**FigS1**

Based on pipeline set up in <https://github.com/CalabreseLab/Airn_Xist_manuscript/tree/main/RIPanalysis> we got the RBP\_fc\_rpm\_2igg\_2reps files, which identify the peak locations that have rpm >2xIGG in >=2 replicates. Then IGG rpm are subtracted from each RBP to get normalized rpm: RBP\_less\_igg.

Peaks were ranked and only the top 1000 ranked peaks were kept. Generate bedfiles for the locations of the top 1000 peaks. Code: top1kbedfile\_gen.R

Manually rename: u2af65\_fc\_rpm\_2igg\_top1k.bed to u2af65\_fc\_rpm\_2igg\_2reps\_top1k.bed and nxf1\_fc\_rpm\_2igg\_top1k.bed to nxf1\_fc\_rpm\_2igg\_2reps\_top1k.bed to be consistent with other RBPs.

Get sequences using bedtools getfasta from mm10 genome (GRCm38.primary\_assembly.genome.fa). These sequence files are named as: RBP\_fc\_rpm\_2igg\_2reps\_top1k.fa. Code: run\_getseq.sh and getseq.sh

Generate control sequences with the same number and length of the top 1000 ranked peaks but randomized with the single nucleotide bias of the mm10 primary assembly (counting the percentages of A,T,G,C across the mm10 genome). These control sequence files are named: control\_RBP\_fc\_rpm\_2igg\_2reps\_top1k.fa. Code: control\_seq\_gen.py

Input the top 1000 ranked peaks with its corresponding control sequences to streme (meme module) to get top10 motifs for each RBP. Code: top1kstreme.sh and run\_top1kstreme.sh

The output streme\_RBP/streme.html are the results. Renamed the streme.html to the corresponding RBP name and save into top1kmotif\_html folder.