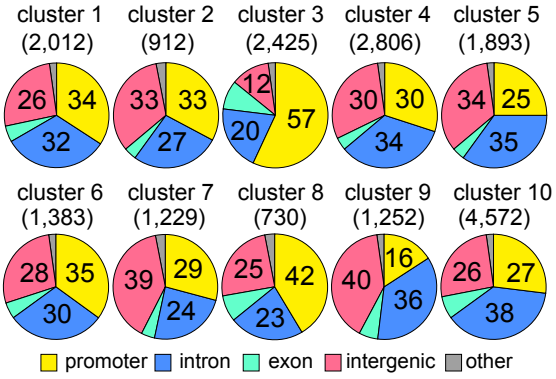

Fig. S4. Genomic distribution of the 10 K-means clusters in Fig. 4*A*. Pie charts showing the genome-wide distribution of the 10 LP clustered sites. All sites have distributions similar to LP (Fig. 1*A*) with the exceptions of cluster 3, which is highly promoter associated and rich in PAX5 (Fig. 4*A*), and cluster 9, which is deficient in promoter association, and rich in DNA looping factors (Fig. 4*A*).

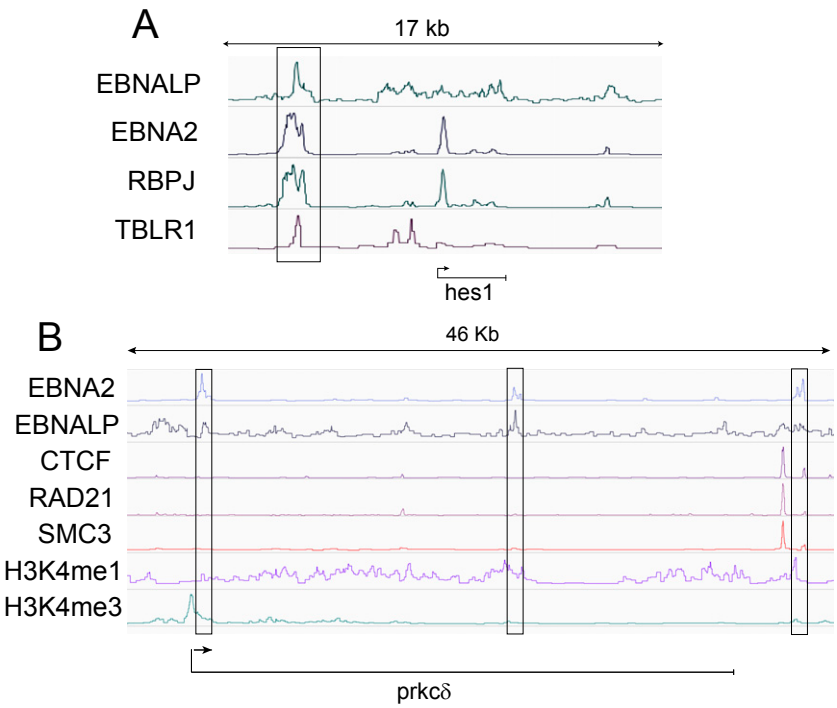

Fig. S5. ChIP-seq tracks at *hes1* (*A*) and *prkcd* (*B*) loci. (*A*) Schematic of the ChIP-seq peaks for LP, E2, RBPJ, and TBLR1 at the ∼17-kb *hes1* locus. The LP site overlaps with E2, RBPJ, and TBLR1 at −5 kb of the *hes1* ORF as previously detected (1). The *hes1* TSS is indicated by an arrow. (*B*) Three LP sites are present at the promoter, coding region, and immediately downstream of the *prkcδ* ORF (indicated at *Bottom*). The TSS is indicated by an arrow.

1. Portal D, et al. (2011) EBV nuclear antigen EBNALP dismisses transcription repressors NCoR and RBPJ from enhancers and EBNA2 increases NCoR-deficient RBPJ DNA binding. *Proc Natl Acad Sci USA* 108(19):7808–7813.

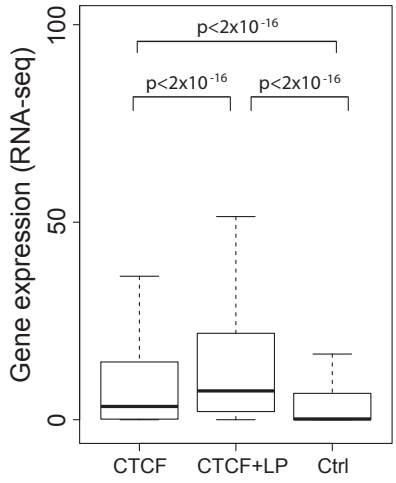

Fig. S6. Box plots comparing RNA-seq gene expression of genes having CTCF and an overlapping LP site (CTCF+LP) at their promoter versus genes having only CTCF (CTCF). Genes with overlapping CTCF and LP sites were significantly more expressed than genes having only CTCF sites ($P < 2 \times 10^{-16}$), indicating that LP localization with CTCF increases expression, by enabling long-distance DNA interaction with enhancers, or by relocalizing CTCF-associated repressors. Both gene sets were significantly more expressed than a randomly selected control gene set (Ctrl) (*P* values are indicated in the figure).

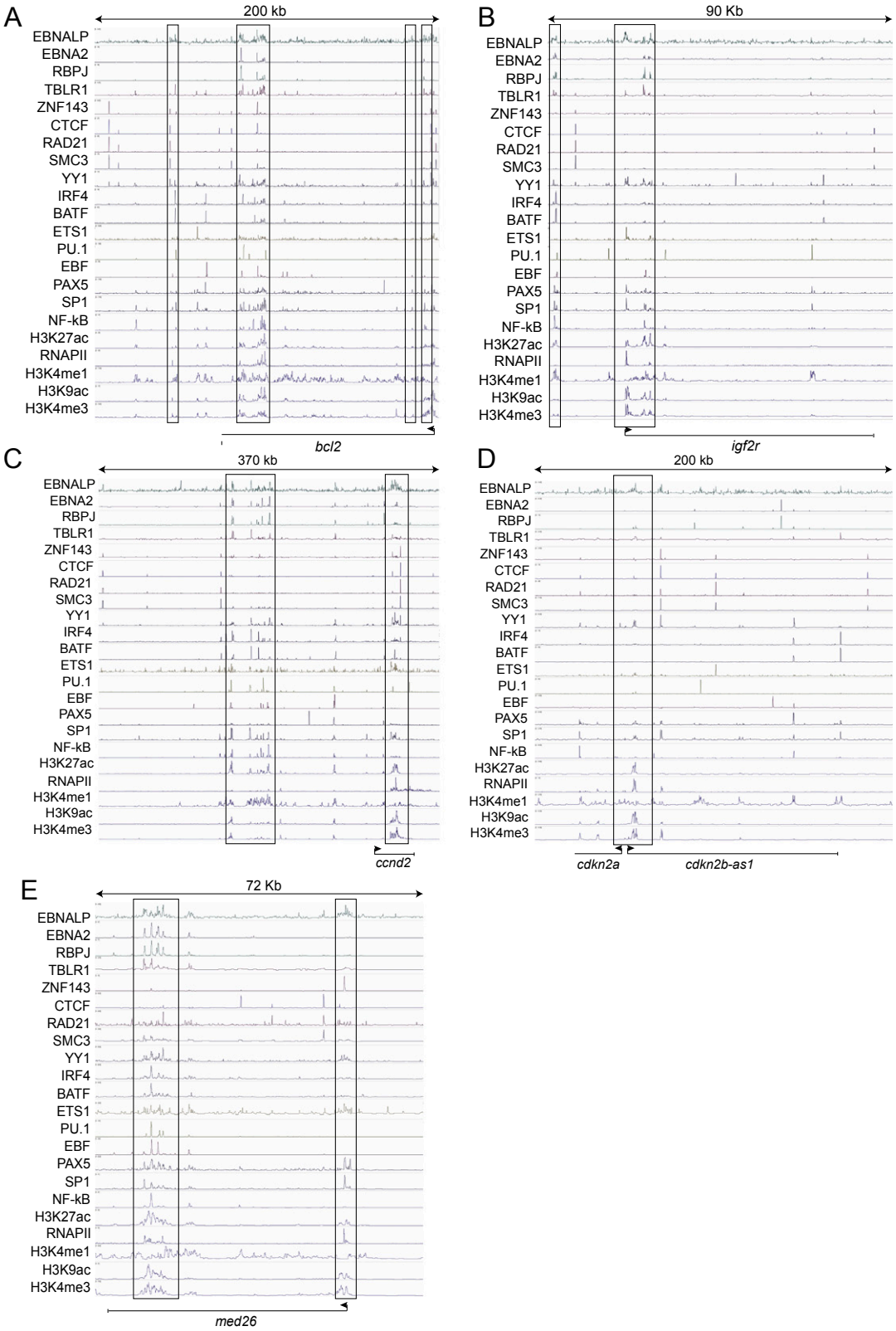

Fig. S7. ChIP-seq tracks for the TFs and epigenetic modifications at the *bcl2*, *igf2r*, *ccnd2*, *cdkn2A*, and *med26* loci. Schematic of the ChIP-seq peaks of the indicated TFs, epigenetic marks, LP, and E2 at biologically relevant loci showing LP's presence and likely role in regulating *bcl2* (*A*), *igf2r* (*B*), *ccnd2* (C), *cdkn2A* (*D*), and *med26* (*E*) expression. ORFs are indicated at the bottom. TSSs are indicated by arrows. The approximate sizes of the loci are indicated in kilobases at *Top*. The boxes indicate LP peaks.