**DATA ACCESS**

Sequence data generated for this paper can be found at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110681>

Coordinate files used to generate the figures for this paper can be found at:

<https://github.com/CEGRcode/ChIP-exo-5.0>

All data was deduplicated

Analysis was performed on the GUI **ScriptManager v.010**, which is available for download at:

<https://github.com/CEGRcode/scriptmanager>

**Generation of coordinate files**

* TAB files were generated using **BAM to scIDX** from ScriptManager v0.10.
  + Read 1
* Peak pair files (.gff) and meme.txt (position weight matrix of motif) were generated from the TAB files using the shell script **standard\_XO\_pipeline.sh**
  + Genetrack.py
    - Smooth=5, exclusion=10, filter=1
  + Cwpair\_gff.py
    - Up=0, down=80
  + meme -dna -minw 10 -maxw 20 -nmotifs 3 -time 600 -mod zoops -revcomp -oc <output folder name> <input fasta file>
    - expansion=80, topsite=500
  + fimo --oc <output folder name> --verbosity 1 --thresh 1.0E-4 <input MEME file> <input fasta file>
* The ChIP-exo 5.0 dataset of each protein was mapped to its cognate motif output from FIMO using **Tag Pileup** from ScriptManager v0.10.
  + Read 1; Combined Strands, 0 bp tag shift, 1 bp bin size, set tags to be equal; sliding window 3
  + Sorts were performed on the central 200 bp around the motif using **Sort BED by CDT** from ScriptManager v0.10.
  + An estimate of the number of bound motifs was made by visual inspection, then the number of sites was rounded; expect for Reb1 where the previously published list of primary bound sites was used (Rhee HS. Pugh BF. (2011) Cell). All datasets were linked to the ChIP-exo 5.0 sort of each protein.

**Heatmaps and Composite Plots of ChIP-exo**

* Heatmaps and composite plots were generated using **Tag Pileup** from ScriptManager v0.10.
  + Read 1; Separate Strands, 0 bp tag shift, 1 bp bin size, set tags to be equal; sliding window 3
  + Sorts were performed on the central 200 bp around the motif of the **ChIP-exo 5.0 dataset** for each protein using **Sort BED by CDT** from ScriptManager v0.10.
  + CDT files visualized in Java Treeview
  + Composites were constructed in Prism 7