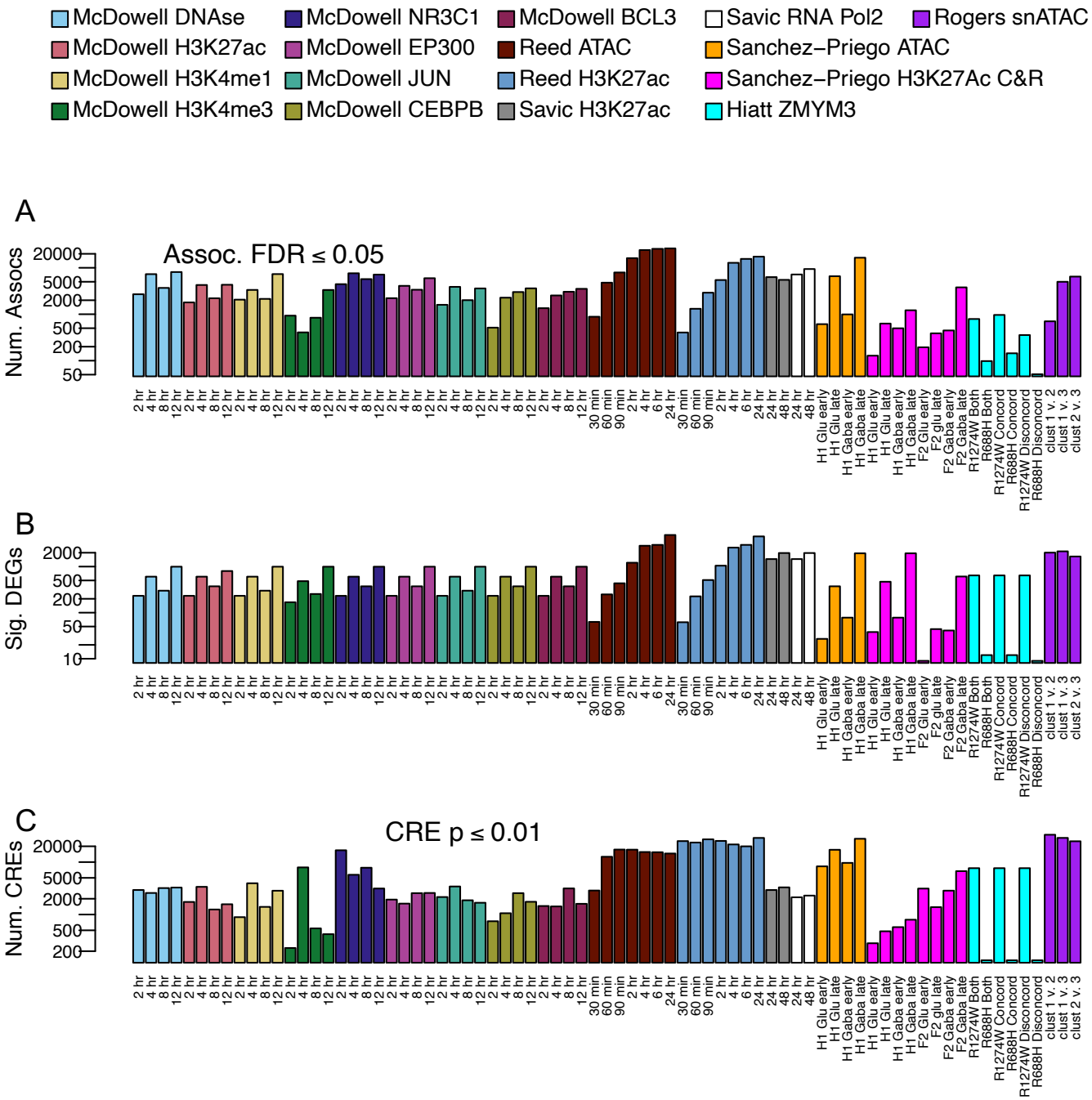
Supplemental Figures and Tables

Supplemental Table 1

Data set nickname	PMID	Data Types	Description
McDowell et al.	30097539	RNA-seq, DNAse hypersensitivity, ChIP-seq of H3K4me3, H3K4me1, H3K27ac, NRC31,EP300, JUN, CEBPB, BCL3	A549 cells were treated with dexamethasone (a NR3C1 agonist) and data was collected at various time points. NR3C1 is the glucocorticoid receptor and plays key roles in immunity and metabolism.
Reed et al.	36323252	RNA-seq, ATAC-seq, H3K27ac ChIP-seq, HiC	THP1 monocytes that had been differentiated into macropahges were stimulated with interferon gamma (IFGN) and lipopolysaccharide (LPS) to convert into a pro-inflammatory state. IFGN and LPS are known potent activators of macrophages. Data was collected at various time points.
Savic et al.	27401066	RNA-seq, RNA Pol2 ChIP-seq, H3K27ac ChIP-seq	HT29 cells were treated with rosiglitazone, a peroxisome proliferator–activated receptor (PPAR) agonist. Natural ligands of PPARs inlcuded fatty acids and Vitamin B3. PPARs play roles in metabloism, differentiation, and tumorigenesis. Data was collected at 24 and 48 hours post treatment.
Hiatt et al.	36586412	RNA-seq, CETCh-seq of ZMYM3	HepG2 cells were modifed via super-exon insertion to either include a variant or reference base at two separate postions in ZMYM3. The super-exons intoduced a 3X FLAG tag to allow anti-FLAG antibody pull down (CETCh-seq). The two tested variants are R1274W and R688H. R1274W is deleterious leading to developmental and cognitive impariment. R688H is likely benign.
Sanchez-Priego et al.	35649373	RNA-seq, ATAC-seq, H3k27ac Cut and Run	Human pluripotent stem cells from two donors, H1 and F2, were differntiated into glutamatergic and GABAergic neruons. These were subjected to stimulation with KCl and data was collected at early and late time points. KCl induces calcium influx to neruons, mimicking the response to sustained action potential firing.
Rogers et al.	38232730	single-nuclei RNA-seq and ATAC-seq	A neuronal precursor cell line, XCL4, was differentiated to a mixed neuronal population via the BrainPhys (Bardy et al., 2015) protocol. Data was collected at the start and after seven and 14 days. Single-nuclei RNA-seq and ATAC-seq was performed using a 10X Genomics multiomics kit.

Supplemental Table 1. Data set characteristics. Descriptions of the data sets used as inputs to DegCre are provided.

Supplemental Figure 1



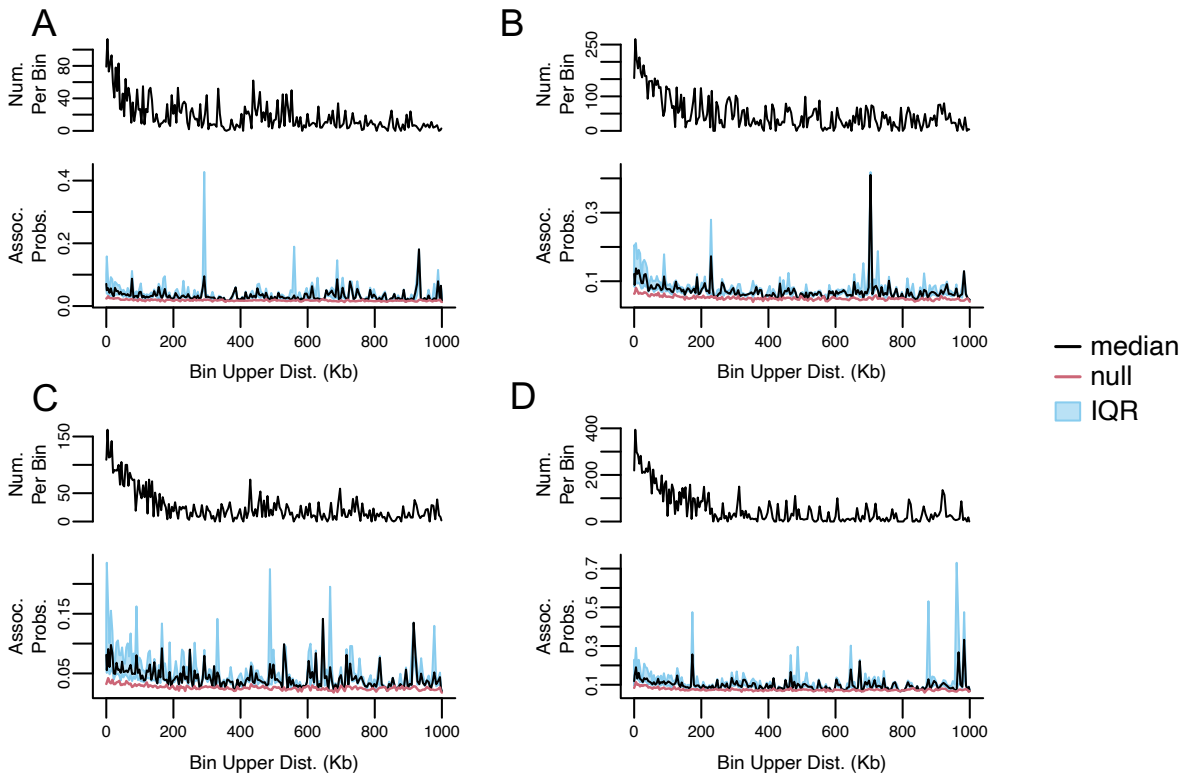
Supplemental Figure 1. Numbers of DegCre associations and input data features across all experiments. A. The bars indicate that number of DegCre associations that with an association FDR less than equal to the the 0.05. Experimental conditions are listed in the x-axis. Bar color indicates the data set and data type as shown in the legend. B.) Same plotting conventions as A.) but for the number of significantly differentially expressed genes (DEGs). Each analysis used a separate alpha value for DEG significance (Supplemental Table 1), chosen by an optimization algorithm (See Methods). C.) Same plotting conventions as A.) but for the number of CRE differential regions below a raw p-value of 0.01. Note that DegCre does not use this or any p-value cutoff.

Supplemental Table 2

Exp. Set	Condition	Optimal DEG alpha	Exp. Set	Condition	Optimal DEG alpha
McDowell DNAse	2 hr	0.05	Reed ATAC	90 min	0.001
McDowell DNAse	4 hr	0.05	Reed ATAC	2 hr	0.05
McDowell DNAse	8 hr	0.01	Reed ATAC	4 hr	0.05
McDowell DNAse	12 hr	0.05	Reed ATAC	6 hr	0.01
McDowell H3K27ac	2 hr	0.05	Reed ATAC	24 hr	0.05
McDowell H3K27ac	4 hr	0.05	Reed H3K27ac	30 min	0.001
McDowell H3K27ac	8 hr	0.05	Reed H3K27ac	60 min	0.01
McDowell H3K27ac	12 hr	0.003	Reed H3K27ac	90 min	0.01
McDowell H3K4me1	2 hr	0.05	Reed H3K27ac	2 hr	0.01
McDowell H3K4me1	4 hr	0.05	Reed H3K27ac	4 hr	0.01
McDowell H3K4me1	8 hr	0.01	Reed H3K27ac	6 hr	0.01
McDowell H3K4me1	12 hr	0.05	Reed H3K27ac	24 hr	0.01
McDowell H3K4me3	2 hr	0.001	Savic H3K27ac	24 hr	0.05
McDowell H3K4me3	4 hr	0.003	Savic H3K27ac	48 hr	0.05
McDowell H3K4me3	8 hr	0.003	Savic RNA Pol2	24 hr	0.05
McDowell H3K4me3	12 hr	0.05	Savic RNA Pol2	48 hr	0.05
McDowell NR3C1	2 hr	0.05	Sanchez-Priego ATAC	H1 Glu early	0.003
McDowell NR3C1	4 hr	0.05	Sanchez-Priego ATAC	H1 Glu late	0.003
McDowell NR3C1	8 hr	0.05	Sanchez-Priego ATAC	H1 Gaba early	0.003
McDowell NR3C1	12 hr	0.05	Sanchez-Priego ATAC	H1 Gaba late	0.05
McDowell EP300	2 hr	0.05	Sanchez-Priego H3K27Ac C&R	H1 Glu early	0.2
McDowell EP300	4 hr	0.05	Sanchez-Priego H3K27Ac C&R	H1 Glu late	0.05
McDowell EP300	8 hr	0.05	Sanchez-Priego H3K27Ac C&R	H1 Gaba early	0.005
McDowell EP300	12 hr	0.05	Sanchez-Priego H3K27Ac C&R	H1 Gaba late	0.05
McDowell JUN	2 hr	0.05	Sanchez-Priego H3K27Ac C&R	F2 Glu early	5.00E-04
McDowell JUN	4 hr	0.05	Sanchez-Priego H3K27Ac C&R	F2 glu late	5.00E-04
McDowell JUN	8 hr	0.01	Sanchez-Priego H3K27Ac C&R	F2 Gaba early	0.05
McDowell JUN	12 hr	0.05	Sanchez-Priego H3K27Ac C&R	F2 Gaba late	0.05
McDowell CEBPB	2 hr	0.05	Hiatt ZMYM3	R1274W Both	0.05
McDowell CEBPB	4 hr	0.05	Hiatt ZMYM3	R688H Both	0.005
McDowell CEBPB	8 hr	0.05	Hiatt ZMYM3	R1274W Concord	0.05
McDowell CEBPB	12 hr	0.05	Hiatt ZMYM3	R688H Concord	0.005
McDowell BCL3	2 hr	0.05	Hiatt ZMYM3	R1274W Discord	0.05
McDowell BCL3	4 hr	0.05	Hiatt ZMYM3	R688H Discord	1.00E-04
McDowell BCL3	8 hr	0.05	Rogers snATAC	clust 1 v. 2	0.05
McDowell BCL3	12 hr	0.05	Rogers snATAC	clust 1 v. 3	0.05
Reed ATAC	30 min	0.003	Rogers snATAC	clust 2 v. 3	0.05
Reed ATAC	60 min	0.05			

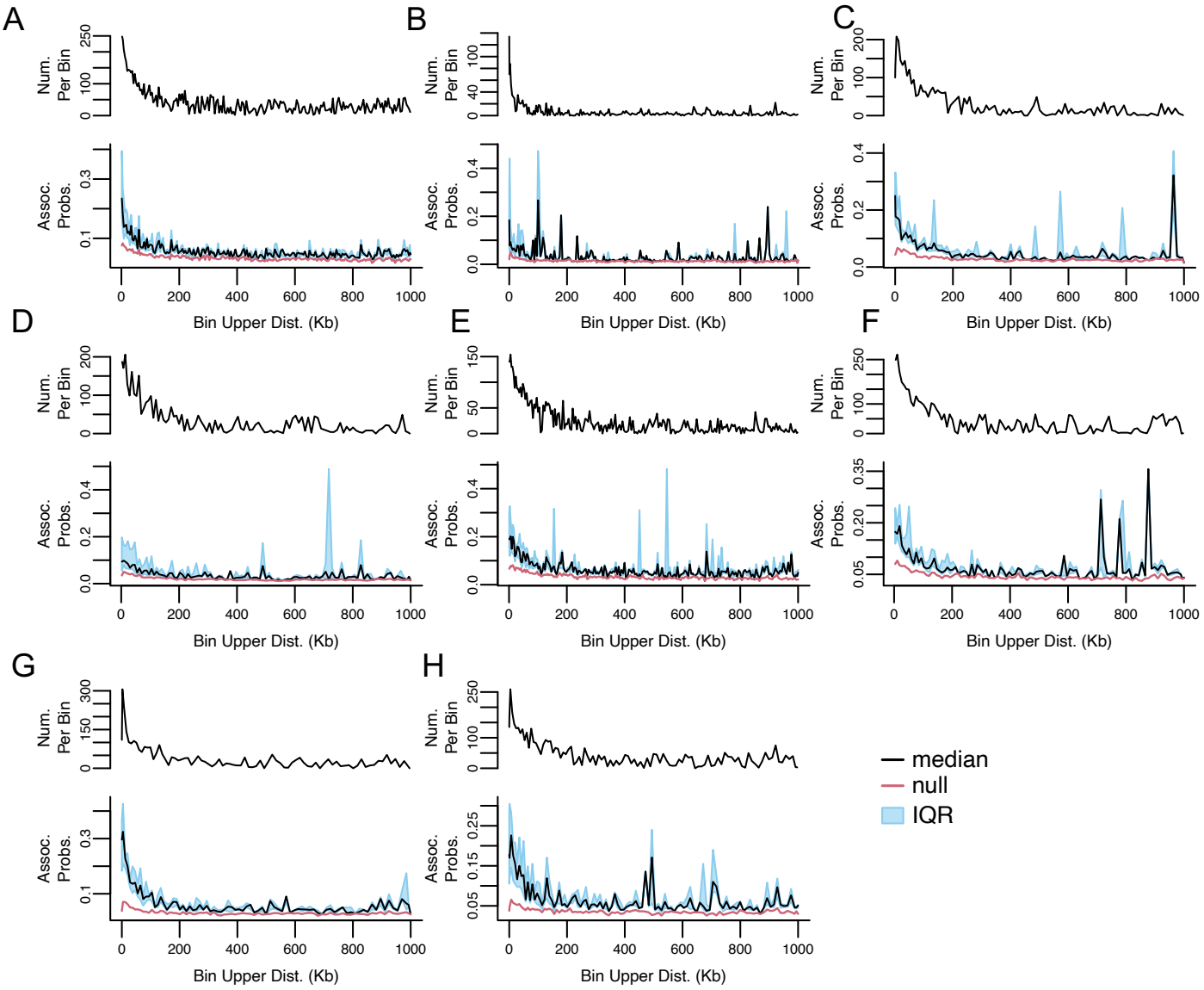
Supplemental Table 2. Selected significance thresholds for DEGs. The alpha selected by the DegCre optimization method for significance for DEGs for each experiment is shown.

Supplemental Figure 2



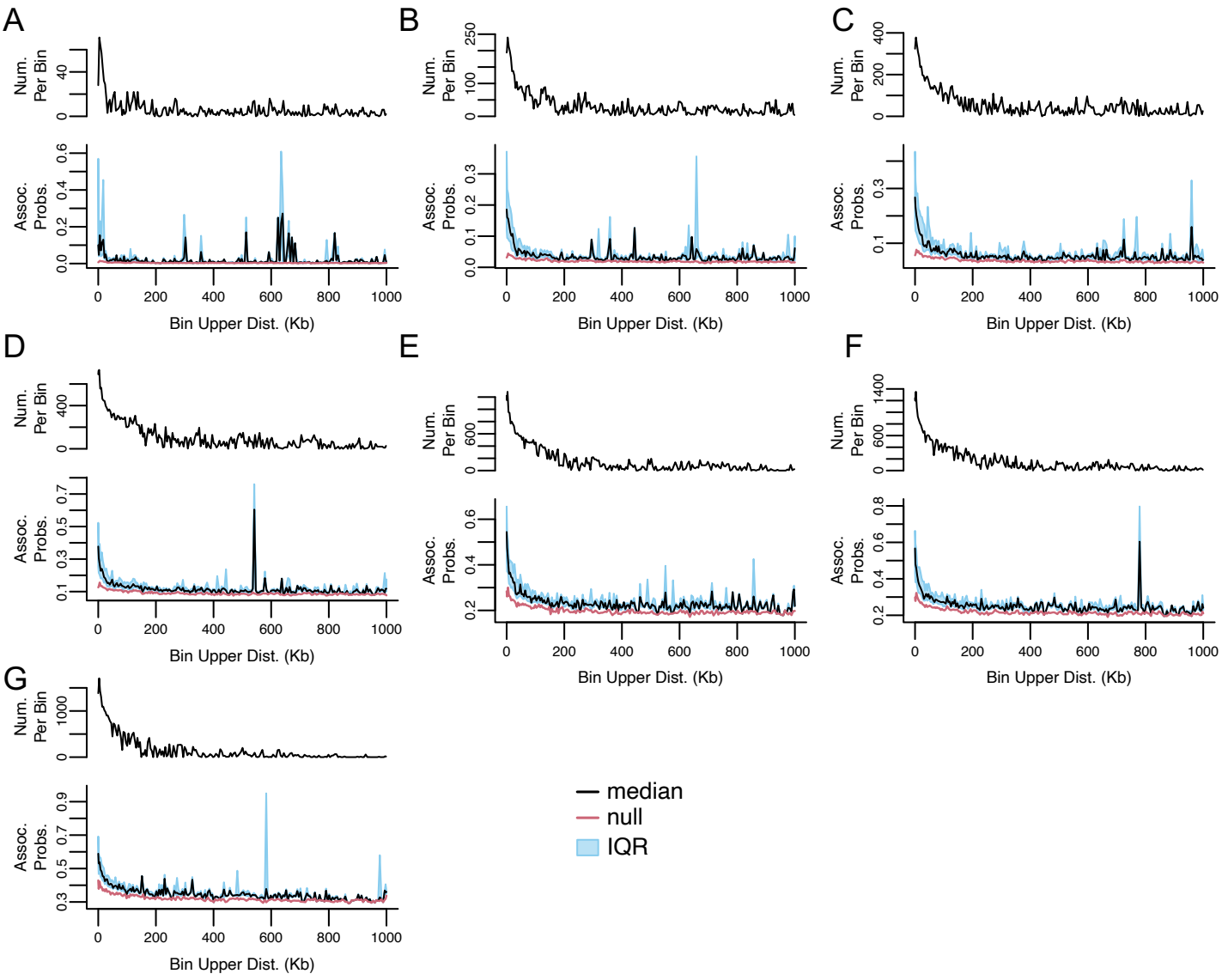
Supplemental Figure 2. DegCre associations probabilities versus bin distance for DNAse hypersensitivity data from McDowell et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on DNAse hypersensitivity data from McDowell et al. from: A.) two hours, B.) four hours, C.) eight hours, and D.) 12 hours.

Supplemental Figure 3



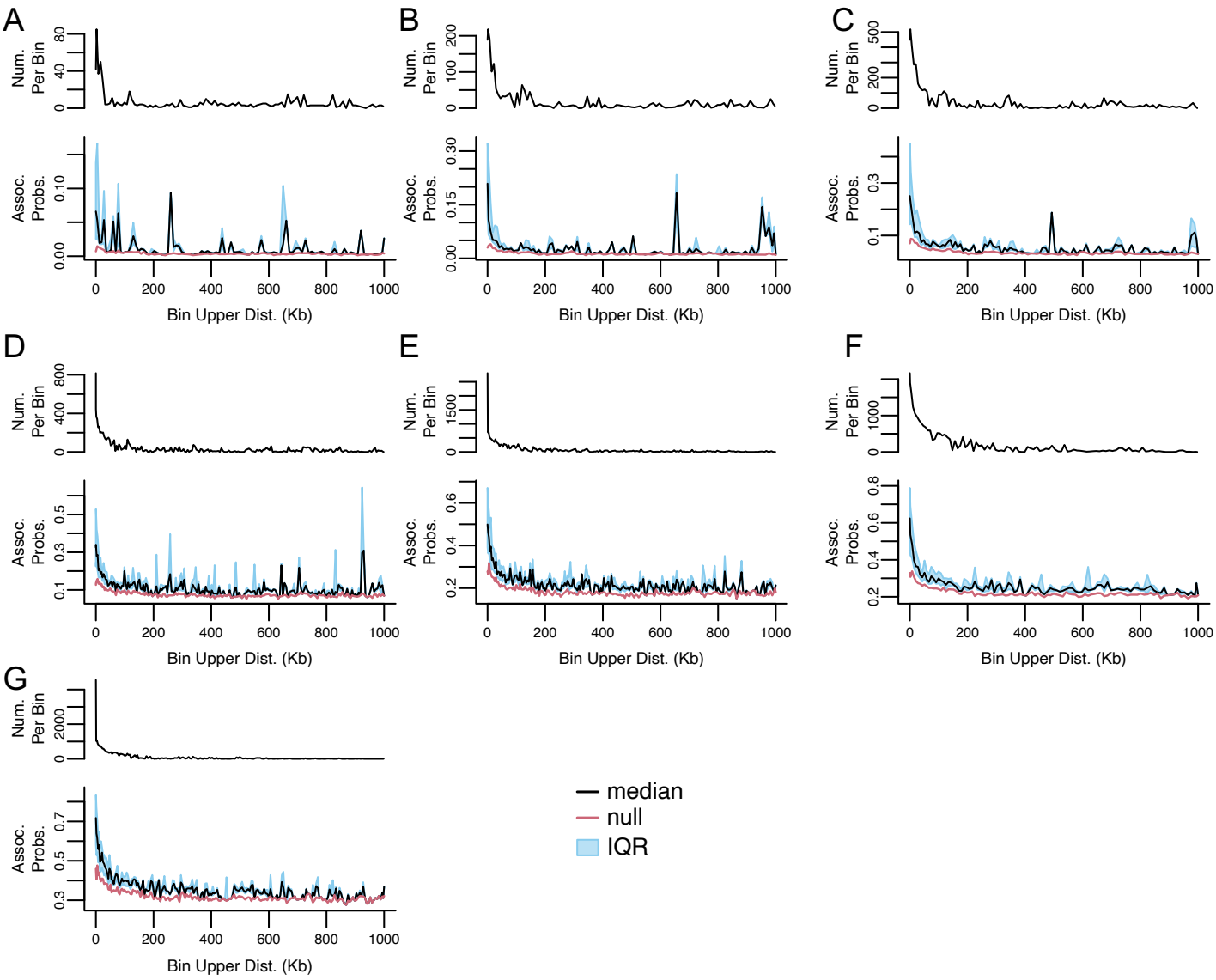
Supplemental Figure 3. DegCre associations probabilities versus bin distance for various data types at eight hours from McDowell et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that pass FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on ChIP-seq data at eight hours from McDowell et al. from: A.) NR3C1, B.) H3K4me3, C.) H3K27ac, D.) H3Kme1, E.) EP300, F.) CEBPB, G.) JUN, and H.) BCL3.

Supplemental Figure 4



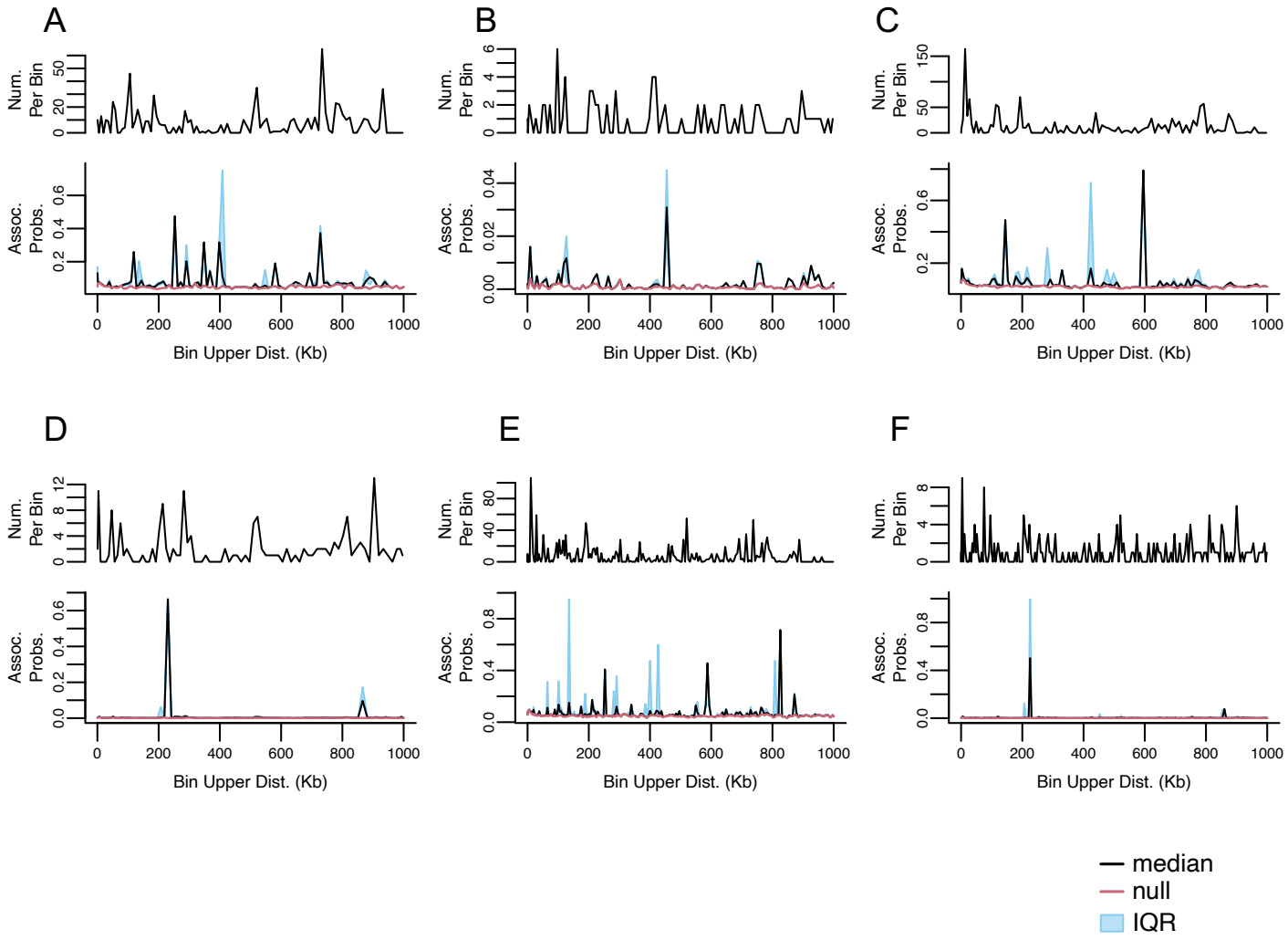
Supplemental Figure 4. DegCre associations probabilities versus bin distance for ATAC-seq data from Reed et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that pass FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on ATAC-seq data from Reed et al. from: A.) 30 minutes, B.) 60 minutes, C.) 90 minutes, D.) two hours, E.) four hours, F.) six hours, and G.) 24 hours.

Supplemental Figure 5



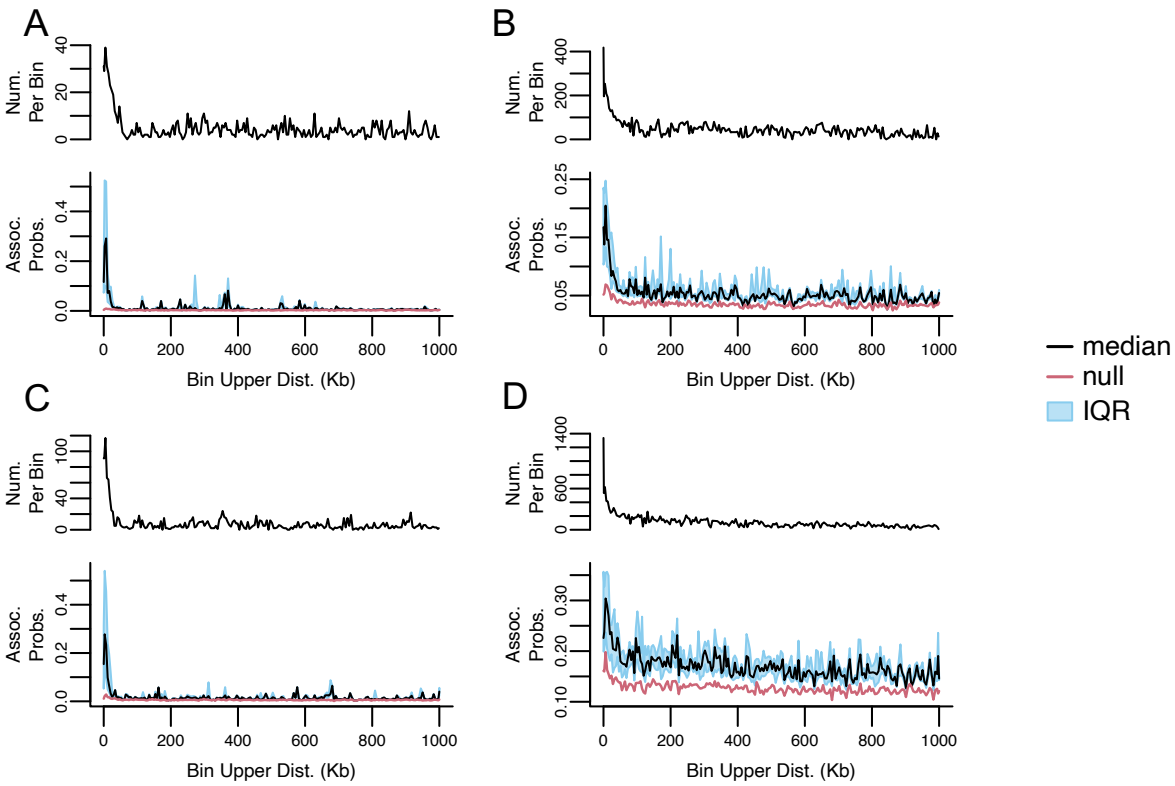
Supplemental Figure 5. DegCre associations probabilities versus bin distance for H3K27ac ChIP-seq data from Reed et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that pass FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on H3K27ac ChIP-seq data from Reed et al. from: A.) 30 minutes, B.) 60 minutes, C.) 90 minutes, D.) two hours, E.) four hours, F.) six hours, and G.) 24 hours.

Supplemental Figure 6



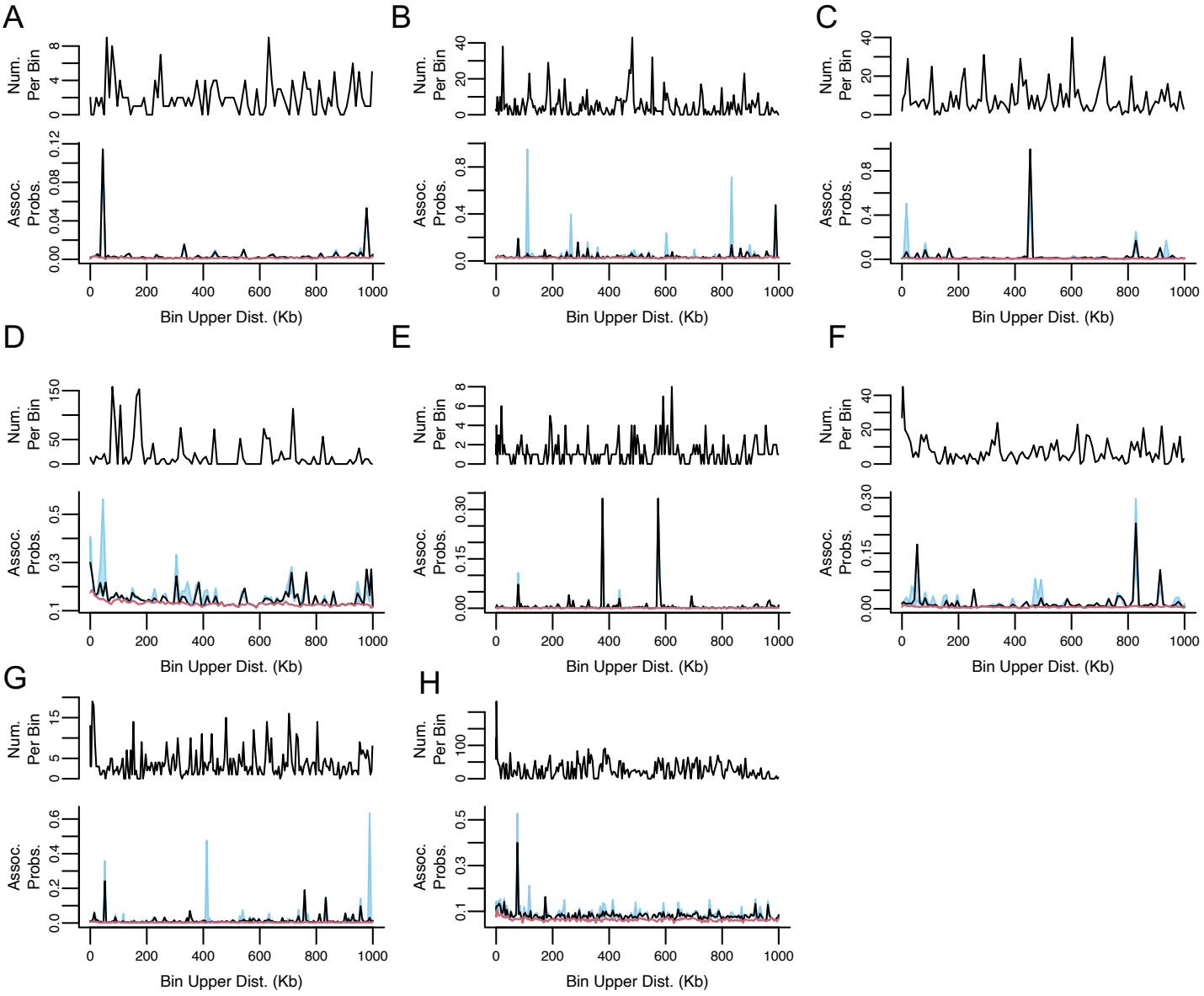
Supplemental Figure 6. DegCre associations probabilities versus bin distance for ZMYM3 mutants CETCh-seq from Hiatt et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on ZMYM3 CETCh-seq data from Hiatt et al. from: A.) R1274W concordant, B.) R688H concordant, C.) R1274W discordant, D.) R688H discordant, E.) R1274W both, and F.) R688H both,

Supplemental Figure 7



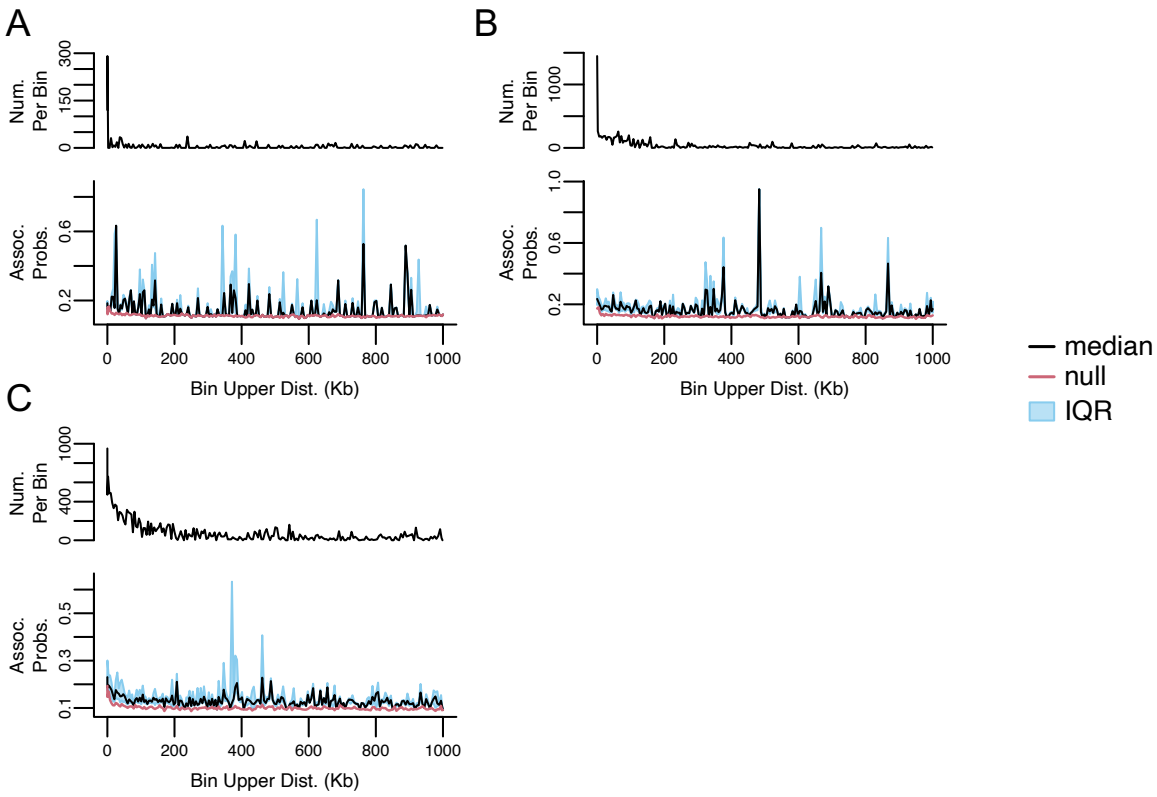
Supplemental Figure 7. DegCre associations probabilities versus bin distance for ATAC data from Sanchez-Priego et al. A-D. The black line in the upper plot half displays the number of DegCre associations per bin that FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on ATAC data from Sanchez-Preigo et al. from: A.) H1 Gluta early, B.) H1 Gluta late, C.) H1 GABA early, and D.) H1 GABA late.

Supplemental Figure 8



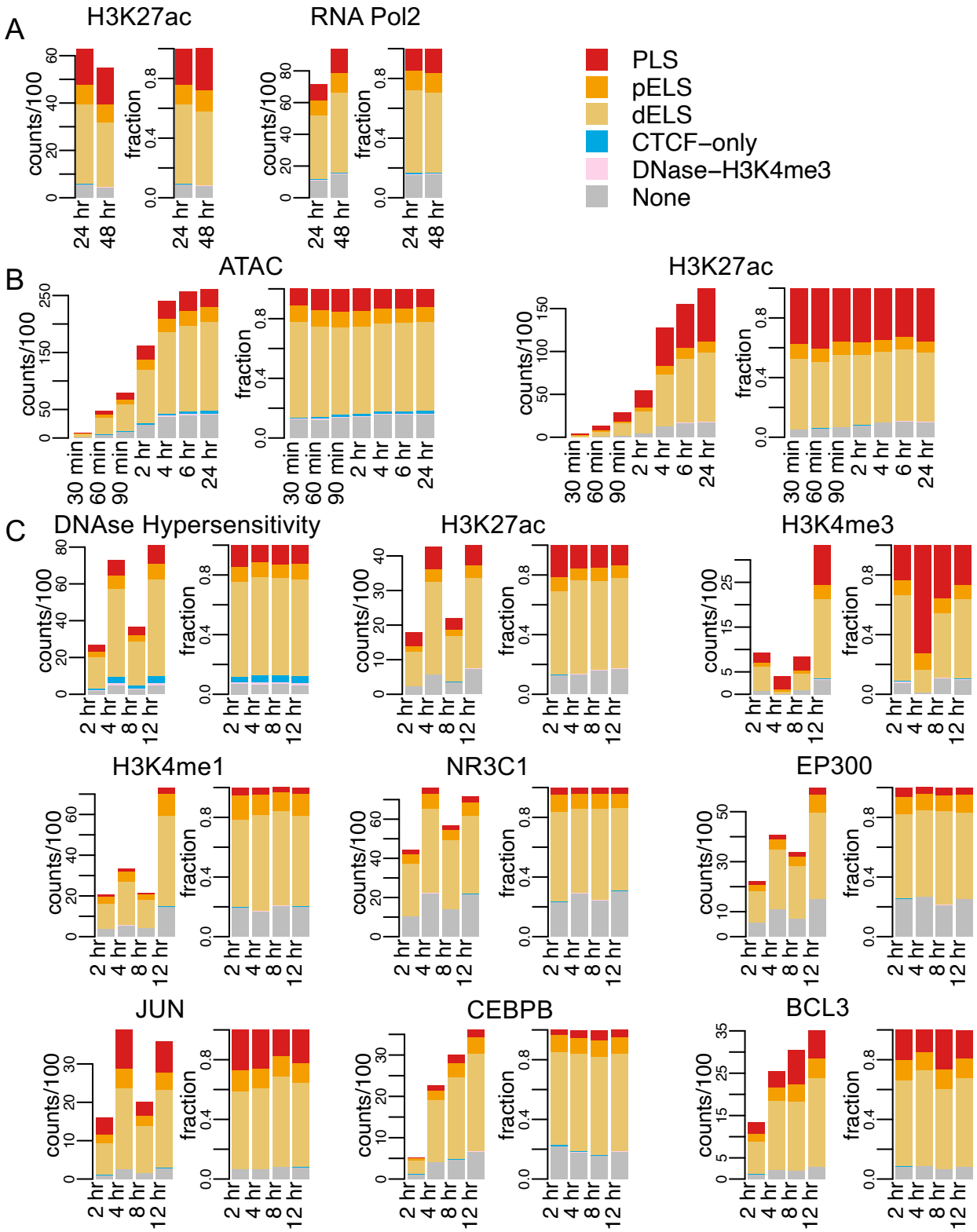
Supplemental Figure 8. DegCre associations probabilities versus bin distance for H3k27ac Cut and Run data from Sanchez-Priego et al. A-H The black line in the upper plot half displays the number of DegCre associations per bin that pass FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on H3K27ac Cut and Run data from Sanchez-Priego et al. from: A.) H1 Glu early, B.) H1 Glu late, C.) H1 GABA early, D.) H1 GABA late, E.) F2 Glu early, F.) F2 Glu late, G.) F2 GABA early, and H.) F2 GABA late.

Supplemental Figure 9



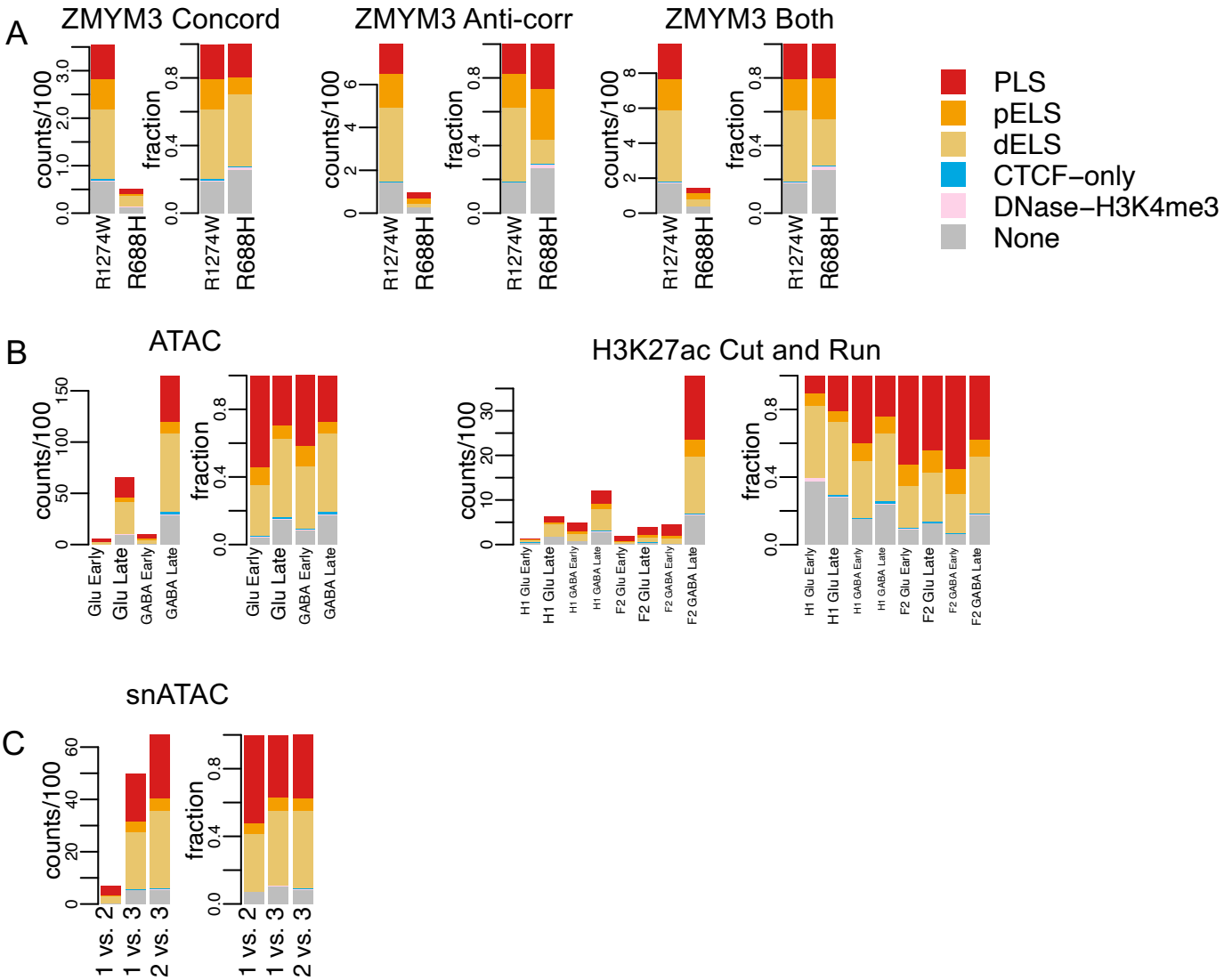
Supplemental Figure 9. DegCre associations probabilities versus bin distance for single-nuclei ATAC data from Rogers et al. A-C. The black line in the upper plot half displays the number of DegCre associations per bin that FDR less than or equal to 0.05. The bottom half displays the per bin DegCre association probability. The common x-axis shows for each bin the association distance from TSS to CRE. Each bin comprises a range of distances with the upper bound of that range plotted here. The black line indicates the median value for each bin and the blue region indicates the inner quartile range (IQR). The red line shows the per bin probability considering only the bin distance, used as the null in the DegCre FDR calculation. Plots are based on ATAC data from Rogers et al. from: A.) cluster 1 vs, 2, B.) cluster 1 vs. 3, and C.) cluster 2 vs. 3.

Supplemental Figure 10



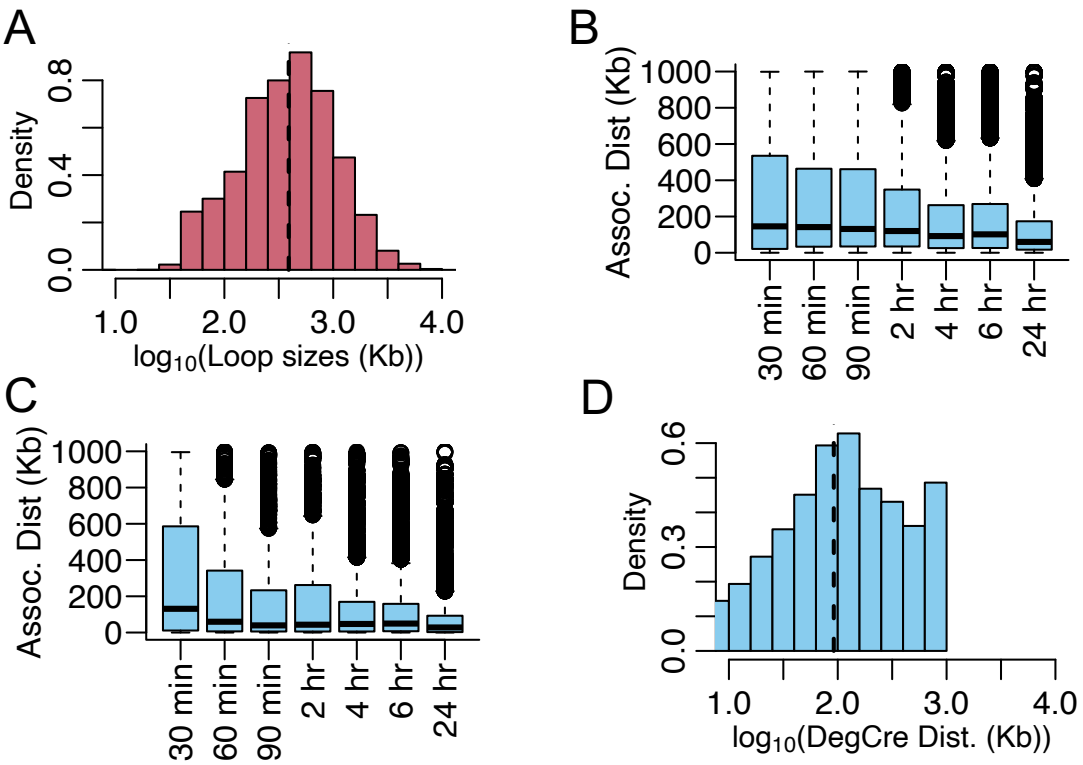
Supplemental Figure 10. Annotation of CREs in significant DegCre associations to ENCODE cCREs. A. For data from Savic et al. the barplots show the number and fraction of overlap with ENCODE cCREs for CREs with a DegCre association based on the data type indicated. Only DegCre associations with an FDR less than 0.05 were considered. B. Same conventions as A but for data from Reed et al. C. Same conventions as A but for data from McDowell et al.

Supplemental Figure 11



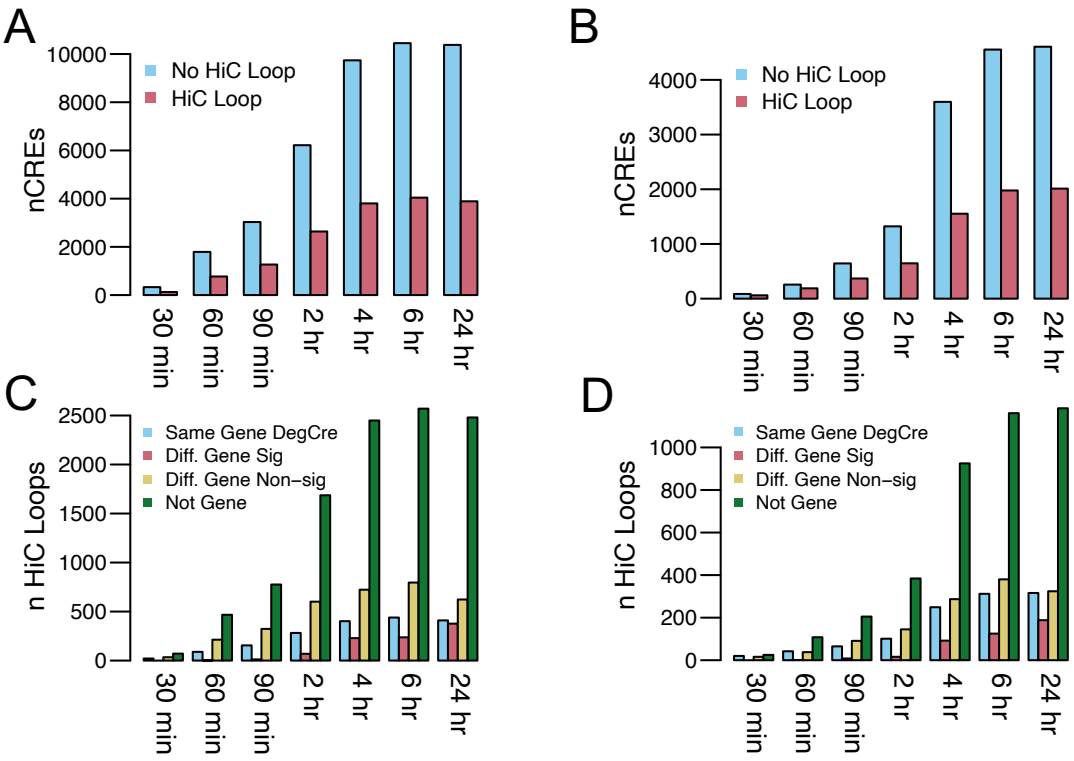
Supplemental Figure 11. Annotation of CREs in significant DegCre associations to ENCODE cCREs. A. For data from Hiatt et al. the barplots show the number and fraction of overlap with ENCODE cCREs for CREs with a DegCre association based on the data type indicated. Only DegCre associations with an FDR less than 0.05 were considered. B. Same conventions as A but for data from Sanchez-Priego et al. C. Same conventions as A but for data from Rogers et al.

Supplemental Figure 12



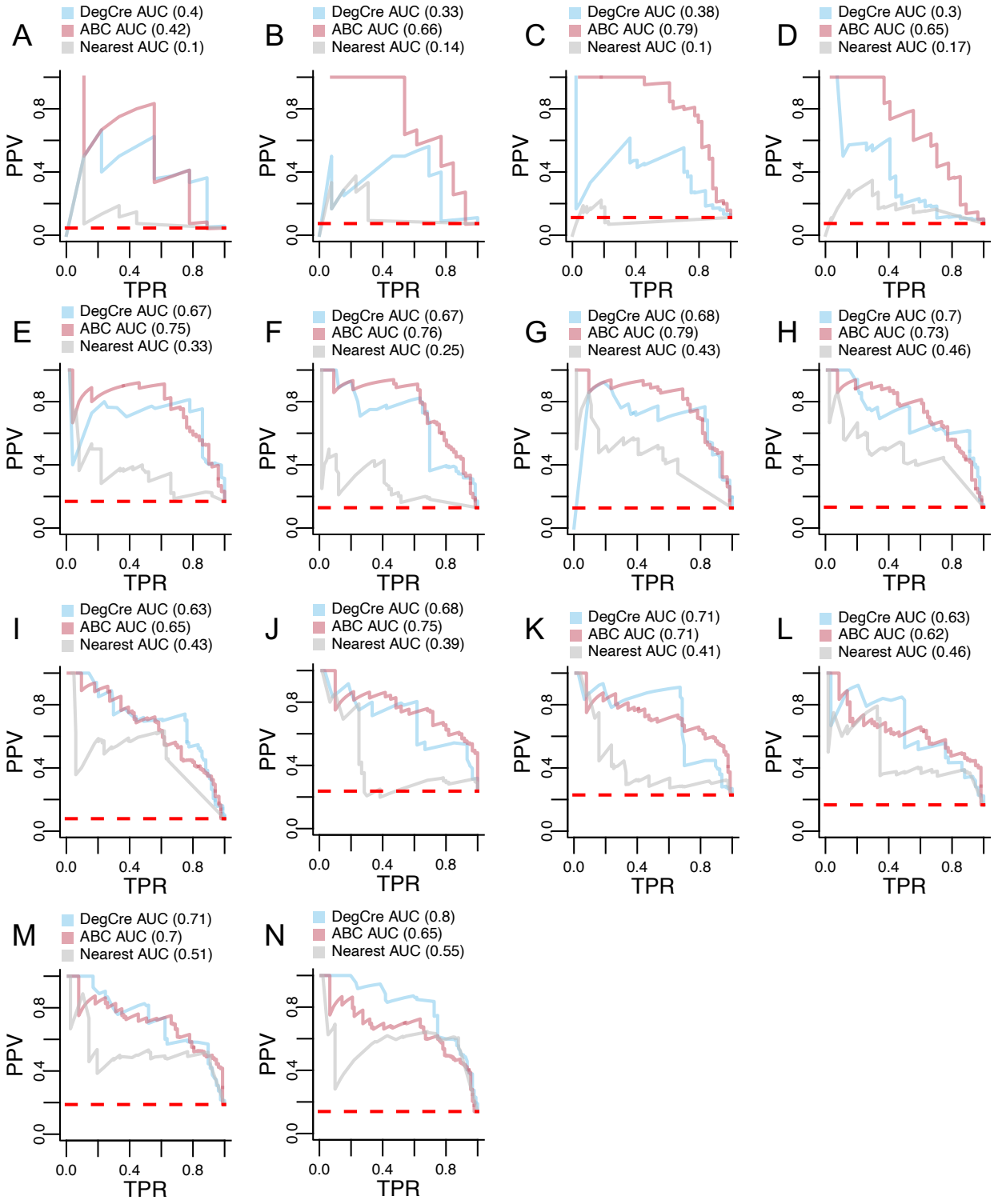
Supplemental Figure 12. HiC Loop and DegCre association distance properties. A. Histogram of the HiC loop size (distance from anchor-to-anchor midpoints) for Reed et al. data. Dashed line represents the median, 390 kb. B. Boxplot of association distances (from Cre to gene TSS) of DegCre associations from ATAC-seq data from Reed et al. C. Boxplot of association distance of DegCre associations from H3K27Ac ChIP-seq data from Reed et al. D. Histogram of DegCre association distance at the two-hour time point for ATAC-seq data from Reed et al. The dashed line indicates the median, 91,949 bp.

Supplemental Figure 13



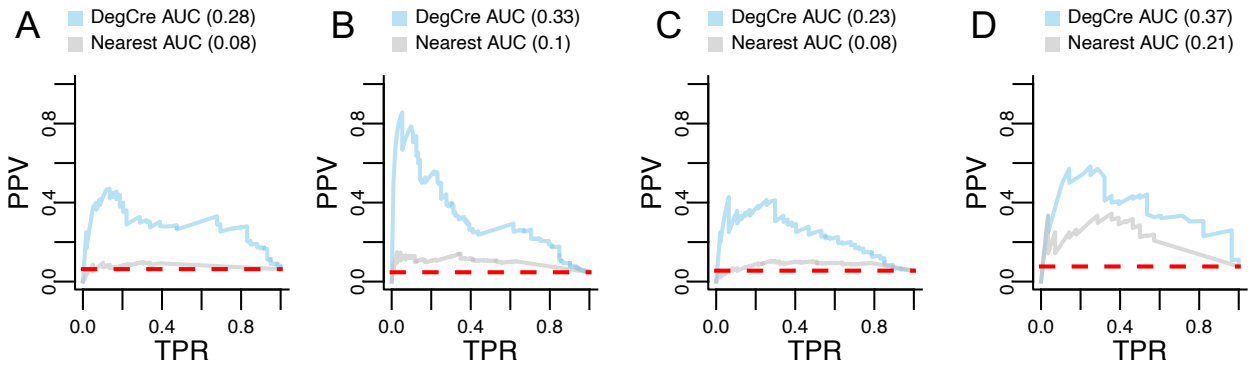
Supplemental Figure 13. Comparison of CRE to DEG agreement in HiC Loops and DegCre Associations. A. For every ATAC-seq CRE from Reed et al (defined as regions with a DegCre association with FDR <= 0.05), the bar show the number that also have one anchor of a HiC loop or not. B. Same as A but for CREs derived from H3K27Ac data. C. For each ATAC-seq CRE with at least one HiC loop, the bars show number fitting the listed categories of overlap types. “Same Gene DegCre” means that the other HiC anchor overlaps the same gene as the DegCre association. ”Diff. Gene Sig” means that the other HiC anchor overlaps the TSS of a different gene as the DegCre association that is significantly differentially expressed (passes FDR cutoff used by DegCre calculation). ”Diff. Gene Non-sig” means that the other HiC anchor overlaps TSS of a different gene as the DegCre association that is not significantly differentially expressed. “Not Gene” means that the other anchor does not overlap any gene TSS. D. Same as C except for DegCre associations derived from H3K27Ac data.

Supplemental Figure 14



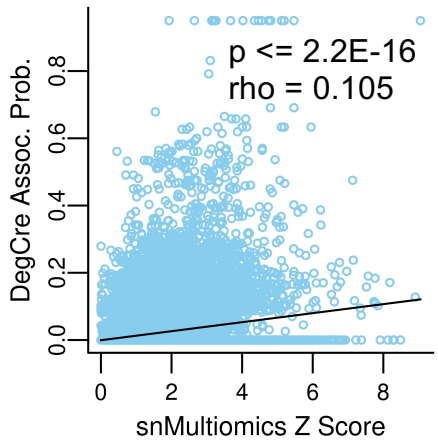
Supplemental Figure 14. Evaluation of DegCre associations with CRISPR data from Nasser et al. Each panel is a Precision-Recall plot using CRISPR perturbation results from Nasser et al. as the standard. PPV= Positive predictive value (Precision) and TPR = True positive rate (Recall). Dashed red line indicates “zero-skill”. ABC scores from Nasser et al. are plotted as well. A model in which a CRE is assigned to the nearest DEG (passing adjusted p-value cutoff), “Nearest”, is also shown. Only plots from data sets that had at least 25 overlaps with a positive CRISPR association are shown. Plots are based on A-D.) Savic et al. H3K27Ac ChIP-seq at 24 and 48 hrs, and RNA Pol2 ChIP-seq at 24 and 48 hrs, E-I.) Reed et al. ATAC-seq data at 90 min, 2 hrs, 4 hrs, 6 hrs, and 24 hrs, J-N.) Reed et al. H3K27Ac ChIP-seq data at 90 min, 2 hrs, 4 hrs, 6 hrs, and 24 hrs.

Supplemental Figure 15



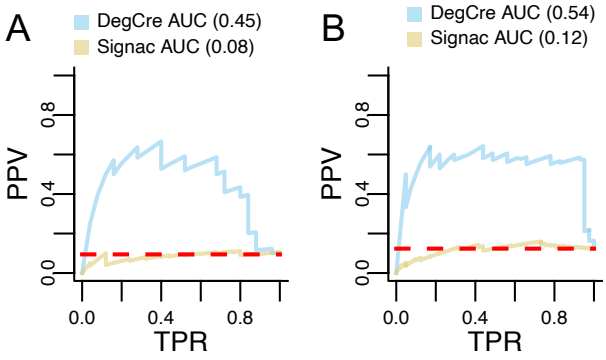
Supplemental Figure 15. Evaluation of DegCre associations from selected data sets with CRISPR data from Gasperini et al. Each panel is a Precision-Recall plot using CRISPR perturbation results from Gasperini et al. as the standard. PPV= Positive predictive value (Precision) and TPR = True positive rate (Recall). Dashed red line indicates “zero-skill”. A model in which a CRE is assigned to the nearest DEG (passing adjusted p-value cutoff), “Nearest”, is also shown. Selected plots from data sets that had at least 25 overlaps with a positive CRISPR association are shown. Plots are based on A.) Savic et al. RNA Pol2 ChIP-seq at 24 hrs, B.) Reed et al. ATAC at 2 hrs, C.) Sanchez-Priego et al. ATAC-seq from H1 Gaba late, and D.) McDowell et al. BCL3 ChIP-seq at 12 hrs.

Supplemental Figure 16



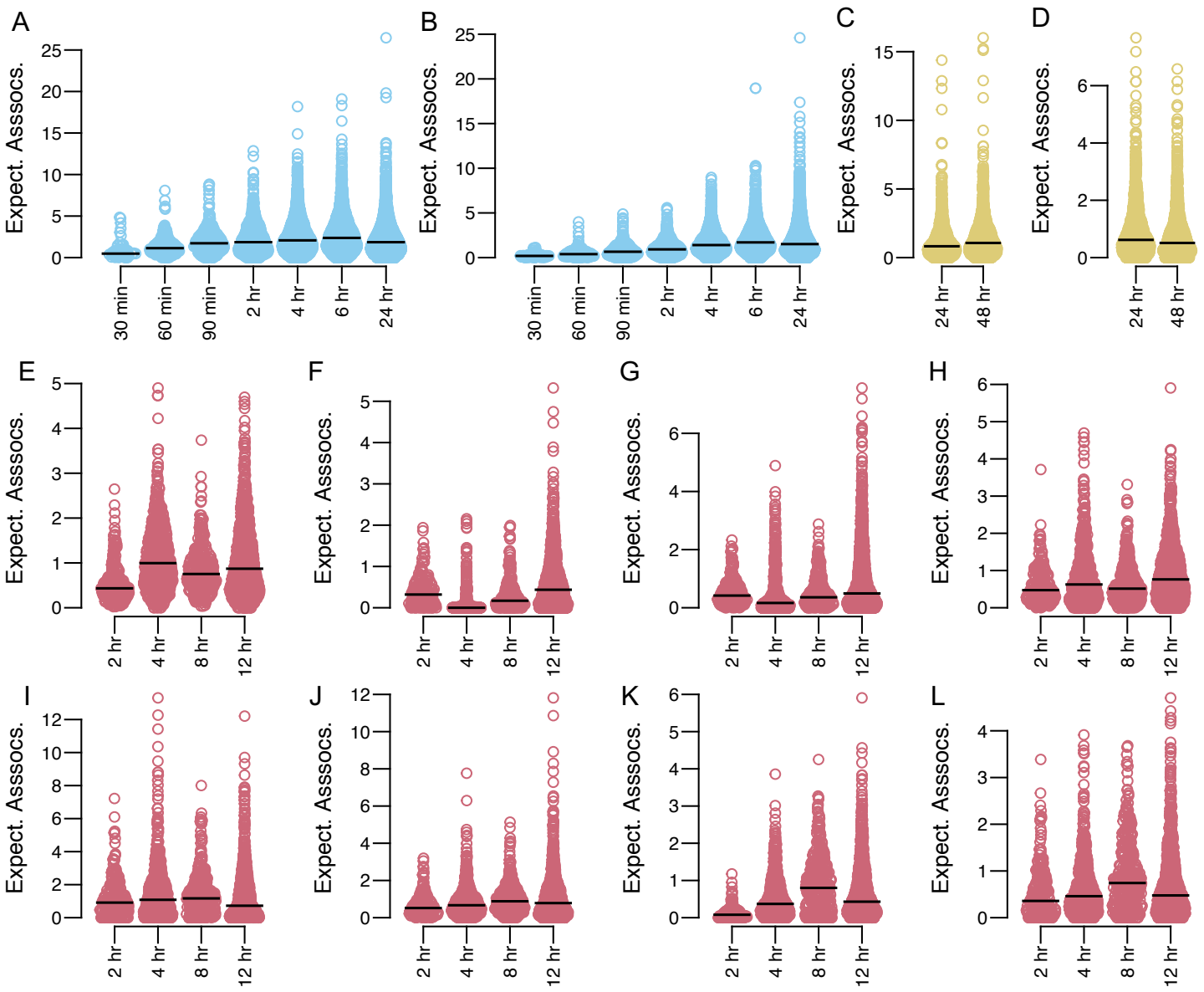
Supplemental Figure 16. Correlation of DegCre and Signac association scores. The Signac sn-multiomics association scores are plotted versus the DegCre association probabilities. The p-value and rho value are from a Spearman correlations test.

Supplemental Figure 17



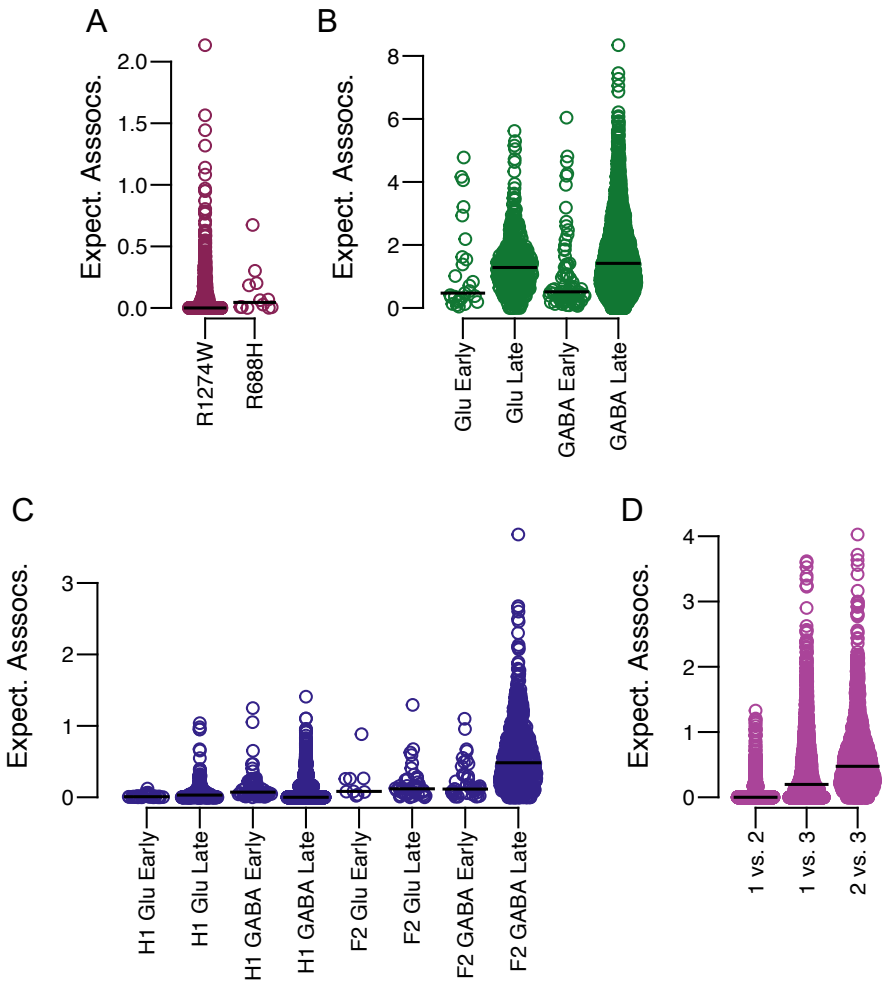
Supplemental Figure 17. More comparisons of DegCre associations from single-nuclei multiomics to Signac. Each panel is a Precision-Recall plot using CRISPR perturbation results from Gasperini et al. as the standard. PPV= Positive predictive value (Precision) and TPR = True positive rate (Recall). Dashed red line indicates “zero-skill”. DegCre associations were calculated on pseudo-bulked RNA and ATAC data. Signac (Stuart et al.) was applied to single-nuclei data to generate linkage scores. Predictions are based on data from A.) cluster 2 versus cluster 1, and B.) cluster 3 versus cluster 1.

Supplemental Figure 18



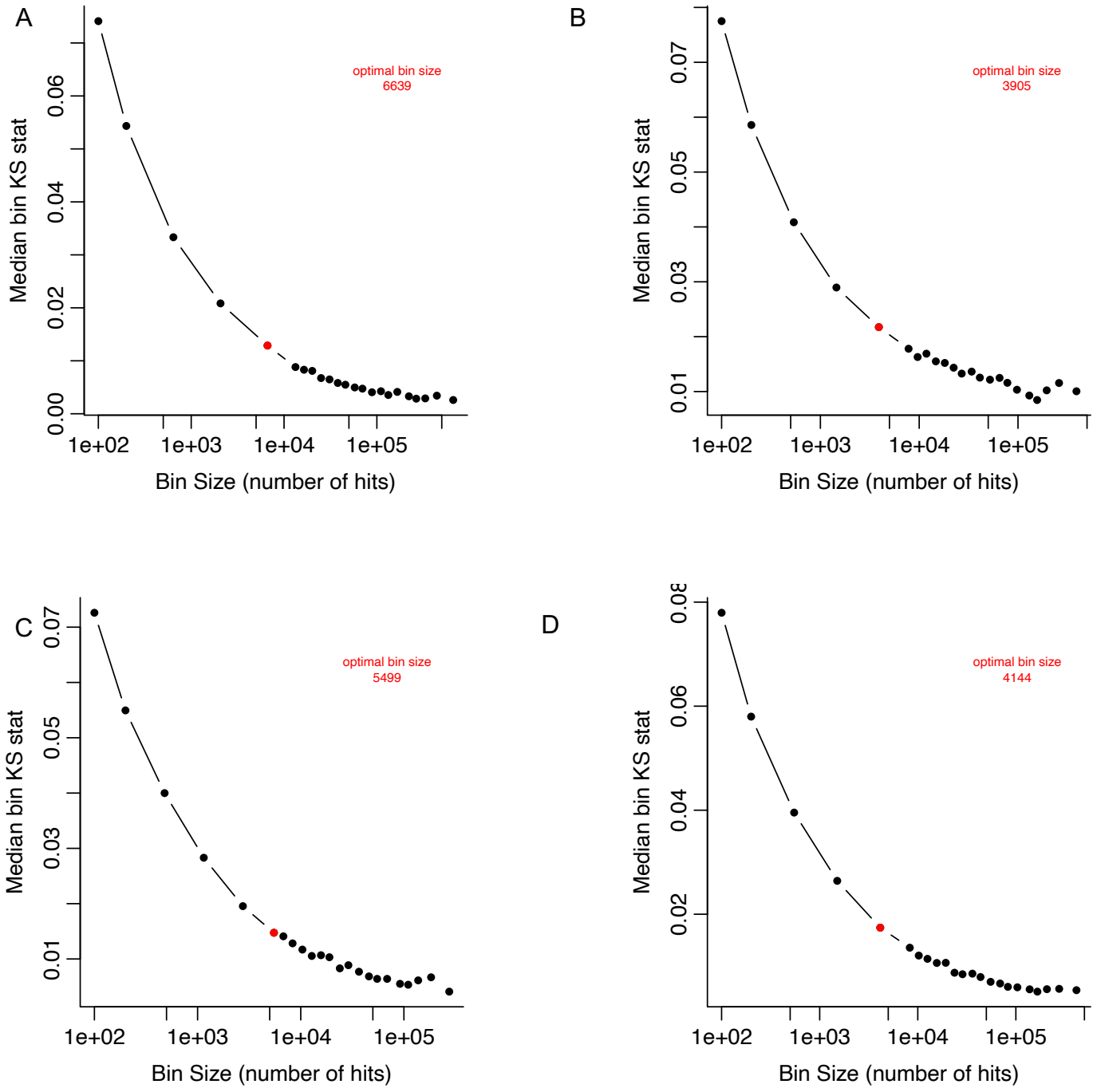
Supplemental Figure 18. Distribution of expected DegCre associations per DEG across data sets and time point. Each boxplot shows the expected DegCre associations for each significant DEG (FDR less than or equal to 0.05) at each time point. The black line indicates the median value. The jitter width of the points is proportional to the local point density. Data are from: A.) Reed et al. ATAC-seq, B.) Reed et al. H3K27ac ChIP-seq, C.) Savic et al. RNA-Pol2 ChIP-seq, D.) Savic et al. H3K27ac ChIP-seq, E.) McDowell et al. DNAse Hypersensitivity, F.) McDowell et al. H3K4me3 ChIP-seq, G.) McDowell et al. H3K4me1 ChIP-seq, H.) McDowell et al. H3K27ac ChIP-seq, I.) McDowell et al. NR3C1 ChIP-seq, J.) McDowell et al. EP300 ChIP-seq, K.) McDowell et al. CEBPB ChIP-seq, and L.) McDowell et al. BCL3 ChIP-seq.

Supplemental Figure 19



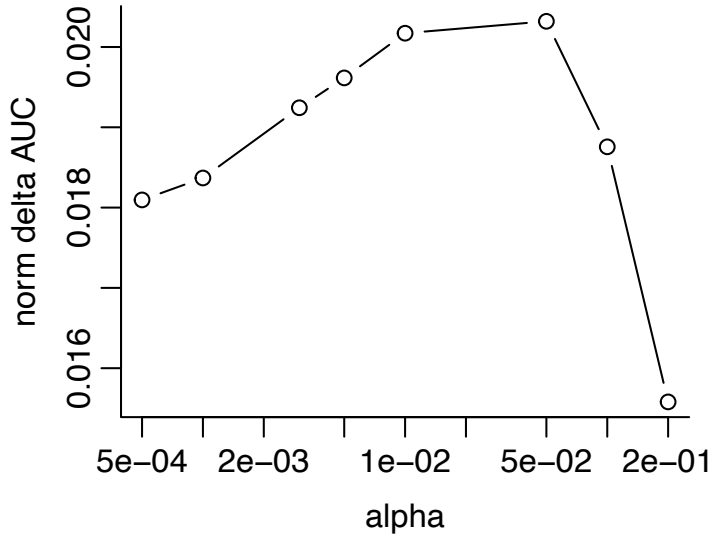
Supplemental Figure 19. More distributions of expected DegCre associations per DEG across data sets. Each boxplot shows the expected DegCre associations for each significant DEG (FDR less than or equal to 0.05) at each time point. The black line indicates the median value. The jitter width of the points is proportional to the local point density. Data are from: A.) Hiatt et al. ZMYM3 ChIP-seq anti-correlated analysis, B.) Sanchez-Priego et al. ATAC-seq, C.) Sanchez-Priego et al. H3K27Ac Cut and Run, D.) Rogers et al. single-nuclei ATAC-seq.

Supplemental Figure 20



Supplemental Figure 20. Distance bin size optimization plots. Each plot shows the median KS statistic across bins for the CRE p-values compared to the global distribution on the y-axis. The x-axis show the number of associations per bin. The red dot shows the selected bin size. Plots are from: A.) ATAC-seq data at two hours from Reed et al., B.) NR3C1 ChiP-seq data at two hours from McDowell et al., C.) H3K27ac ChIP-seq data at 24 hours from Savic et al., and D.) ATAC-seq data from Sanchez-Priego et al. from H1 GABA late.

Supplemental Figure 21



Supplemental Figure 21. Results of DEG α optimization. The plot shows the difference between the Precision-Recall AUC of the DegCre predictions at each DEG alpha value from the AUC of shuffled predictions divided by 1 – AUC of the shuffled predictions. The Precision-Recall is calculated from the expected precision of the DegCre associations, which is the average association probability for all associations passing a given threshold. Plot is generated from ATAC-sea data at two hours from Reed et al.