

Table S1: Details of the *in vivo* dataset. The table shows the class index from the DeepSea dataset (zero-based index), the transcription factor name, the cell type, the number of training sequences, and the original ENCODE file name for each transcription factor in the *in vivo* dataset.

| Index | Name   | Cell type | # of sequences | ENCODE file name  |
|-------|--------|-----------|----------------|---|
| 592   | ARID3A | K562      | 27652          | wgEncode.AwgTfbsSydhK562Arid3asc8821IggrabUniPk.narrowPeak.gz |
| 602   | CEBPB  | K562      | 85354          | wgEncode.AwgTfbsSydhK562CebpbIggrabUniPk.narrowPeak.gz        |
| 344   | FOSL1  | K562      | 19724          | wgEncode.AwgTfbsHaibK562Fosl1sc183V0416101UniPk.narrowPeak.gz |
| 345   | GABP   | K562      | 34194          | wgEncode.AwgTfbsHaibK562GabpV0416101UniPk.narrowPeak.gz       |
| 635   | MAFK   | K562      | 43528          | wgEncode.AwgTfbsSydhK562Mafkab50322IggrabUniPk.narrowPeak.gz  |
| 636   | MAX    | K562      | 87290          | wgEncode.AwgTfbsSydhK562MaxIggrabUniPk.narrowPeak.gz          |
| 349   | MEF2A  | K562      | 9792           | wgEncode.AwgTfbsHaibK562Mef2aV0416101UniPk.narrowPeak.gz      |
| 642   | NFYB   | K562      | 22758          | wgEncode.AwgTfbsSydhK562NfybUniPk.narrowPeak.gz               |
| 359   | SP1    | K562      | 17450          | wgEncode.AwgTfbsHaibK562Sp1Pcr1xUniPk.narrowPeak.gz           |
| 361   | SRF    | K562      | 7528           | wgEncode.AwgTfbsHaibK562SrfV0416101UniPk.narrowPeak.gz        |
| 661   | STAT1  | K562      | 4516           | wgEncode.AwgTfbsSydhK562Stat1Hfg30UniPk.narrowPeak.gz         |
| 369   | YY1    | K562      | 31146          | wgEncode.AwgTfbsHaibK562Yy1V0416101UniPk.narrowPeak.gz        |