SUPPLEMENTAL MATERIALS

CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins

Daniel Savic[1], E. Christopher Partridge[1], Kimberly M. Newberry[1], Sophia B. Smith[2], Sarah K. Meadows[1], Brian S. Roberts[1], Mark Mackiewicz[1], Eric M. Mendenhall[1,2] and Richard M. Myers[1]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

[2]University of Alabama at Huntsville, Huntsville, AL, USA

TABLE OF CONTENTS

Supplemental Table 1. Gene expression of transcription factors in HepG2 and MCF7 cells. RPKM values for tested transcription factors in HepG2 and MCF7 cells are displayed.

Supplemental Table 2. pFETCh homology arm and gRNA designs. A summary of all gRNA sequences and homology arm sequences for all DNA-binding proteins and across all DNA-binding proteins is given. For gRNAs, the species, number of gRNAs used, sequence of all gRNAs, on-target (Doench et al. 2014) scores and off-target (Hsu et al. 2013) scores are shown. For 5-prime and 3-prime homology arms, the size of each arm, the type of homology arm (how it was generated; PCR or gBlock) and the entire sequences are given. The DNA sequences that are deleted, their number, including notes, as well as synonymous variants added to 5-prime homology arm at exonic ATF1 sequences spanning the gRNA site to prevent Cas9 re-cleavage events.

Supplemental Table 3. PCR primer sequences. Primer sequences for cloning of both 5-prime and 3-prime homology arms are shown. Sequences in lowercase denote pFETCh plasmid sequences required for cloning (Gibson Assembly). Homology arms generated through gBlock synthesis are also denoted. PCR validation primers are further given for all DNA binding proteins tested.

Supplemental Table 4. Validation of cloned PCR sequences spanning 5-prime and 3-prime homologous recombination sites. The raw Sanger sequence data for cloned PCR fragments spanning the 5-prime (Endogenous sequence - FLAG tag) and 3-prime (Neomycin - Endogenous sequence) homologous recombination sites for all DNA-binding proteins is shown. Sequences mapping to the endogenous locus are shown in black while sequences mapping to the pFETCh plasmid are shown in blue.

Supplemental Table 5. ChIP-seq data summary. A summary of all ChIP-seq experiments is given, including (from left to right) the cell line, the ChIP-seq experiment name, antibody used, the number of aligned reads, the number of binding sites identified, normalized strand coefficients (NSC), relative strand correlations (RSC), quality scores, the number (Number) and percent (%) of peaks that are shared between technical CETCh-seq replicates (Tech Rep), the number (Number) and percent (%) of peaks that are shared between biological CETCh-seq replicates (Bio Rep), the number (Number) and percent (%) of peaks shared between CETCh-seq and standard ChIP-seq replicate experiments using transcription factor antibodies (WT Rep1 and WT Rep2) and the number (Number) and percent (%) of peaks that are shared between CETCh-seq experiments and CRISPR-modified ChIP-seq replicate experiments using transcription factor antibodies (Modified).

Supplemental Figure 1. PCR validation of homologous recombination. PCR validation of CRISPR-mediated homologous recombination for all HepG2 and MCF7 experiments. (A) A schematic of primer locations at transcription factor genetic loci is given. Below, 5'- and 3'-end homologous recombination predicted PCR band sizes are displayed. (B-H) Gel images for all validations are given. From left to right, the gel images display ladder, 5' end and 3' end homology PCR products. Names of each transcription factor and biological replicate (Rep1 and Rep2) are shown above each gel. Band sizes are given at the left of each gel and asterisks mark the synonymous band sizes across rows of gels. HepG2 results are given in (B)-(F), while MCF7 RAD21 validation data is displayed in (G) and (H).

Supplemental Figure 2. Western blot validation of FLAG tagged transcription factors. Protein validation experiments are displayed. Western blot images in (A)-(C) and IP western blot images in (D) and (E) are given for all HepG2 and MCF7 experiments (cell line denoted below western image). From left to right, westerns show protein size ladders, experimental and control western experiments. Red arrows mark the location of relevant bands. (A) HepG2 RAD21 validation using FLAG antibody (FLAG) and RAD21 antibody (RAD21) in CRISPR-modified cell lines. A ~130kDa band, the predicted size of full-length RAD21, is seen using both antibodies. (B) MCF7 RAD21 validation using FLAG antibody (FLAG) in CRISPR-modified MCF7 cell lines and FLAG antibody in wild-type MCF7 cells (WT FLAG). A ~130kDa band is only observed in the CRISPR-modified MCF7 cells. (C) HepG2 GABPA validation using FLAG antibody (FLAG) and GABPA antibody (GABPA) in CRISPR-modified cell lines. Several bands ~50-60 kDa in line with GABPA protein sizes are visible using both antibodies. (D) HepG2 CREB1 IP western validation using FLAG antibody (FLAG IP) or IgG (IgG IP) for IP pulldown followed by FLAG antibody blotting. A ~37kDa and ~46KDa bands, sizes identified for the CREB1 protein, are only observed in the FLAG IP pulldown experiment. (E) HepG2 ATF1 IP western validation using FLAG antibody (FLAG IP) or IgG (IgG IP) IP pulldown followed by FLAG antibody blotting. A ~40kDa band, the predicted size of ATF1, is only observed in the FLAG IP pulldown experiment.

Supplemental Figure 3. *RAD21* Sanger sequencing of PCR amplicons representing homologous recombination site in HepG2 cells. (A) The endogenous sequence of the 3-prime end of *RAD21* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram trace of PCR amplicons spanning 3-prime homologous recombination sites. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graph below trace.

Supplemental Figure 4. *CREB1* Sanger sequencing of PCR amplicons representing homologous recombination site in HepG2 cells. (A) The endogenous sequence of the 3-prime end of *CREB1* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram traces of PCR amplicons spanning 3-prime homologous recombination sites. Two replicate transfection experiments are shown. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graphs below traces.

Supplemental Figure 5. *ATF1* Sanger sequencing of PCR amplicons representing homologous recombination site in HepG2 cells. (A) The endogenous sequence of the 3-prime end of *ATF1* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram trace of PCR amplicons spanning 3-prime homologous recombination sites. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graph below trace.

Supplemental Figure 6. *GABPA* Sanger sequencing of PCR amplicons representing homologous recombination site in HepG2 cells. (A) The endogenous sequence of the 3-prime end of *GABPA* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram trace of PCR amplicons spanning 3-prime homologous recombination sites. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graph below trace.

Supplemental Figure 7. Analysis of CRISPR/Cas9 genome editing disruptions at untagged alleles. (A) The fraction of alleles in the HepG2 polyclonal population that show insertions or deletions (indels, in red) compared to unmodified alleles (in blue). Data is shown for *ATF1*, *RAD21* and *CREB1*. (B) Examples of the low frequency sequence disruptions at the *ATF1* coding sequence are shown. An image of the gRNA orientation at the coding is shown above. The protospacer adjacent motif (PAM) sequence is shown (in yellow) along with the putative nuclease cleavage site (red arrow). Examples of disruptions at amino acid sequences and their associated frequencies are shown below.

Supplemental Figure 8. HepG2 transcription factor motifs. Enriched binding motifs for all HepG2 experiments are shown. The names of ChIP-seq experiments are given above each DNA sequence motif. CETCh-seq experiments are labeled as FLAG while ChIP-seq experiments using transcription factor antibodies in wild-type cells are labeled as WT. Biological replicates are labeled rep 1 or rep 2 for CREB1 data. For ATF1, technical replicates are labeled rep1 and rep2.

Supplemental Figure 9. HepG2 FLAG tag ChIP-seq correlations. (A) Rank correlations of normalized sequence read counts between standard ChIP-seq replicates (WT replicate 1 and WT replicate 2) using transcription factor antibodies in wild-type HepG2 cells. (B) Rank correlations of normalized sequence read counts between standard ChIP-seq data (WT) using transcription factor antibodies in wild-type HepG2 cells and CETCh-seq experiments (FLAG) using FLAG antibodies in CRISPR-modified HepG2 cells. Rank correlations in (A) are given. In (B), average rank correlations for all FLAG and wild-type ChIP-seq replicate pairwise comparisons are displayed. Transcription factor names are given above each plot.

Supplemental Figure 10. HepG2 Epitope tag ChIP-seq technical replicate correlations. Rank correlations of normalized sequence read counts between CETCh-seq technical replicates (FLAG replicate 1 and FLAG replicate 2). Rank correlations are displayed in each plot while tagged transcription factor names are given above plots.

Supplemental Figure 11. HepG2 CREB1 ChIP-seq site correlations with FLAG tag biological replicate. (A) Rank correlations of normalized sequence read counts between a second CETCh-seq biological replicate (FLAG Biological rep 2) and standard ChIP-seq data (WT) using transcription factor antibodies in wild-type HepG2 cells. (B) Rank correlations of normalized sequence read counts between CETCh-seq biological replicates (FLAG Biological rep 1 and FLAG Biological rep 2). Average rank correlations for all FLAG and/or wild-type ChIP-seq replicate pairwise comparisons are displayed on each plot.

Supplemental Figure 12. HepG2 correlations with CRISPR-modified ChIP-seq data. (A) Rank correlations of normalized sequence read counts between a ChIP-seq experiments in CRISPR-modified cells using transcription factor antibodies (Modified cells) and ChIP-seq data in wild-type HepG2 cells using transcription factor antibodies (WT). (B) Rank correlations of normalized sequence read counts between ChIP-seq experiments in CRISPR-modified cells using transcription factor antibodies (Modified cells) and CETCh-seq experiments using FLAG antibodies in CRISPR-modified cells (FLAG). Average rank correlations for all pairwise comparisons are shown. Transcription factor names are given above each plot.

Supplemental Figure 13. **Differentially regulated gene p-value Q-Q plots.** A Q-Q plot of observed versus expected p-values (-log10 transformed) of differentially regulated genes. Differentially regulated gene comparisons (using the DESeq program) of wild-type HepG2 RNA-seq biological replicate data (in black), as well as wild-type HepG2 RNA-seq data versus CRISPR-modified HepG2 cells tagging ATF1 (in blue) or tagging RAD21 (in gray) are given. All comparisons exhibit nearly identical patterns and an overall paucity of significant differentially regulated genes.

Supplemental Figure 14. **RNA-seq read counts across DNA-binding protein mRNA transcripts.** RNA-derived read counts across RAD21 (A) and ATF1 (B) transcripts are displayed. Reads mapping to epitope tag sequences are highlighted in red, suggesting a substantial proportion of RNA molecules are tagged in polyclonal HepG2 cell populations.

Supplemental Figure 15. **Sequence integrity across 5-prime homologous recombination sites on mRNA.** RNA-derived reads spanning 5-prime homologous recombination sites are displayed for RAD21 (A) and ATF1 (B) epitope tagged cells. At each endogenous nucleotide location upstream of the recombined pFETCh construct, the number of reads matching to the reference genome (gray), or that contain mismatches (blue), insertions (red) and deletions (green) are shown. These data support a low level of mis-targeting at the 3-prime ends of the coding regions of RAD21 and ATF1.

Supplemental Figure 16. **Sequence integrity across 3-prime homologous recombination sites on mRNA.** RNA-derived reads spanning 3-prime homologous recombination sites are displayed for RAD21 (A) and ATF1 (B) epitope tagged cells. At each endogenous nucleotide location downstream of the recombined pFETCh construct, the number of reads matching to the reference genome (gray), or that contain mismatches (blue), insertions (red) and deletions (green) are shown. These data support a low level of mis-targeting at the 3-prime ends of the coding regions and the beginning of the 3-prime untranslated regions of RAD21 and ATF1.

Supplemental Figure 17. ***RAD21*** **Sanger sequencing of PCR amplicons representing homologous recombination site in MCF7 cells.** (A) The endogenous sequence of the 3-prime end of *RAD21* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram traces of PCR amplicons spanning 3-prime homologous recombination sites. Two replicate transfection experiments are shown. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graphs below traces.

Supplemental Figure 18. MCF7 transcription factor motifs. Enriched binding motifs for all RAD21 MCF7 experiments. The names of RAD21 ChIP-seq experiments are given above each motif. Data is shown for CETCh-seq FLAG antibody biological replicate experiments and standard ChIP-seq experiments in wild-type MCF7 cells using transcription factor antibodies.

Supplemental Figure 19. MCF7 RAD21 FLAG tag ChIP-seq correlations. (A) Rank correlations of normalized sequence read counts between standard ChIP-seq replicates (WT replicate 1 and WT replicate 2) using transcription factor antibodies in wild-type MCF7 cells. (B) Rank correlations of normalized sequence read counts between standard ChIP-seq data (WT) using transcription factor antibodies in wild-type MCF7 cells and CETCh-seq experiments (FLAG) using FLAG antibodies in CRISPR-modified MCF7 cells. (C) Rank correlations of normalized sequence read counts between CETCh-seq technical replicates (FLAG rep 1 and FLAG rep 2). In (A) and (C), rank correlations are displayed. In (B), average rank correlations for all FLAG and wild-type ChIP-seq replicate pairwise comparisons are given.

Supplemental Figure 20. MCF7 ChIP-seq site correlations with FLAG tag biological replicate. (A) Rank correlations of normalized sequence read counts between a second MCF7 CETCh-seq biological replicate (FLAG Biological rep 2) and standard ChIP-seq data (WT) using transcription factor antibodies in wild-type MCF7 cells. (B) Rank correlations of normalized sequence read counts between a CETCh-seq biological replicates (FLAG Biological rep 1 and FLAG Biological rep 2) are given. The average rank correlations for all FLAG and wild-type ChIP-seq replicate pairwise comparisons are displayed.

Supplemental Figure 21. MCF7 correlations with CRISPR-modified ChIP-seq data. (A) Rank correlations of normalized sequence read counts between ChIP-seq experiments in CRISPR-modified MCF7 cells using transcription factor antibodies (Modified cells) and ChIP-seq data in wild-type MCF7 cells using transcription factor antibodies (WT). (B) Rank correlations of normalized sequence read counts between ChIP-seq experiments in CRISPR-modified MCF7 cells using transcription factor antibodies (Modified cells) and CETCh-seq experiments (FLAG) using FLAG antibody in CRISPR-modified cells. Average rank correlations for all FLAG and wild-type ChIP-seq replicate pairwise comparisons in CRISPR-modified cells are given.

Supplemental Figure 22. *GABPA* Sanger sequencing of PCR amplicons representing homologous recombination site in mouse embryonic stem cells. (A) The endogenous sequence of the 3-prime end of *GABPA* targeted for pFETCh homologous recombination (blue area). (B) Schematic of the homologous recombination pFETCh construct. Red arrows point to primers used for amplification of 3-prime homologous recombination sites. (C) Sanger sequencing electropherogram traces of PCR amplicons spanning 3-prime homologous recombination sites. The 3-prime end of the pFETCh construct is shown at the left, while the endogenous 3-prime un-translated region (UTR) is shaded in gray. Quality scores indicating confidence at each base position are indicated by bar graphs below traces.

Supplemental Figure 23. GABPA CETCh-seq experimentation summary in mouse embryonic stem cells. (A) The canonical GABPA binding motif was identified and enriched at CETCh-seq binding sites. In (B) and (C), DNA-binding protein read enrichment tracks for GABPA CETCh-seq experiments on the UCSC Genome Browser are shown at distinct genetic loci. In (B) CETCh-seq data at two large loci are shown, while in (C) data for two smaller genome windows sizes are displayed.

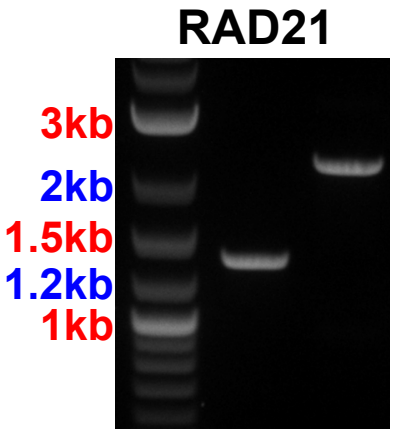
Transcription factor	HepG2 RPKM	MCF7 RPKM
RAD21	40.09	37.55
CREB1	12.84	15.41
NR1H2	7.59	N/A
ATF1	5.99	5.53
GABPA	2.42	N/A

Supplemental Table 1

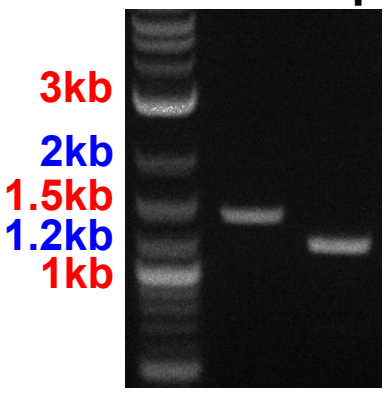
A

TF Name	5' Band Size	3' Band Size
RAD21	1363	2278
CREB1	1428	1190
GABPA	1676	1500
ATF1	1626	1433

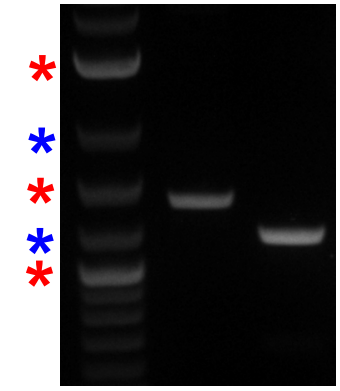
B



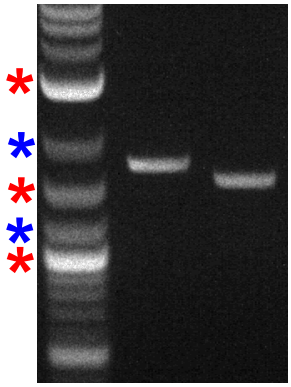
C



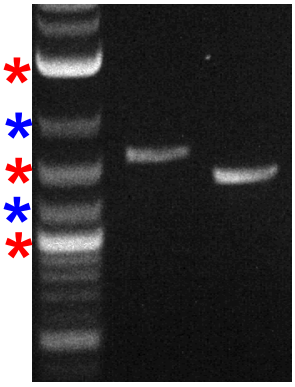
D



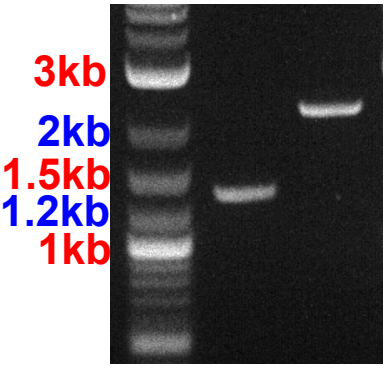
E



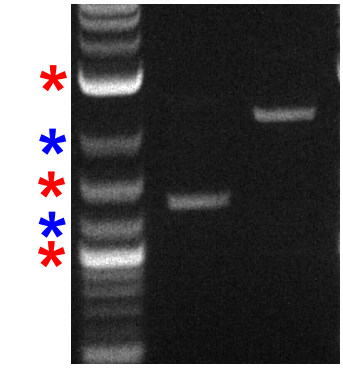
F



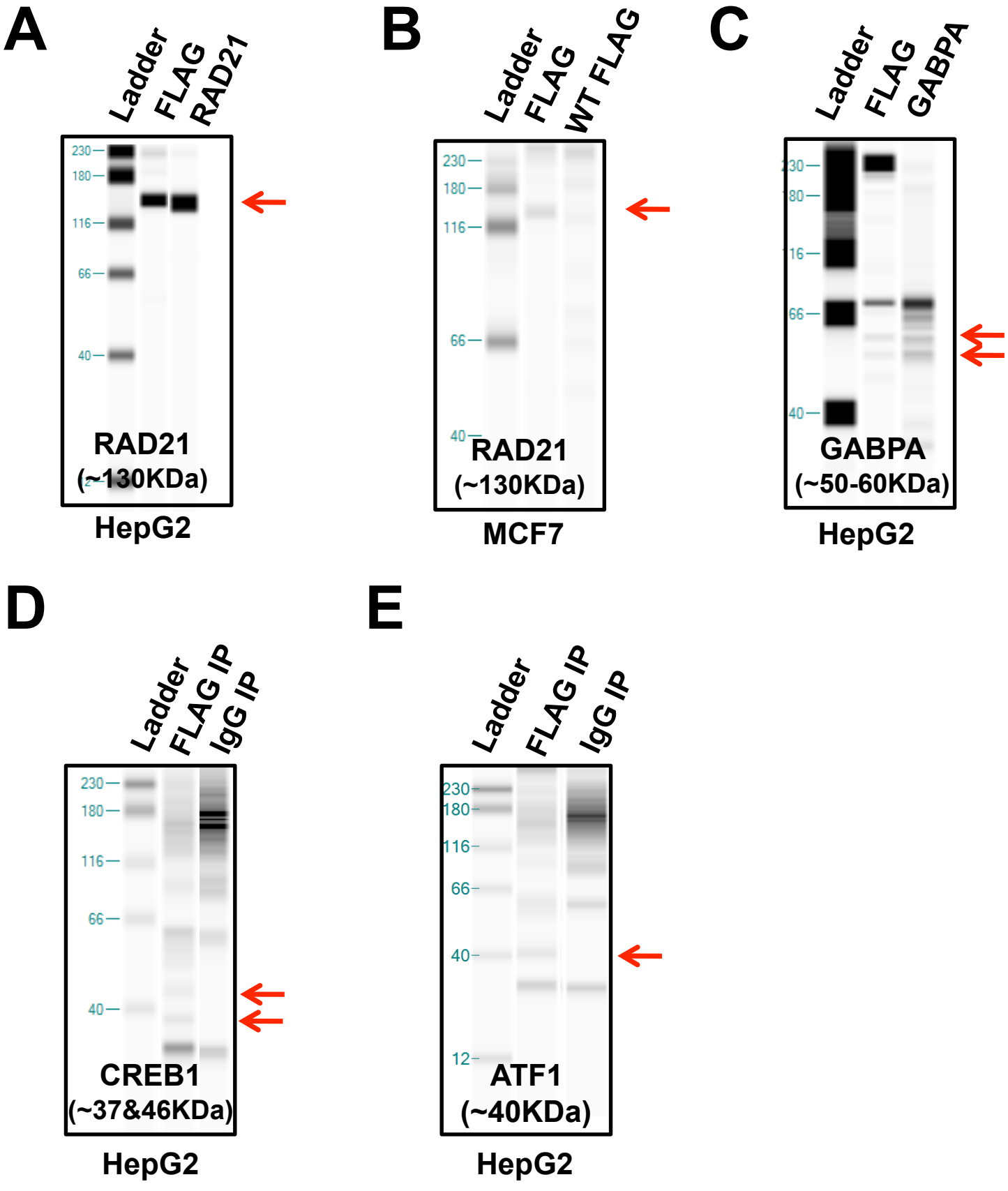
G



H

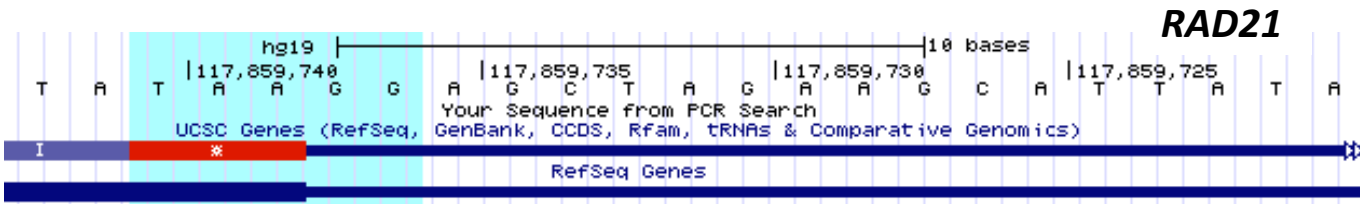


Supplemental Figure 1

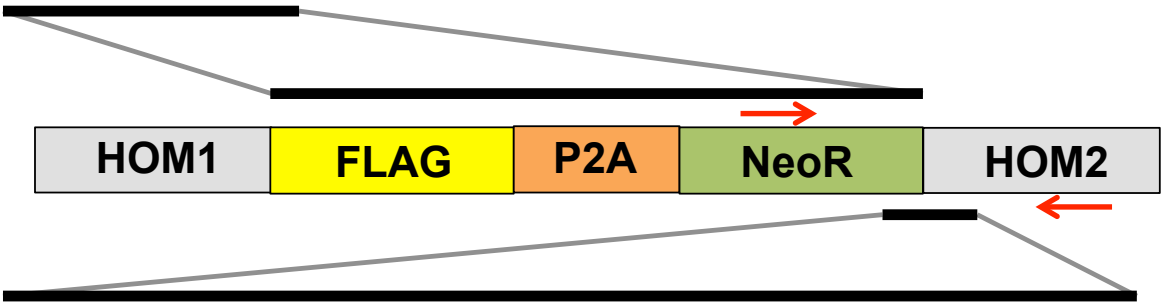


Supplemental Figure 2

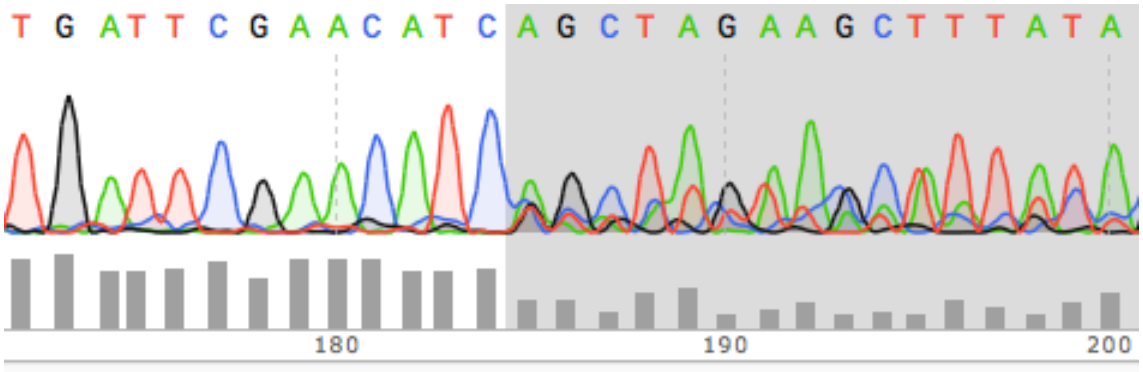
A



B

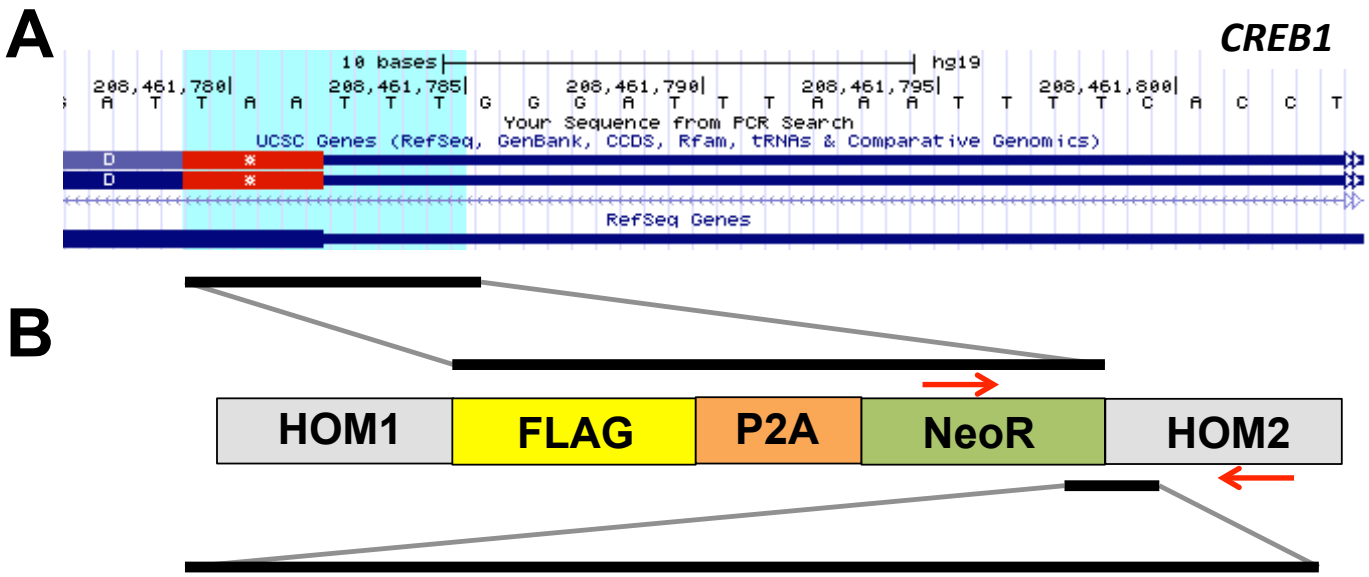


C



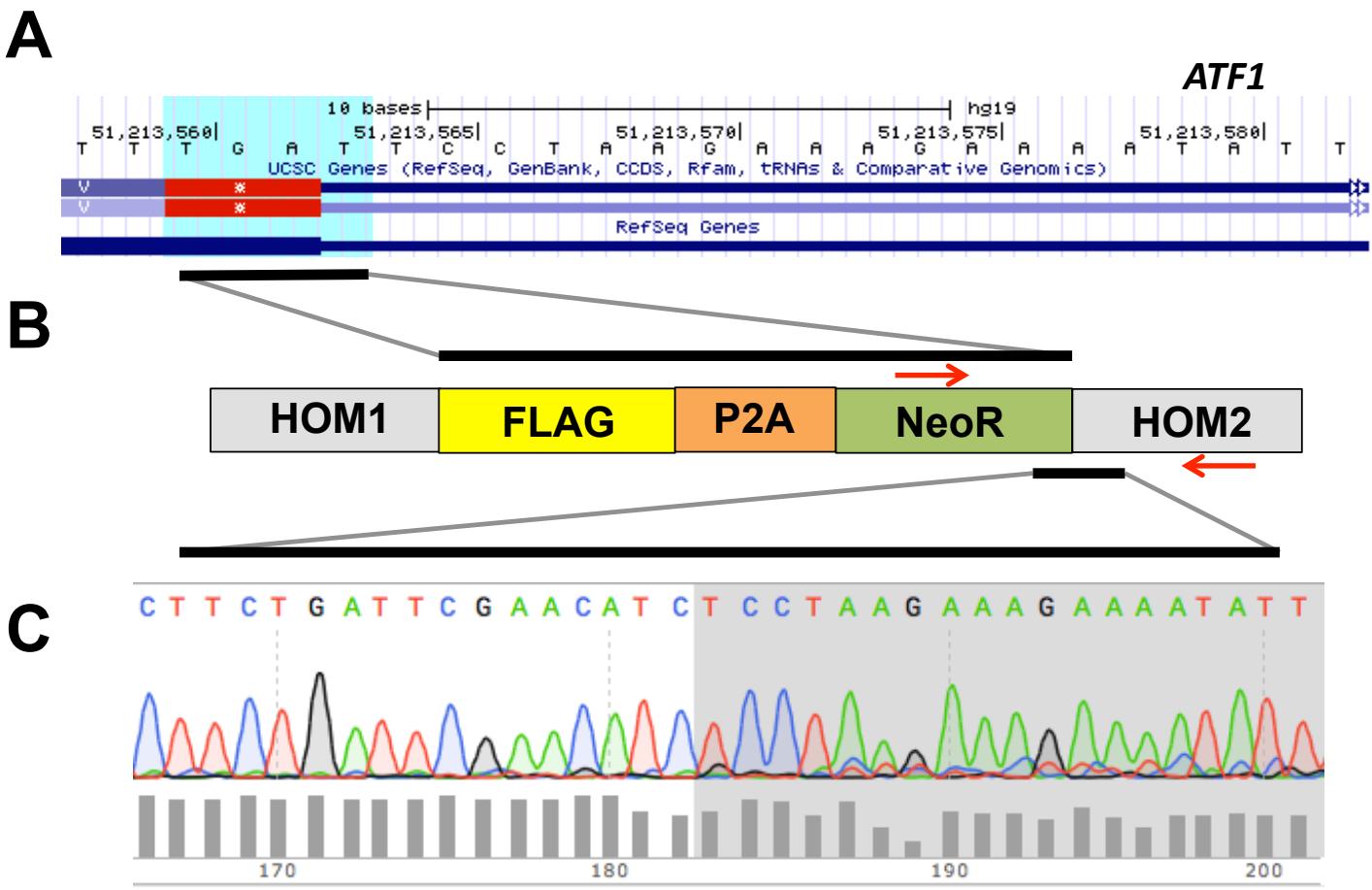
3' End of pFETCh construct Endogenous 3' UTR

Supplemental Figure 3

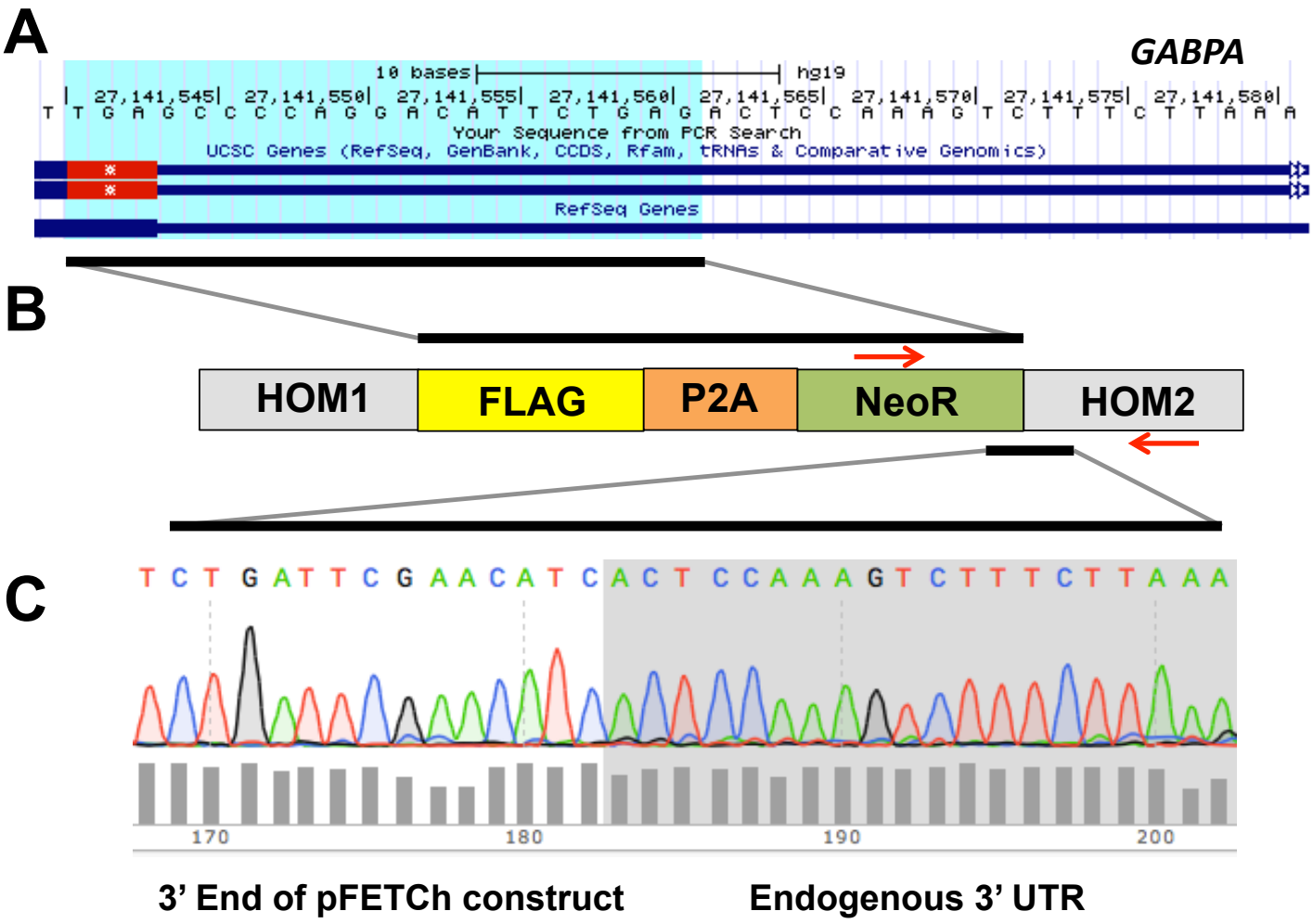


3' End of pFETCh construct **Endogenous 3' UTR**

Supplemental Figure 4

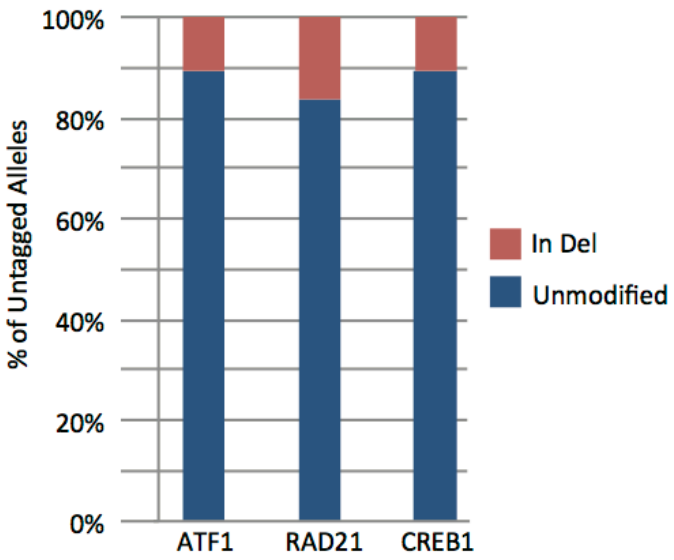


Supplemental Figure 5

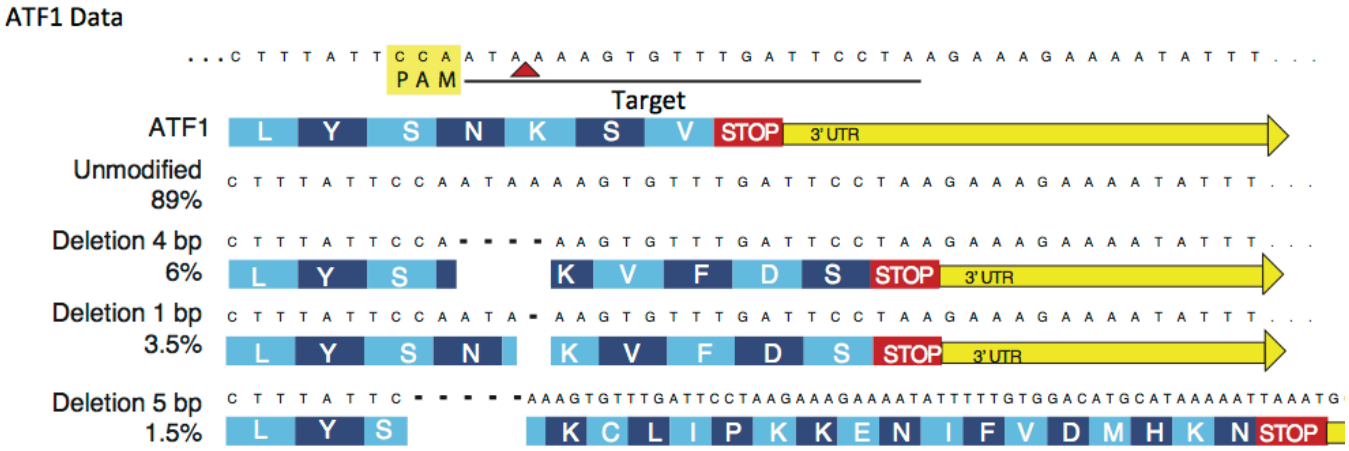


Supplemental Figure 6

A

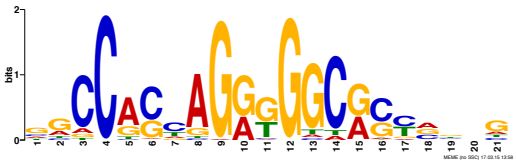


B

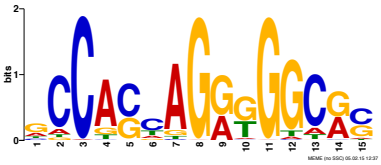


Supplemental Figure 7

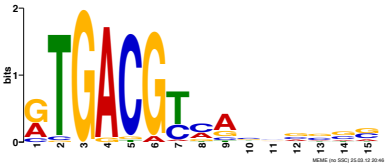
RAD21 WT



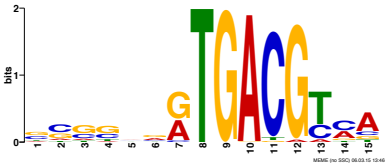
RAD21 Flag



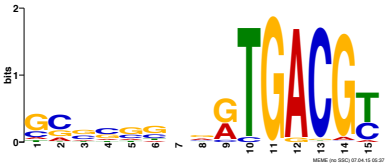
CREB1 WT



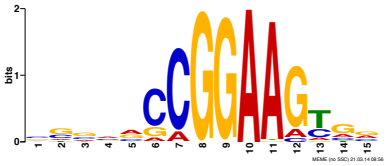
CREB1 Flag rep1



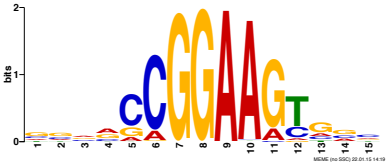
CREB1 Flag rep2



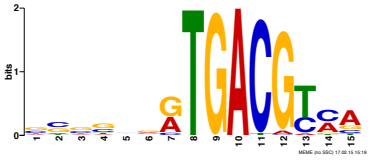
GABPA WT



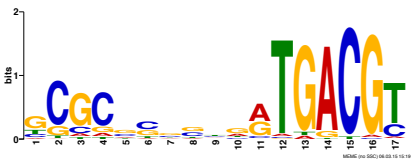
GABPA Flag



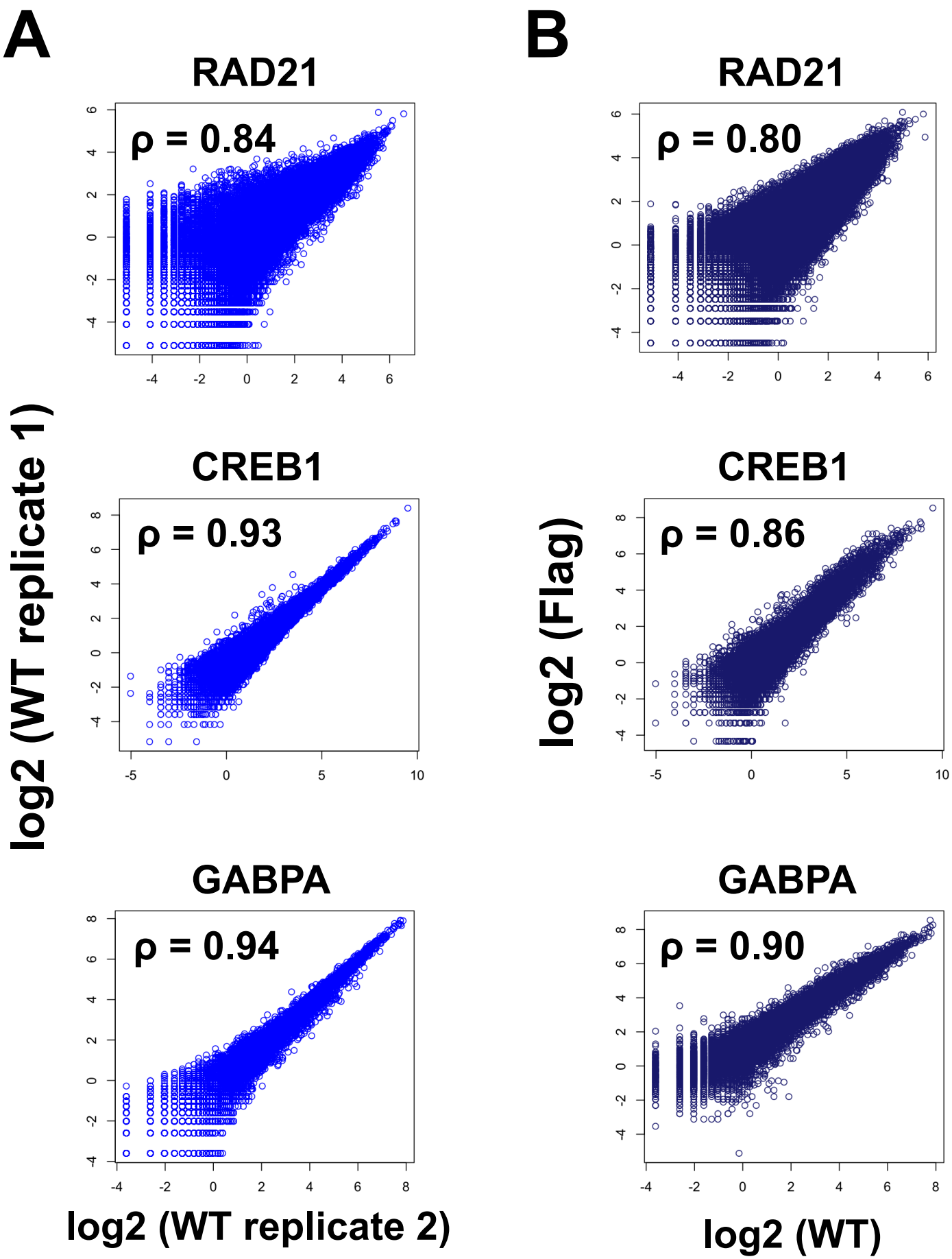
ATF1 Flag rep1



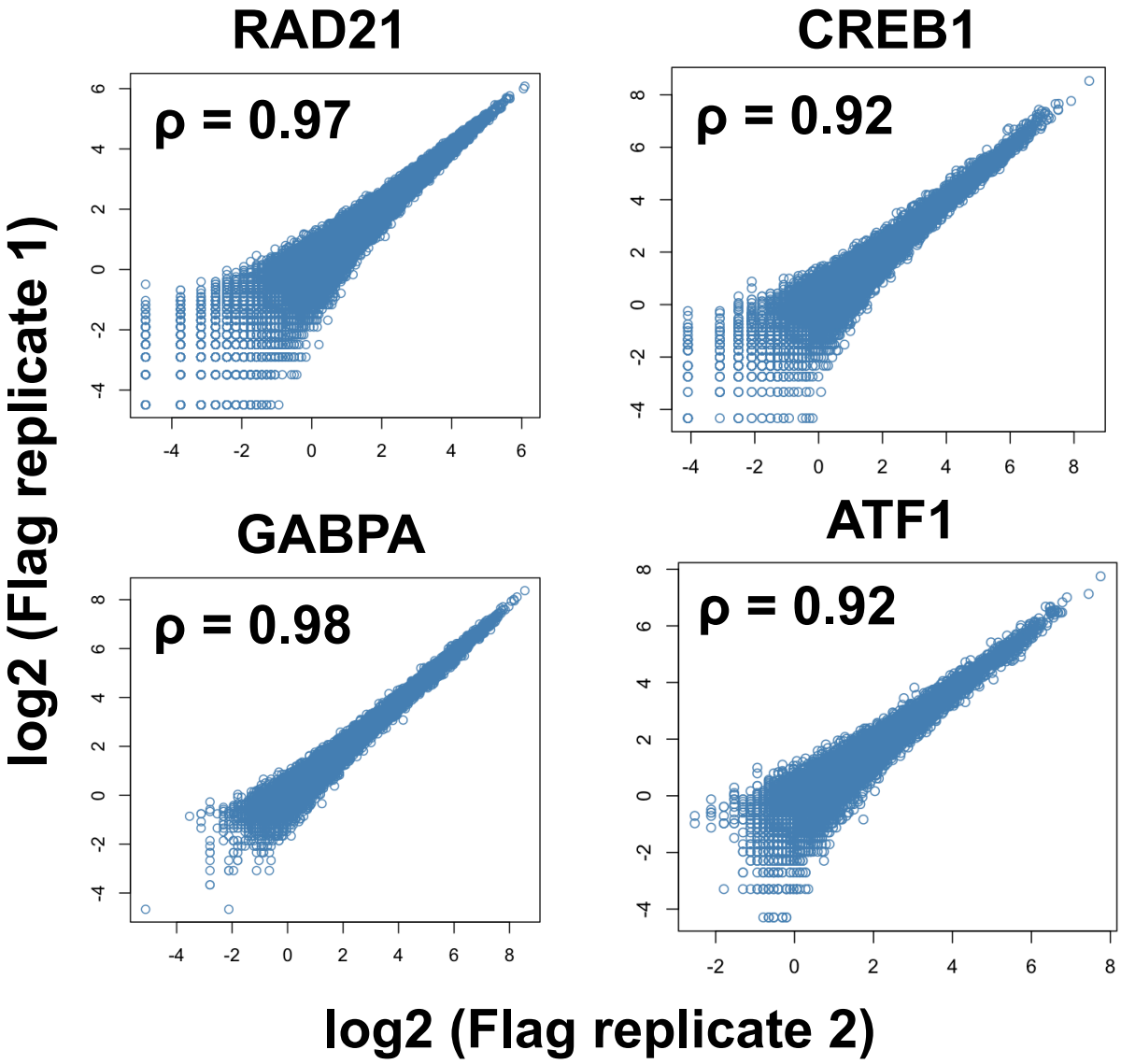
ATF1 Flag rep2



Supplemental Figure 8

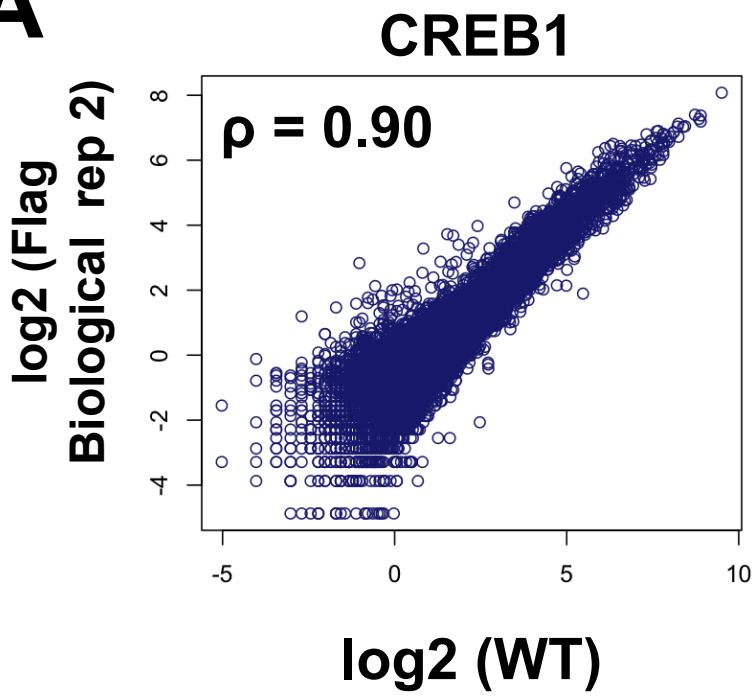


Supplemental Figure 9

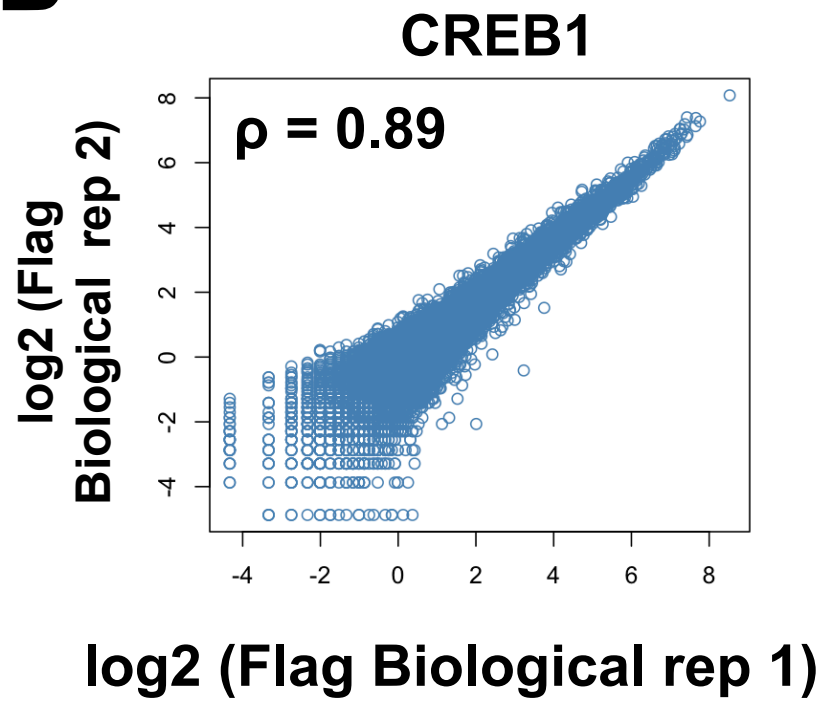


Supplemental Figure 10

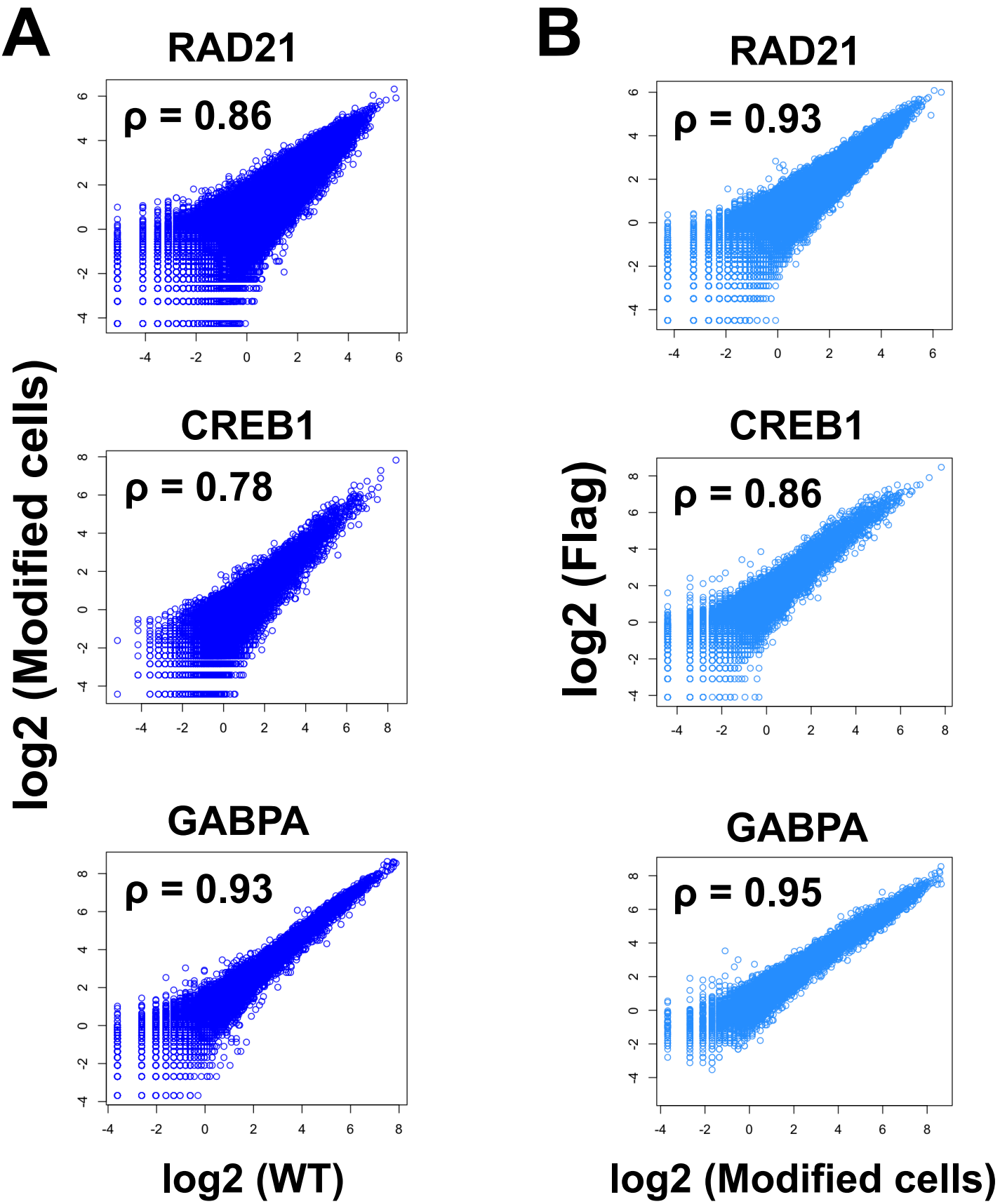
A



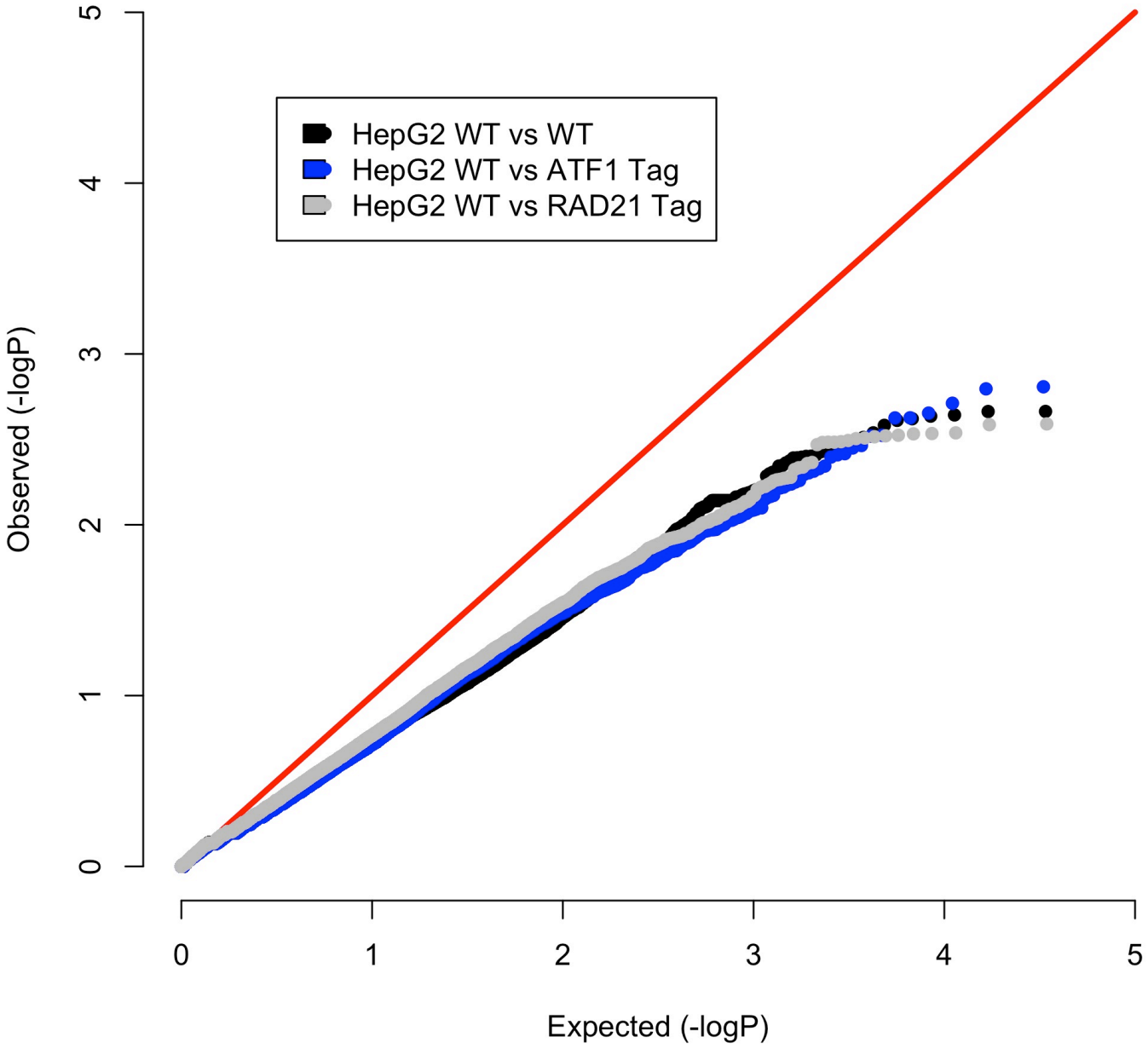
B



Supplemental Figure 11

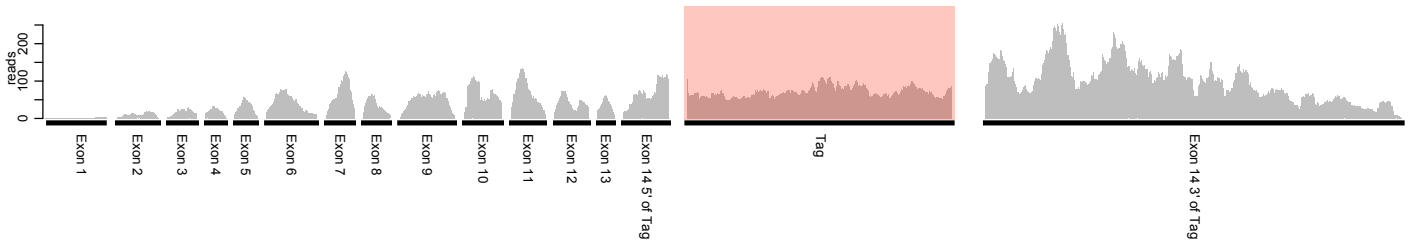


Supplemental Figure 12

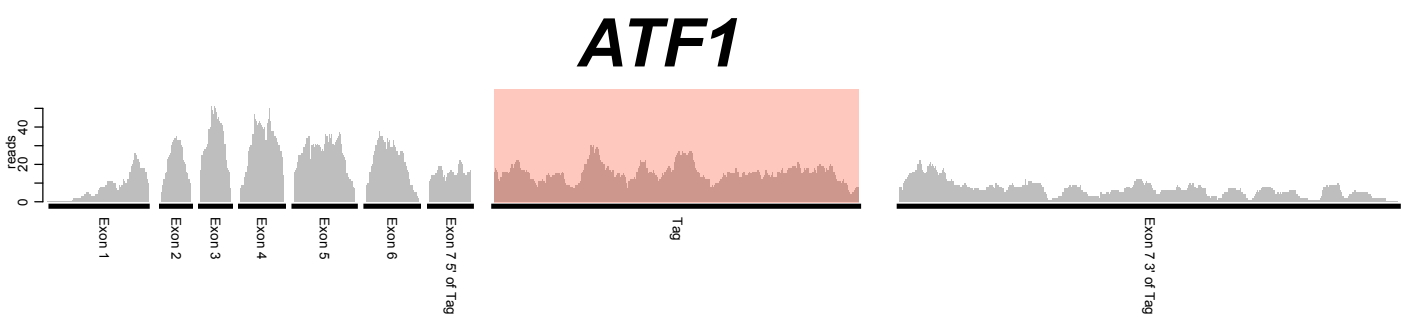


Supplemental Figure 13

A

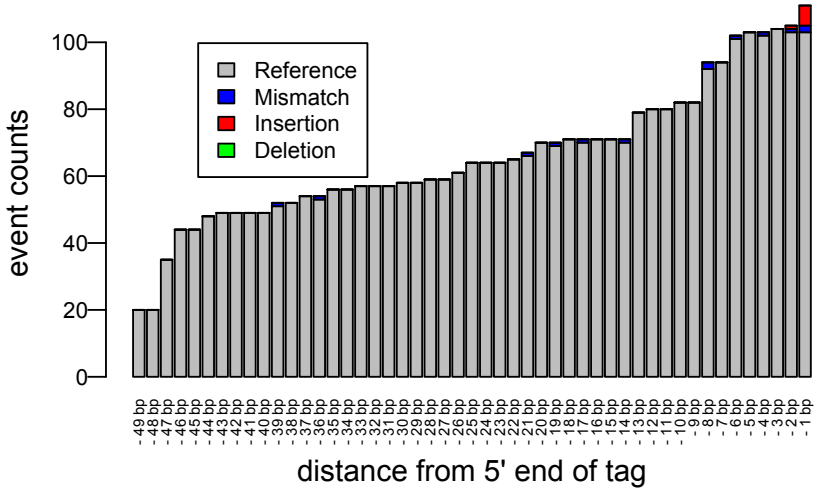


B

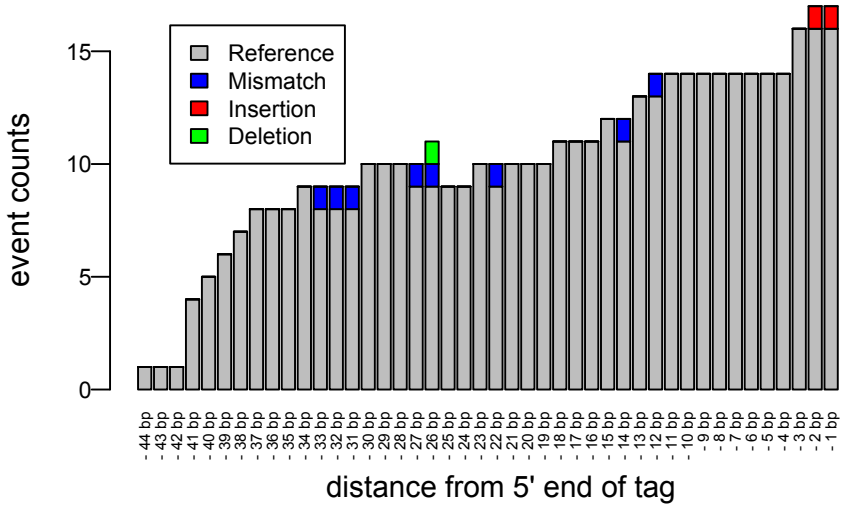


Supplemental Figure 14

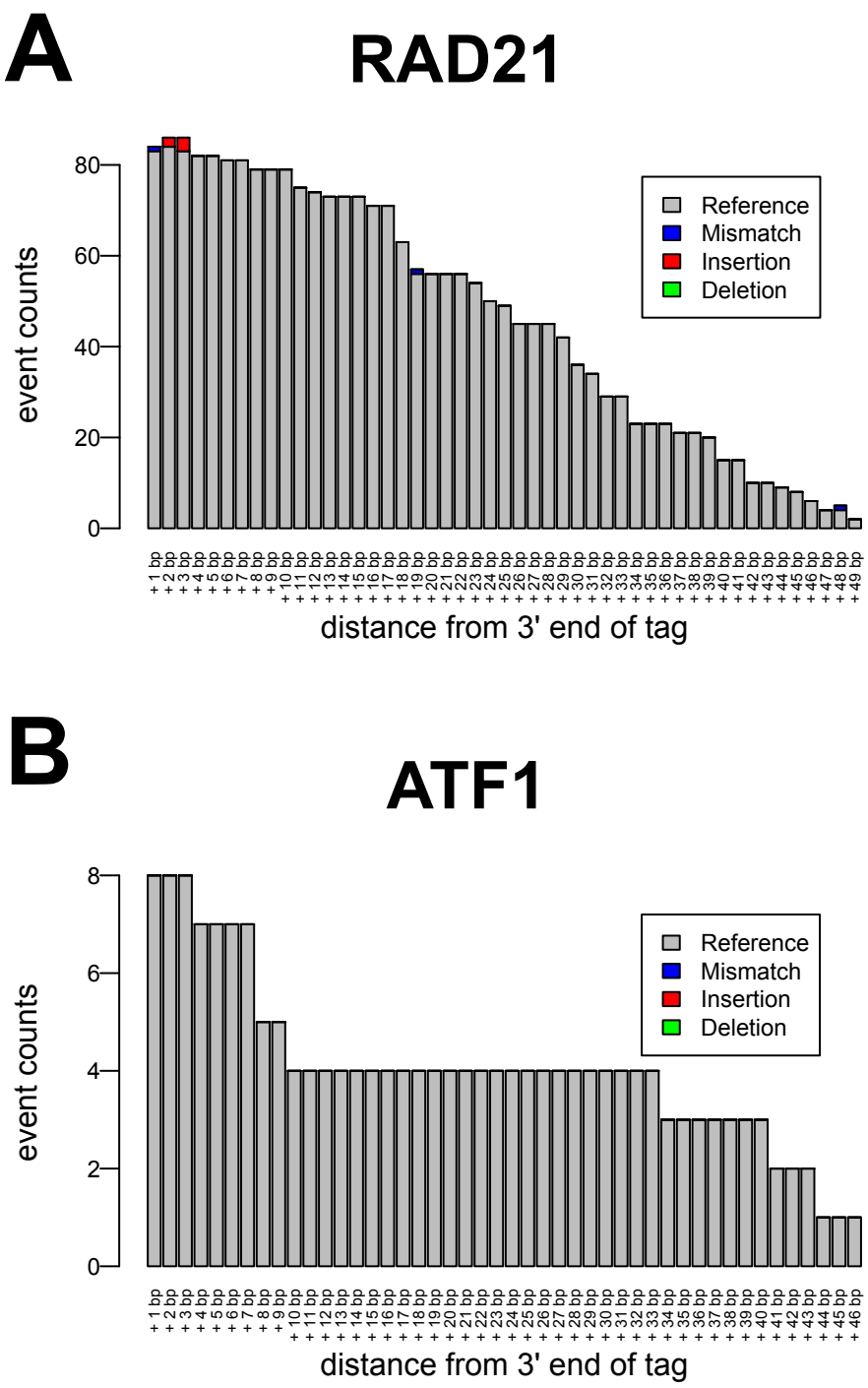
A



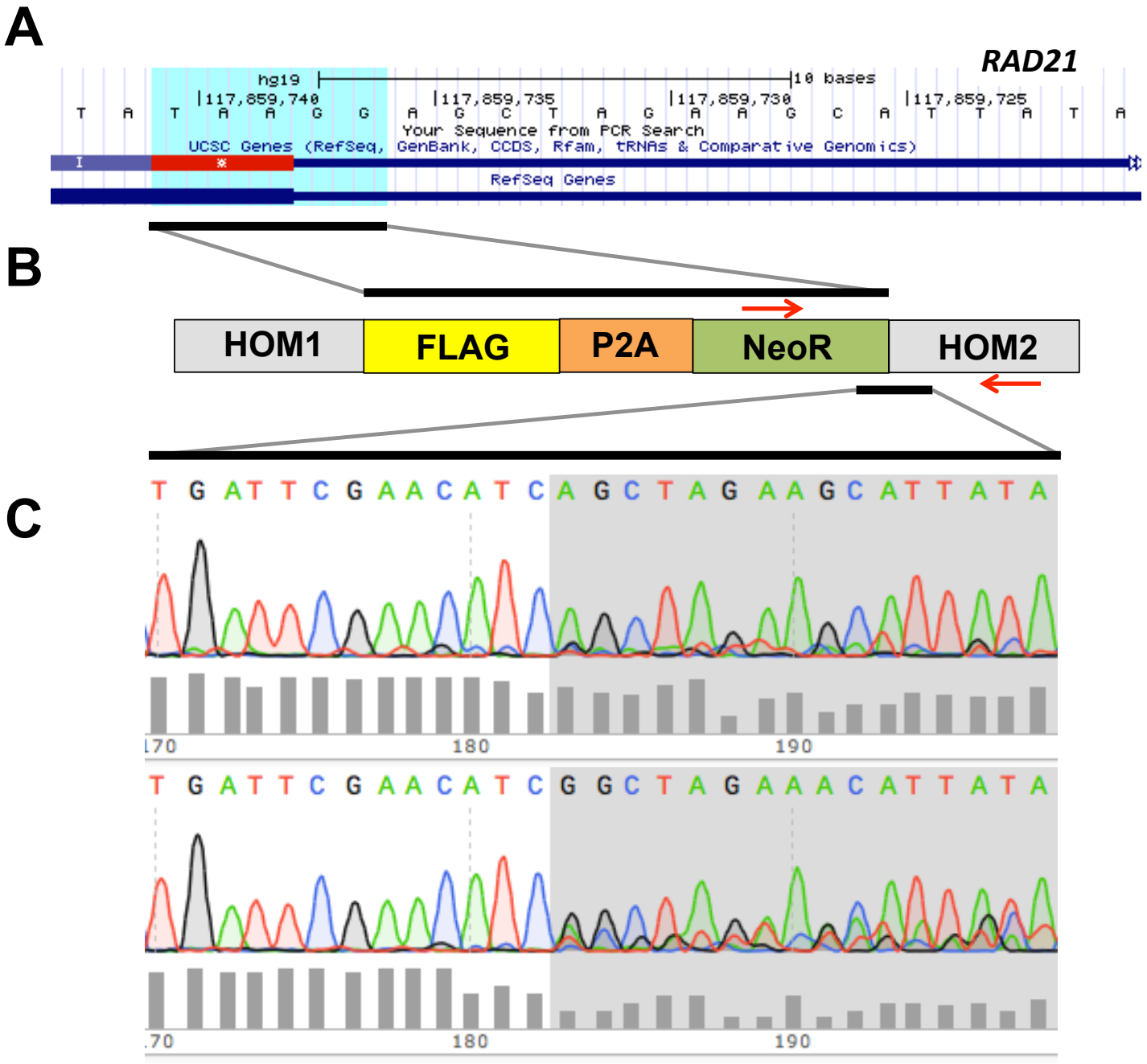
B



Supplemental Figure 15

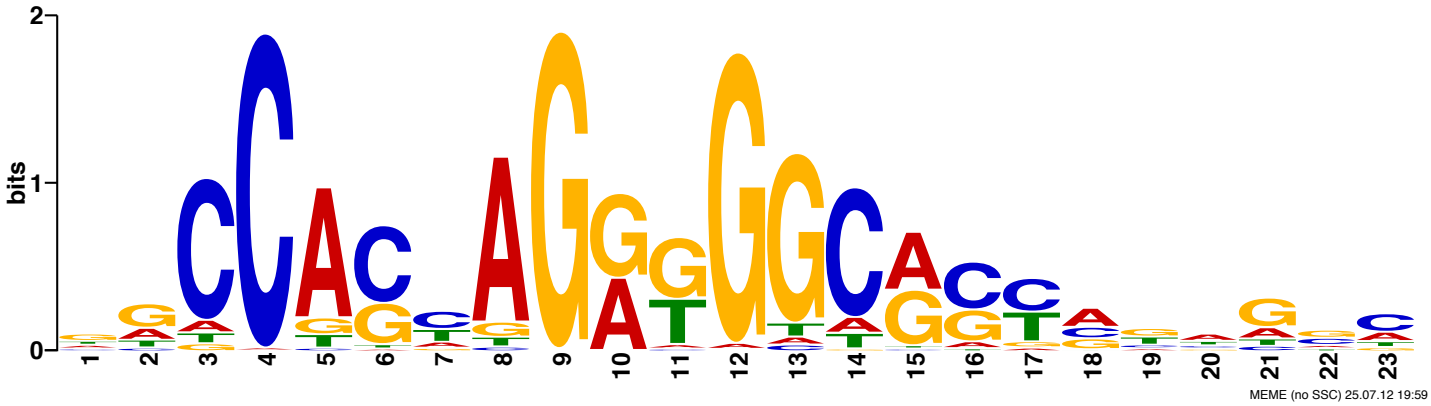


Supplemental Figure 16

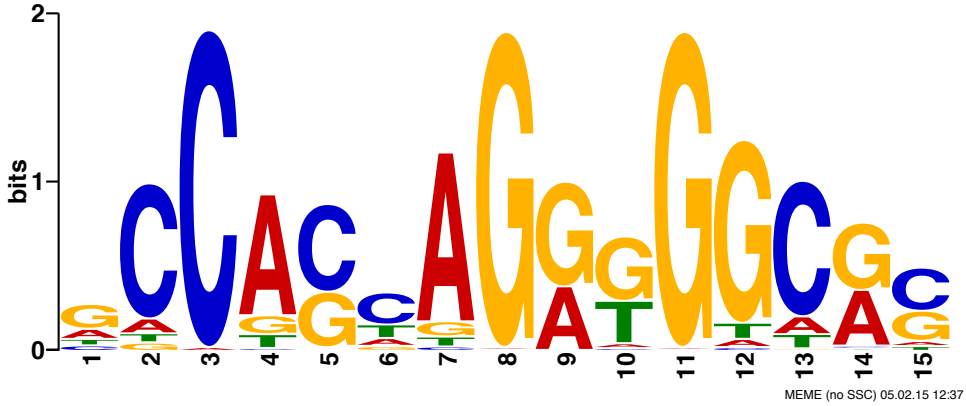


Supplemental Figure 17

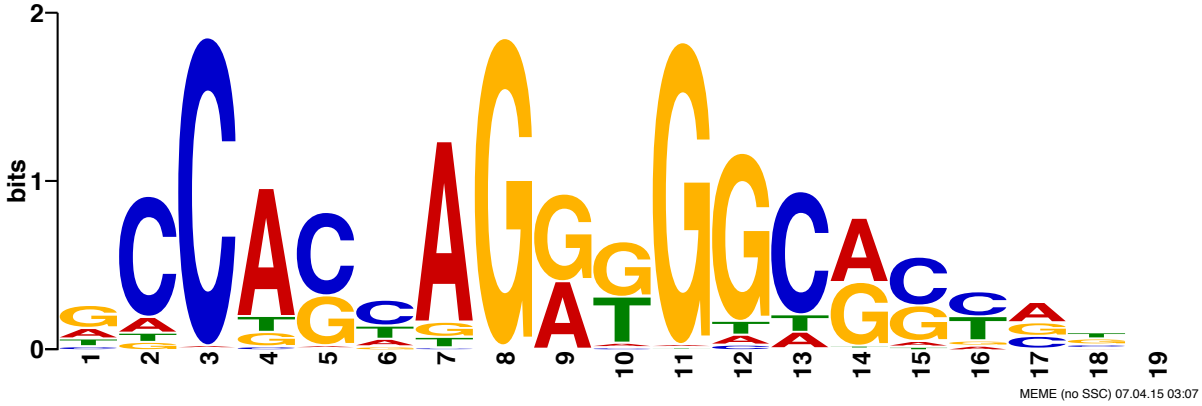
RAD21 WT



RAD21 Flag biological replicate 1

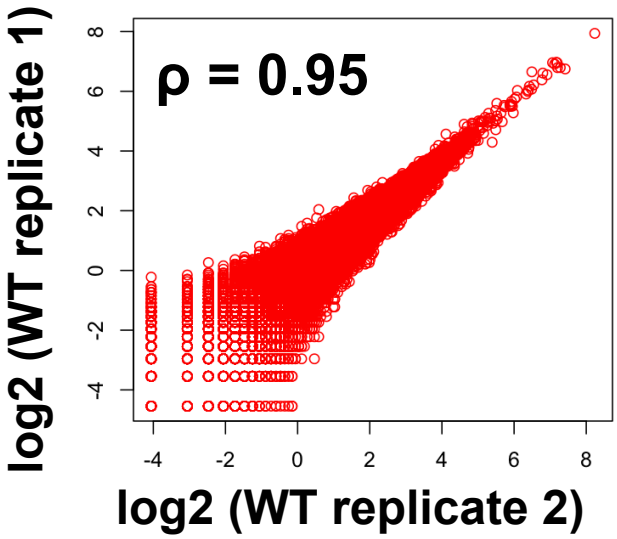


RAD21 Flag biological replicate 2

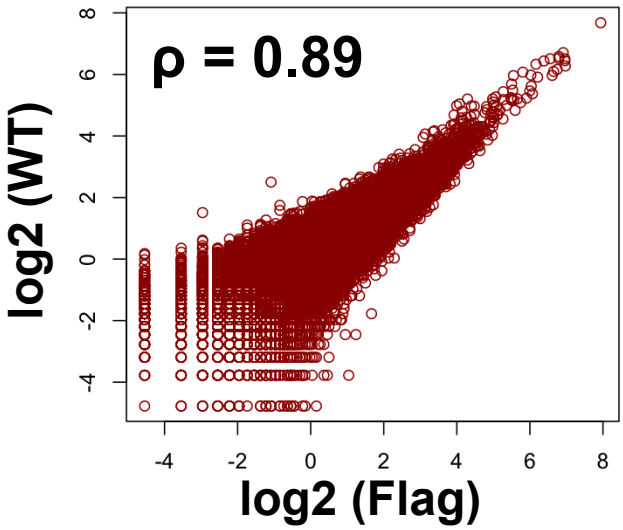


Supplemental Figure 18

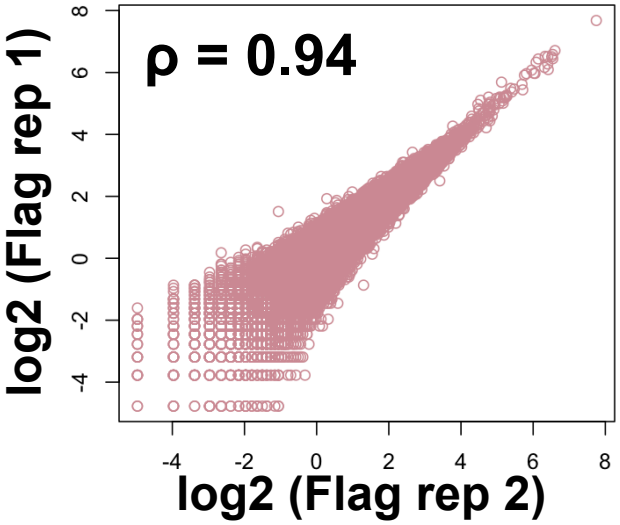
A



B

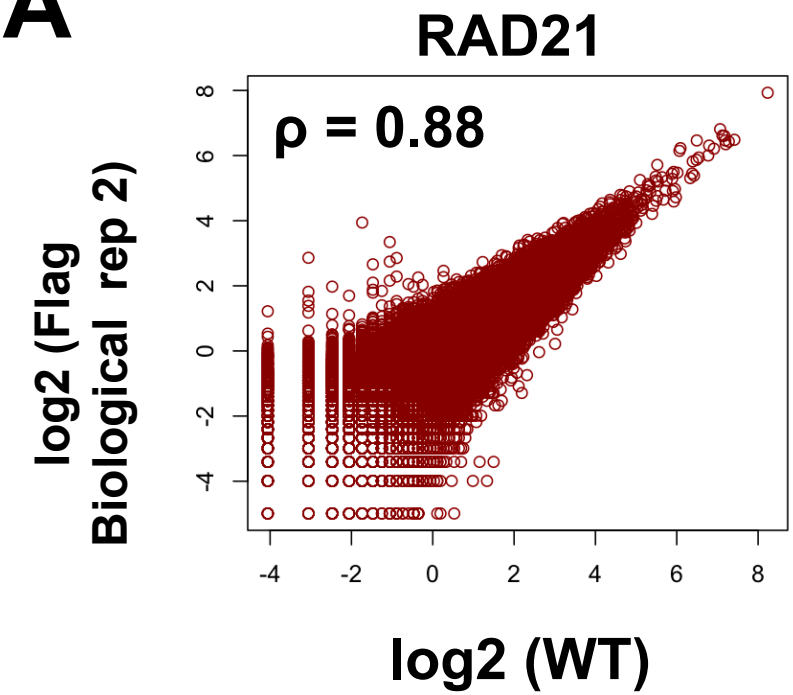


C

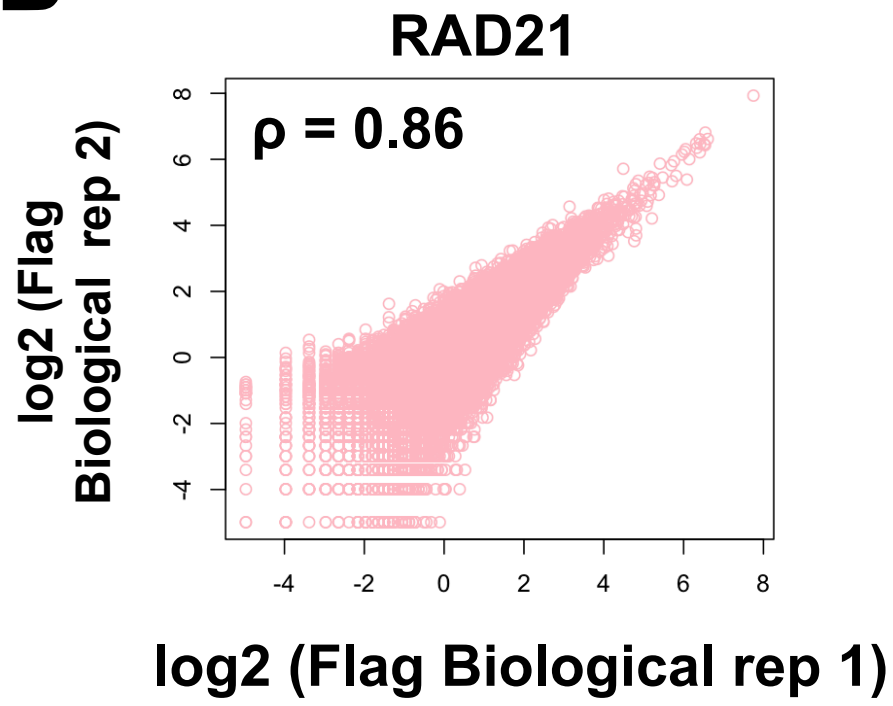


Supplemental Figure 19

A

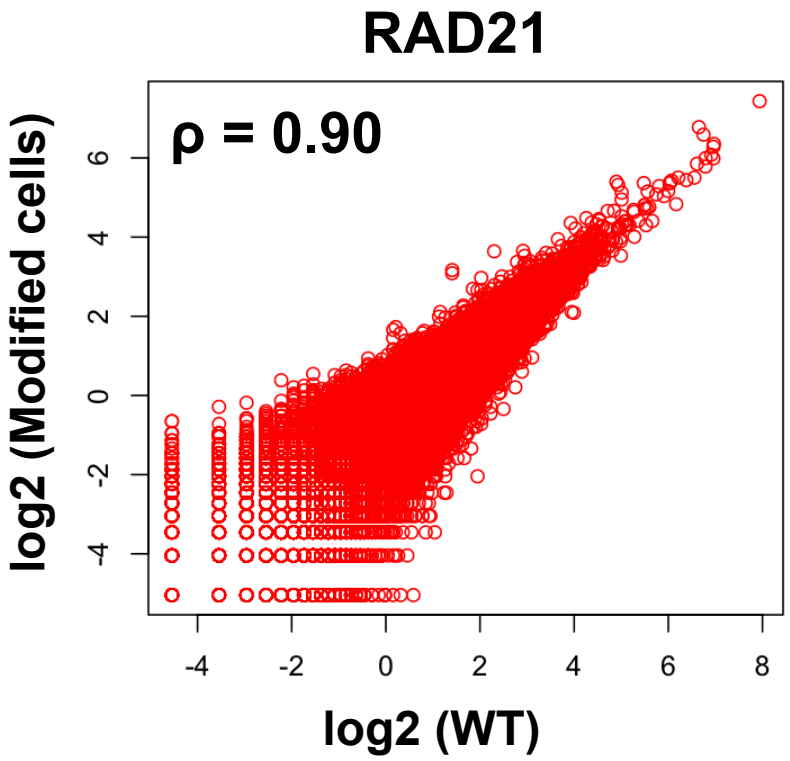


B

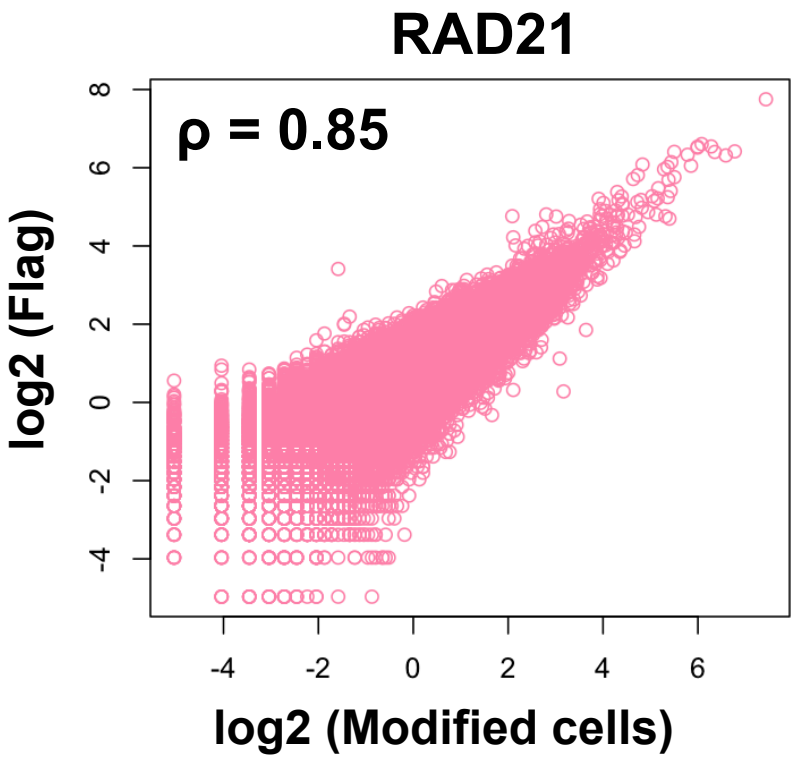


Supplemental Figure 20

A

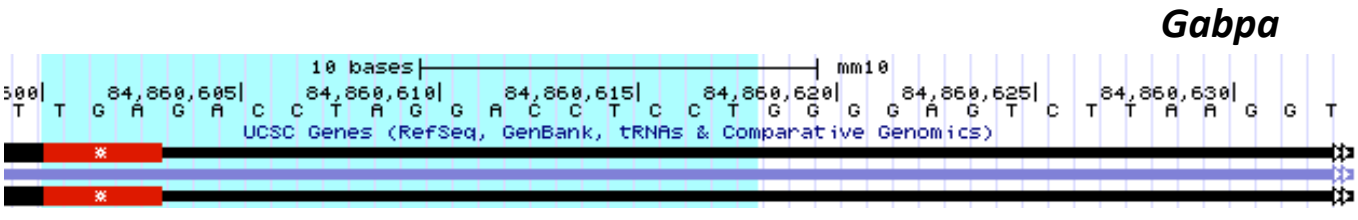


B

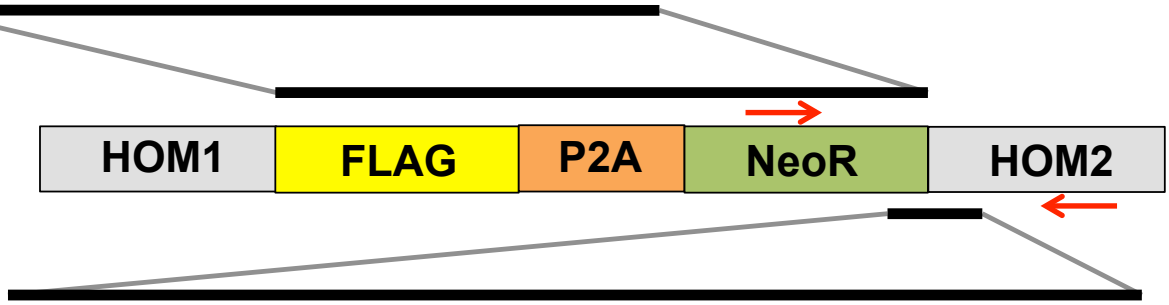


Supplemental Figure 21

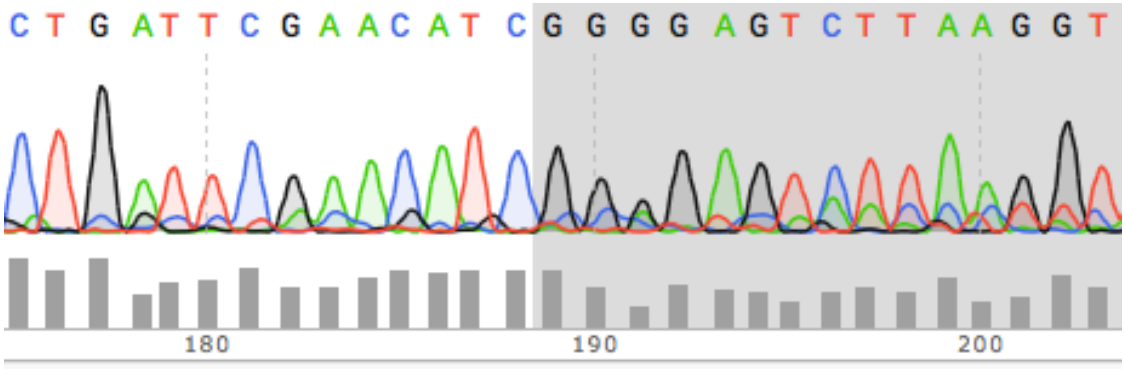
A



B

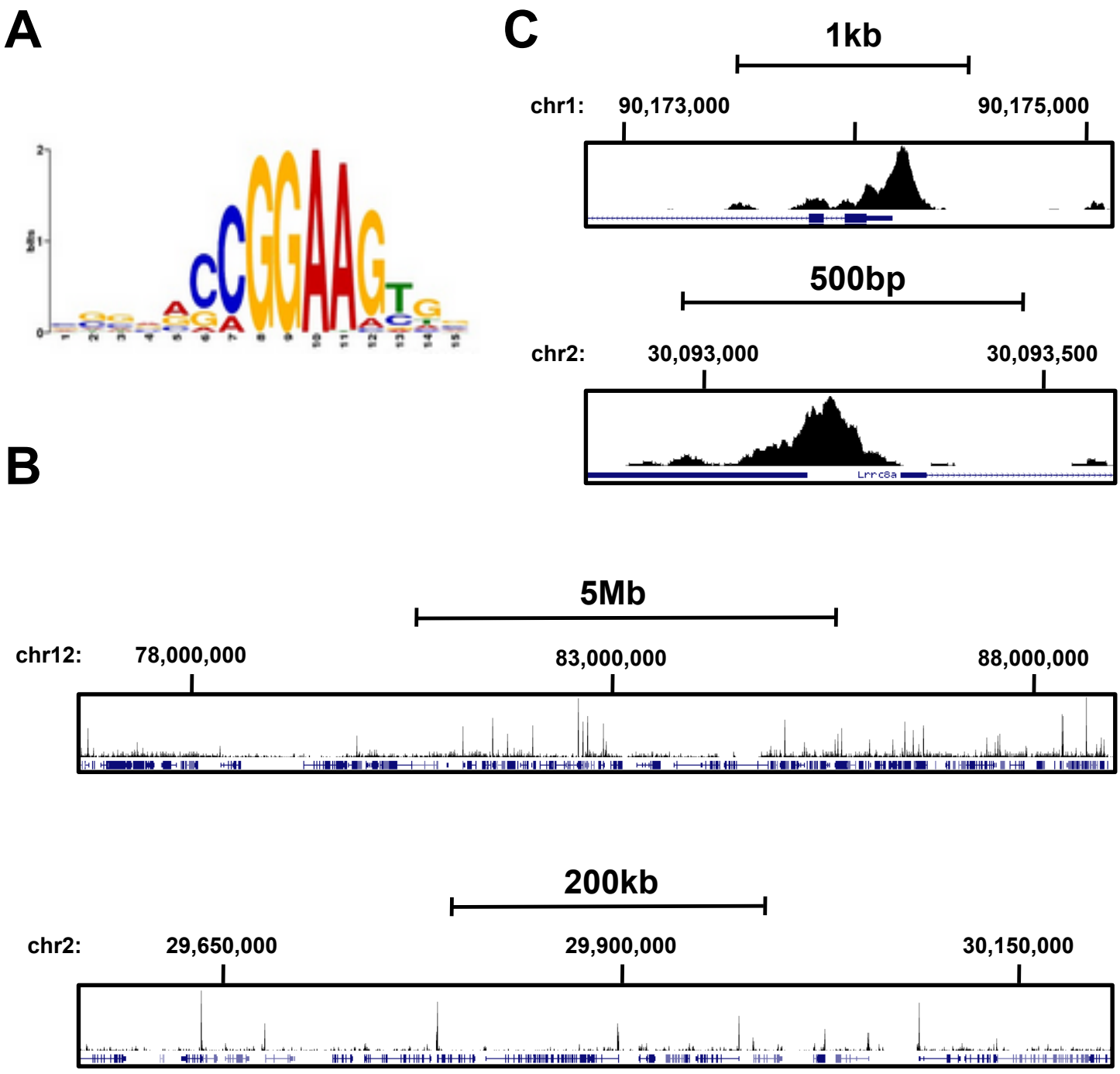


C



3' End of pFETCh construct Endogenous 3' UTR

Supplemental Figure 22



Supplemental Figure 23