Supplemental Material table of contents

Commentary on context-specificity

Much of our analysis was done in a context-agnostic manner, aggregating signals across available experimental systems. As such, our rankings prioritized TR targets with recurrent, rather than context-specific, evidence.

While analyzing context-specific patterns of regulation is desirable as well, this is currently challenging as there are few contexts for which there are sufficient data to apply our aggregation approach. Even when there are multiple data sets for a context (leaving aside the tricky question of how to define “the same context”), these were often generated by the same research group, which weakens their value as independent evidence. Still, it is of interest to get a sense of whether aggregating by context would lead to substantially different results than a generic aggregation. Therefore, in this section we report on a single context with multiple data sets to explore such analyses. We stress that with the data we provide (Supplementary Data S2) others can determine their own definition of a context and perform their own analyses.

In our corpus there are three perturbation and four ChIP-seq experiments for RUNX1 in the context of the K-562 cell line. Much like comparing the similarity of experiments assaying the same versus different TRs, we can contrast experiments for the same TR but in different contexts. The hypothesis is that results from the same context would have greater agreement, pointing to reproducible regulatory profiles. However, it is also expected that considerations such as laboratory protocols or perturbation strategy will contribute significant technical variation.

The K-562 subset of ChIP-seq experiments showed elevated (yet still moderate) similarity to one another relative to other RUNX1 experiments (Supplemental Fig. S22). RUNX1 K-562 ChIP-seq experiment binding scores had a median Pearson’s correlation (r) of 0.36 with one another (median 90/500 overlap) and r = 0.20 (46/500) with non-K-562 RUNX1 experiments, while non-K-562 RUNX1 experiments had r = 0.23 (52/500) with one another. The highest correlation a RUNX1 K-562 experiment had with a RUNX1 non-K-562 experiment involved a fetal liver CD34+ HSPCs experiment (r = 0.36, 147/500). The most similar non-RUNX1 experiments were also carried out using K-562 cells, targeting NEUROD1 and HES1.

All three RUNX1 K-562 perturbation experiments were loss-of-functions, although their similarities aligned with the weak global patterns we observed (Supplemental Fig. S22A). In particular, two of the three experiments showed greater similarity with each other relative to the third, which may be attributable to differences in the experimental designs beyond the perturbation. Finally, the most similar (by p-value overlap) pairing of a K-562 experiment with a non-K-562 context was with a RUNX1 knockdown conducted in an acute myeloid leukemia cell line, suggestive of potential commonalities in RUNX1 regulatory profiles in different leukemic systems. In sum, these comparisons exemplified our conclusions from the full corpus: elevated similarity among experiments targeting the same TR suggests commonalities in targets, but with a degree of noise that motivates aggregation to prioritize the consistent signals.

We next re-calculated the gene rankings, separately aggregating RUNX1 K-562 and non-K-562 experiments. We aimed to gain at least a relativistic sense of the specificity of the RUNX1 K-562 rankings. The K-562 and non-K-562 integrated rankings had a correlation of 0.48, while the perturbation rankings had $r = 0.34$ and $r = 0.55$ for the binding lists. More specifically, cutting the two integrated lists at 1000 revealed 221 common genes (Supplemental Fig S22B). Next, we performed the same comparison, this time comparing the top 1000 genes from the RUNX1 K-562 integrated rankings with the top 1000 genes from each of the human TR integrated rankings. As expected, the ranking generated from all human RUNX1 data (which includes the K-562 experiments) had the greatest overlap (300/1000), while the remaining seven TRs ranged from 55/1000 (MECP2) to 126/1000 (HES1). Thus we observed that the RUNX1 K-562 ranking had some divergence from the RUNX1 non-K-562 ranking, but was still more similar than the rankings generated for the other TRs. This suggests that the K-562 ranking might be recovering at least some context-specific information that isn't obscured by noise due to aggregating fewer data sets.

We now highlight a handful of the identified targets, and again caution against interpreting differences as being purely biological in nature. The top candidate from the RUNX1 K-562 list was *LGALS12* (Supplemental Fig. S22C), which encodes a galectin and whose expression is considered to be enriched in leukemic cell lines by the Human Protein Atlas resource (HPA; Uhlén et al., 2015). This gene was also fairly highly ranked in the non-K562 integrated rankings (607th) and thus also the full RUNX1 integrated ranking (262nd).

Examples of genes that were more specifically prioritized in the K-562 list include *TBX19* (K-562 rank = 58th, non-K-562 rank = 16,900th; Supplemental Fig. S22D), *EXOC3L2* (K-562 rank = 41st, non-K-562 rank = 16,008th), and *ALX4* (K-562 rank = 49th, non-K-562 rank = 15,754th). *CR1* was a RUNX1 curated target much more strongly prioritized in the K-562 analysis (260th) relative to the non-K-562 list (14,060th; Supplemental Fig. S22E), and encouragingly the corresponding low-throughout experiment (a reporter assay) was conducted in an acute myeloid leukemia system (Kim et al., 1999). However, neither of the RUNX1 subset integrated lists outperformed the full RUNX1 integrated ranking in the low-throughput benchmark. Lastly, *TGFBR2* was the gene most prioritized in the non-K-562 ranking (43rd) relative to the K-562 ranking (17,883rd; Supplemental Fig. S22F).

Commentary on aggregate ranking

Our ultimate gene rankings aggregate signals across experiments. Correspondingly, this means that a TR-gene interaction with appreciable signal in only a single experiment is likely to be lower ranked than interactions with signals across experiments. If the aggregation combines multiple contexts, a highly context-specific interaction will receive a worse ranking than one that is common across the contexts. An obvious way to address this is to have sufficient data from one context to do the analysis, as discussed in the previous section. An alternative, suggested by a reviewer, is to use a ranking scheme that prioritizes genes with a single positive finding in any experiment ("winner takes all"), in an attempt to identify context-specific targets.

We first ranked genes *within* experiments by their strength of evidence (binding scores for ChIP-seq, and p-values and absolute fold changes for perturbation). Next, for each TR, we assigned genes to their lowest (best) rank achieved *across* TR-specific experiments. For example, across eight human ASCL1 perturbation experiments, *MYPOP* had a ranking ranging from 2 to 14,383, and thus it was assigned rank 2 in the "minimum ranking". This scheme introduces many ties as each experiment typically had a unique gene at any given rank *i*. Thus, to break ties we applied the same heuristic used on the aggregated perturbation rankings: genes with identical ranks were re-ordered using their average absolute fold change (or average binding score) across TR-specific experiments. The integrated ranking was then calculated from these "minimum rankings" in the same manner as the "aggregate rankings."

In summary, the minimum ranking places greater importance on evidence in individual experiments compared to the aggregate ranking, but still uses information across experiments to resolve ties in a non-arbitrary manner.

We next used the low-throughput resource to identify curated targets that were especially prioritized by the minimum ranking and were performed in comparable biological contexts. A notable example was *G6pc2,* a pancreatic gene with curated mouse NEUROD1 evidence in a pancreatic cell line (Martin et al., 2013) that had a better minimum integrated ranking (4th) than the aggregated ranking (2,623rd). Correspondingly, the NEUROD1 ChIP-seq and perturbation experiments in which *G6pc2* achieved its "best" ranking were both in pancreatic systems. This instance suggests that the minimum ranking scheme has the potential to find relevant context-specific interactions.

However, we regularly found that the biological contexts of the "best" genomic experiments and the low-throughput experiments were not comparable. This does not preclude these genes from being true targets: it is of course possible that a real TR-target interaction could have strong evidence in individual genomic experiments from highly distinct tissues. But we argue that the interpretation of these "mismatched" cases from the minimum ranking (a gene is prioritized in a single perturbation and a single ChIP-seq experiment from distinct contexts) is less intuitive than the aggregate ranking (a gene is prioritized because it generally has evidence across experiments). More concretely, when we applied the low-throughput benchmark to the minimum rankings, they failed to outperform the aggregate rankings in general – 12/16 aggregated integrated rankings outperformed the minimum integrated rankings.

Furthermore, while the minimum ranking may potentially reward true context-specific targets relative to the aggregate rankings, it will also reward outlier behavior that is technical in nature, or simply captures noise. As an example, one of the eight human ASCL1 perturbation experiments had no DEGs (Figure 4B). *SLC9C2* was ranked 21st in this experiment, and thus this gene had a very high minimum ranking for ASCL1, despite not being DE in any of the eight ASCL1 perturbation experiments. Correspondingly, *SLC9C2* outranked genes that did have recurrent DE evidence across experiments but were not exceptionally high in any individual experiment.

Given these caveats, ideally context-specific analysis would be based on aggregation of multiple experiments from the same context, as we performed above for RUNX1 K-562 experiments. We believe this strategy would enrich for specific interactions in a more robust manner than the minimum ranking. However, this requires having sufficient data for each context.

Despite the concerns about whether the minimum ranking approach is heavily contaminated by technical outliers or noise, we make available all effect size matrices used for analysis (Supplemental Data S2), which allows ready ordering of individual experiment evidence. We also provide the summarized rankings with the minimum ranking strategy, which also records the “best” genomic experiment for each ranked gene (Supplemental Data S6).

Low-throughput curated target resources

Our evaluation of the genomic rankings was facilitated by resources that curated biochemical assays probing TR-target interactions. These include low-throughput perturbation assays (such as qPCR), binding assays (ChIP, EMSA), and reporter assays (such as luciferase reporters). The majority of the curated interactions (60%) are uniquely identified from our previous work, Chu et al., 2021 (here, Chu2021), along with additional interactions that have been curated since original publication. For simplicity these interactions are also noted as originating from Chu2021 (a manuscript reporting the increased coverage is in preparation). Additionally, we include targets that have been curated by other resources, which was organized in Chu2021. TRRUST (Han et al., 2018) was by far the largest of these resources. Interactions curated by Chu2021 and/or TRRUST comprise 95% of the low-throughput TR-target relationships in this study. In decreasing order of prevalence, the remaining resources that were included are TFe (Yusuf et al., 2012), ENdb (Bai et al., 2020), CytReg (Carrasco Pro et al., 2018), HTRIdb (Bovolenta et al., 2012), InnateDB (Lynn et al., 2008), TFactS (Essaghir et al., 2010), and ORegAnno (Lesurf et al., 2016).

We further note that multiple of the TCF4 interactions found when combining all of the curated resources were actually regarding TCF7L2, which is frequently referred to as TCF4 (T-Cell Factor 4) in the literature. We took care to remove errors in the resources used arising from confusion between TCF4 and TCF7L2.

We also want to clarify that we made the decision to include all non-curated targets as “negatives” as all of our comparisons are based on relative rather than absolute AUPRCs. We note that Garcia-Alonso et al., (2019) also performed an AUPRC analysis (benchmarking using perturbation data), but for negatives they down-sampled the “non-target/negative” group to match the size of the “target/positive” group in order to (artificially) balance the group sizes. We explored this and found that it provided much higher AUPRCs (in line with what they report) than using the entire list. However, we felt that this framework creates high AUPRCs while suffering from the same biases (the sampled negative set still assumes lack of current evidence means it is a true negative), which we felt would risk misleading readers.

Fold change Purity

To quantify whether a gene was predominantly up/down-regulated across the different types of perturbation experiments, we adopted the Purity metric used in clustering analysis to determine cluster quality. We treated gain-of-function (GoF – overexpressions) and loss-of-function (LoF – knockdowns and knockouts; mutants were excluded) experiments as two clusters in which genes are classified as either up-regulated (FC > 0) or down-regulated (FC < 0):

$$Purity_{g,j} = \frac{1}{N}\sum_{k=1}^{K} \quad max_{\,j}\left|c_k \cap \ t_j\right|$$

Where g indexes genes, j indexes TRs, N is the number of non-mutant experiments in which gene g is measured for TR j, K is the number of clusters (two – GoF and LoF), c_k is a cluster, and t_j is the dominant classification within that cluster. For example, if gene g has FC < 0 in 3/4 GoF experiments (GoF is dominantly down-regulated) and FC < 0 in 4/4 LoF experiments (LoF is dominantly down-regulated) for TR j, then $Purity_{g,j}$ = (3 + 4) / 8 = 0.875.

We further define Signed purity, which considers the dominant class/direction across both clusters: If the dominant direction is the same for GoF and LoF, we multiply Purity by -1 (thus the above example becomes -0.875), otherwise Purity is preserved. In this manner we can distinguish genes that show consistent changes in direction across experiments but unexpected convergence across the perturbation types. In this study, we arbitrarily define consistent genes as having Purity > 0.8.

Supplemental Fig. S12 provides a demonstration of the distribution FC Purity. We stress that this metric is sensitive to sample sizes, and that the sparsity of data (having balanced GoF and LoF experiments in a consistent context) for many TRs warrants a conservative interpretation. Nevertheless, the frequency in which we observed divergent FC directions motivated us to explore this trend and we believe it is an area worthy of further investigation.

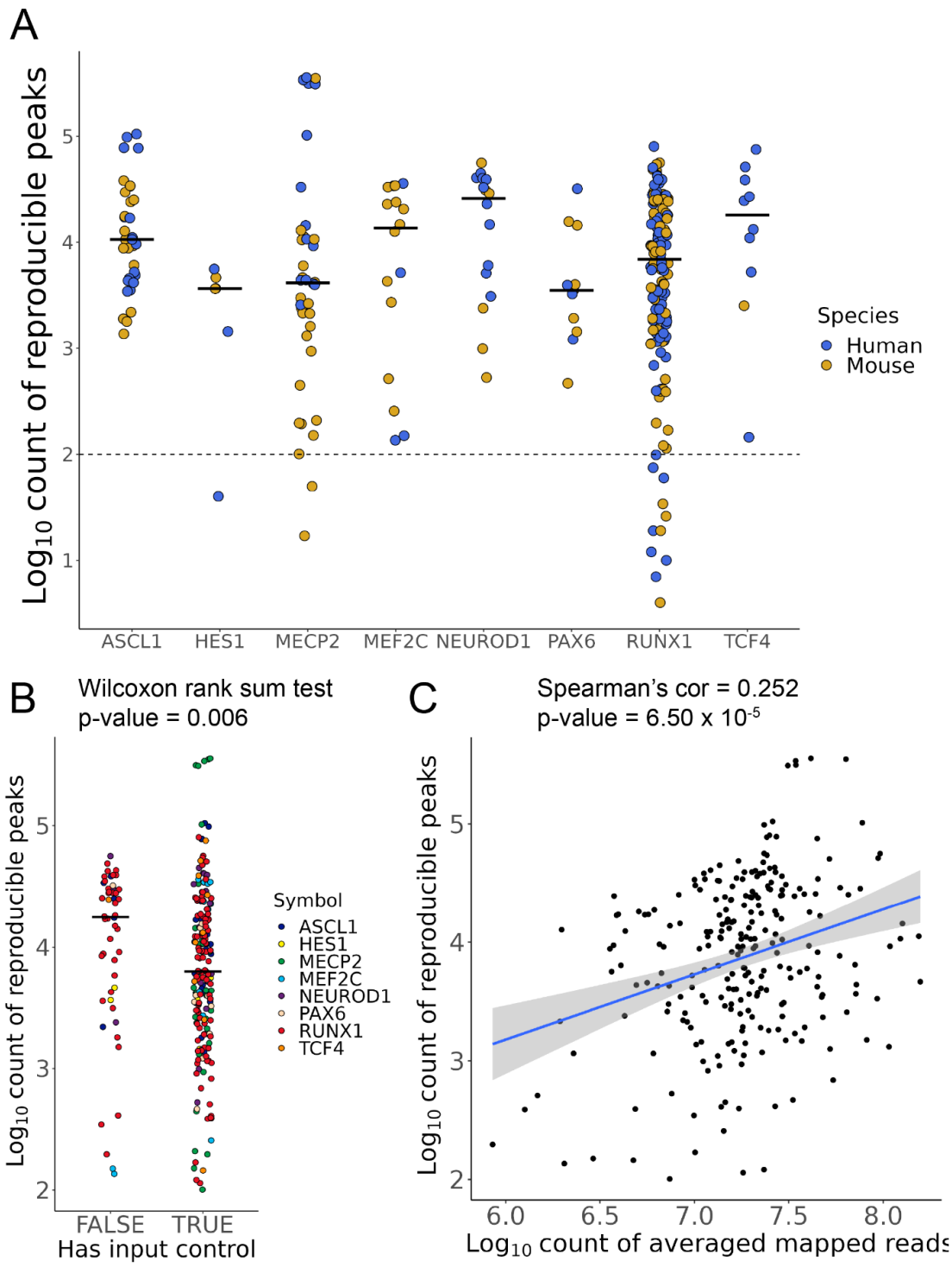


Figure S1. Overview of ChIP-seq experiment count of reproducible peaks. (A) Count of peaks by TR and coloured by species; dashed line represents minimum peaks required for analysis. (B) Count of peaks by presence of at least one input control. (C) Count of peaks by averaging de-duplicated and mapped reads across samples within an experimental unit.

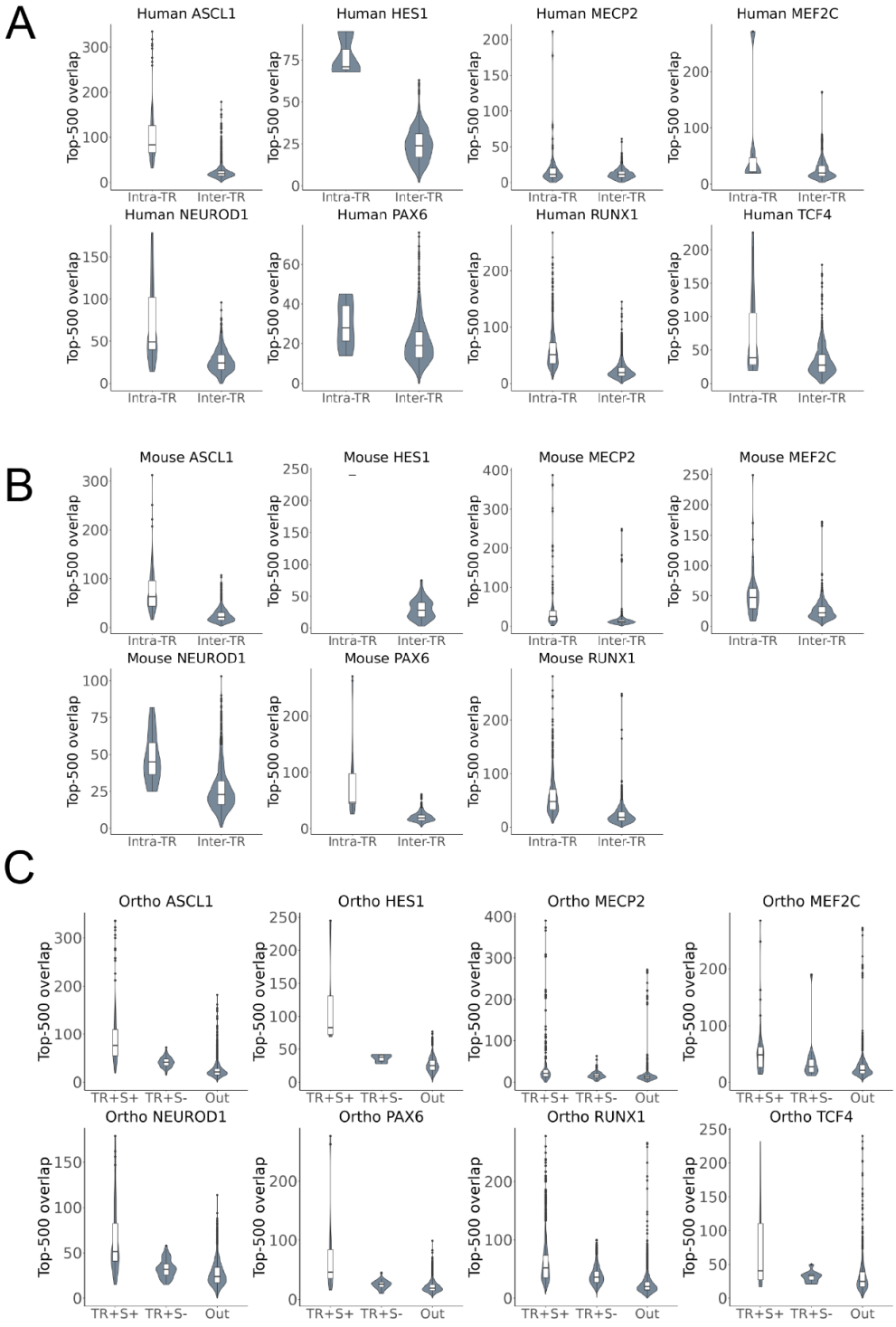


Figure S2. Distribution of the count of overlapping genes when selecting the top 500 genes by binding score when comparing ChIP-seq experiments targeting the same TR (intra-TR) or a different TR (inter-TR) for (A) human, (B) mouse, and (C) both species when using only high-confidence 1:1 orthologous genes. Abbreviations in C: TR+ (intra-TR), S+ (within-species), S- (cross-species), Out (collapsing inter-TR & intra-species with inter-TR & cross-species).

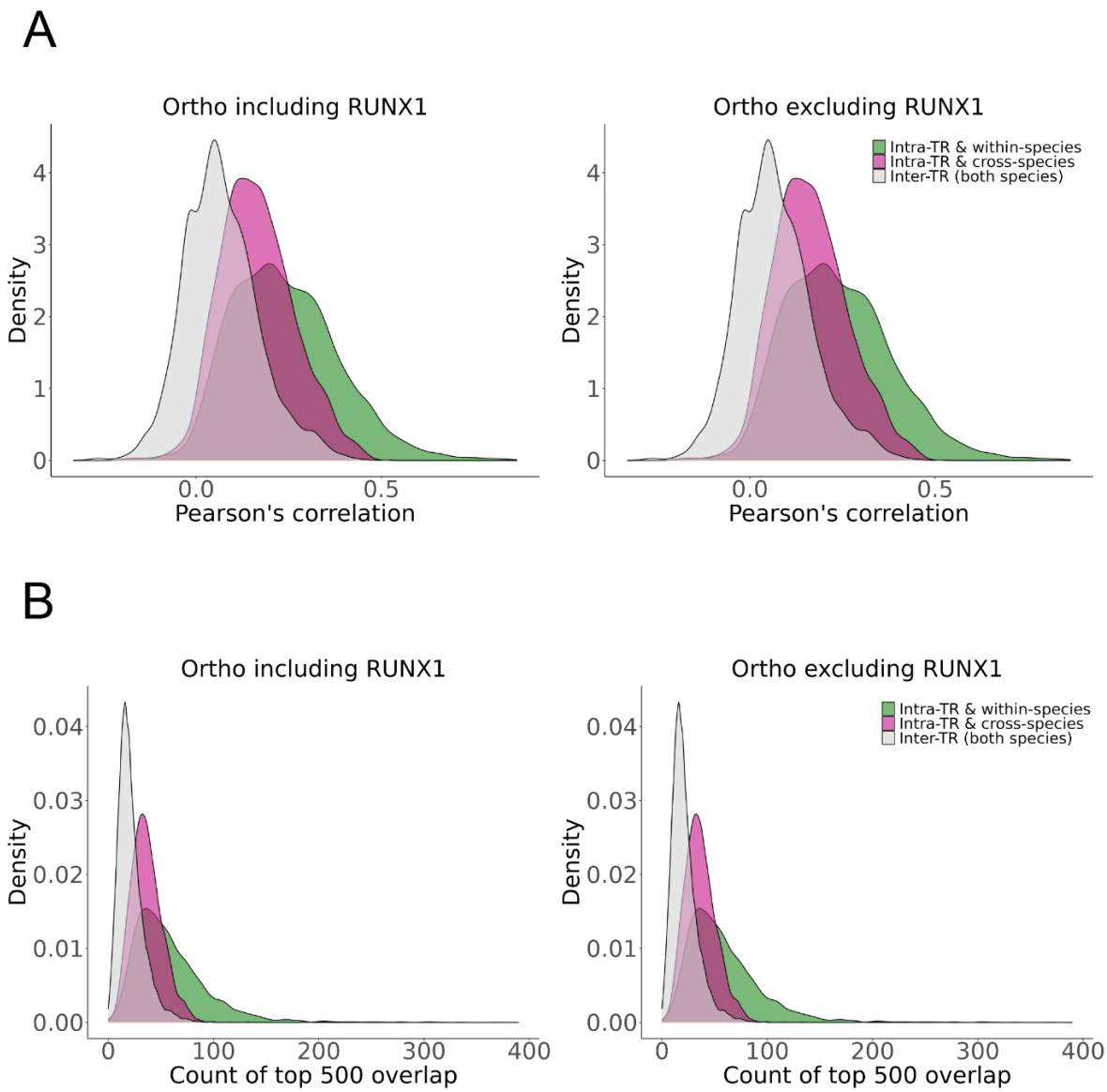


Figure S3. ChIP-seq experiment similarity of orthologs with and without RUNX1. (A) Left panel: As in Figure 2C, the distribution of binding score correlations between ChIP-seq experiments targeting the same TR (intra-TR) versus different TRs (inter-TRs), considering only orthologous genes. Right panel: The same comparison, except excluding all RUNX1 experiments. (B) Same as in A, except showing the density plots of the count of top 500 overlapping genes by binding score between pairs of experiments.

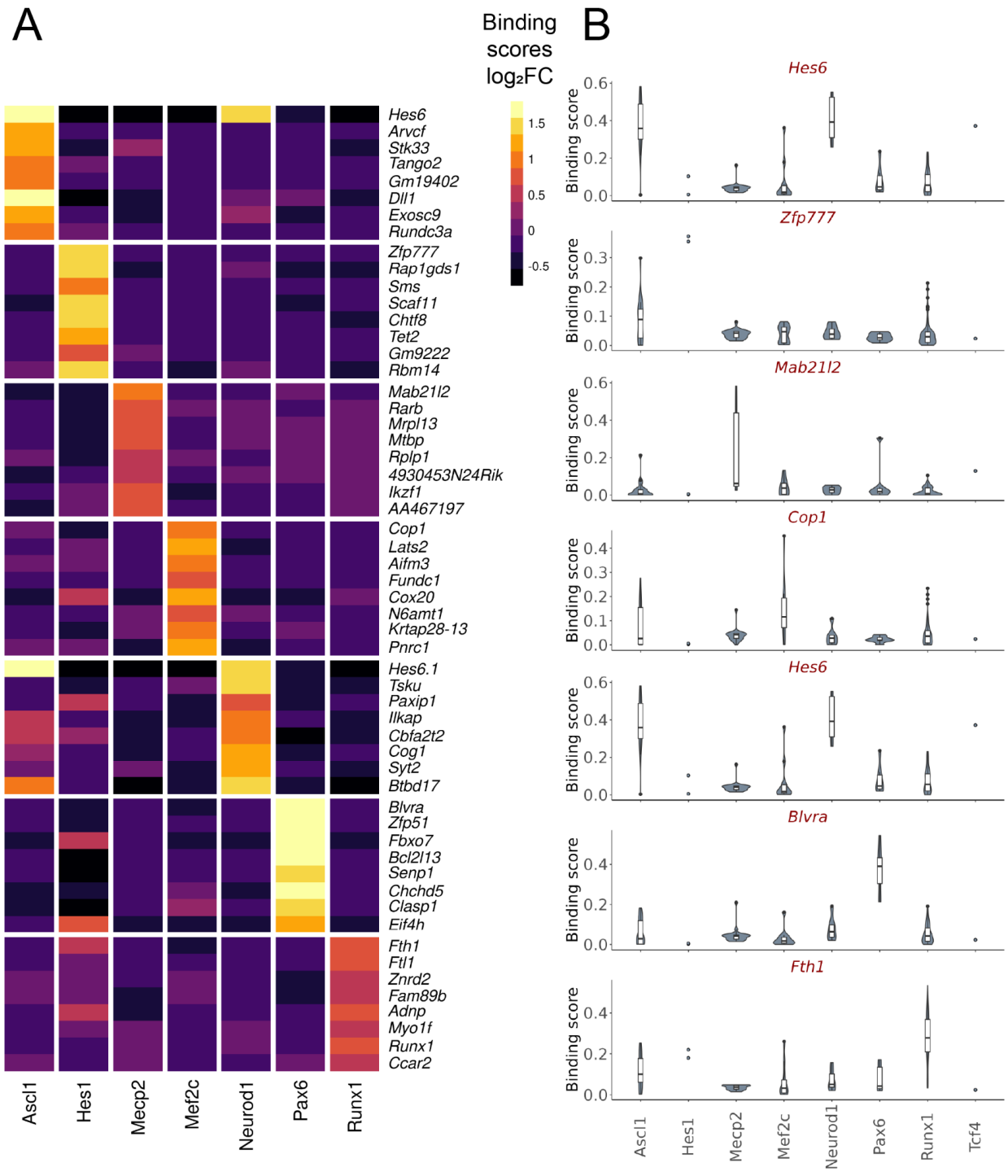


Figure S4. Specifically bound genes in mouse. (A) Log_2fold changes of binding scores for the top eight ranked genes (by p-value) for each TR in mouse experiments using a mixed effect linear model. (B) Distribution of binding scores for the most significant gene for each TR extracted from the same model as in (A). Note that *Hes6* was the most significant gene in both the Ascl1 and the Neurod1 contrast.

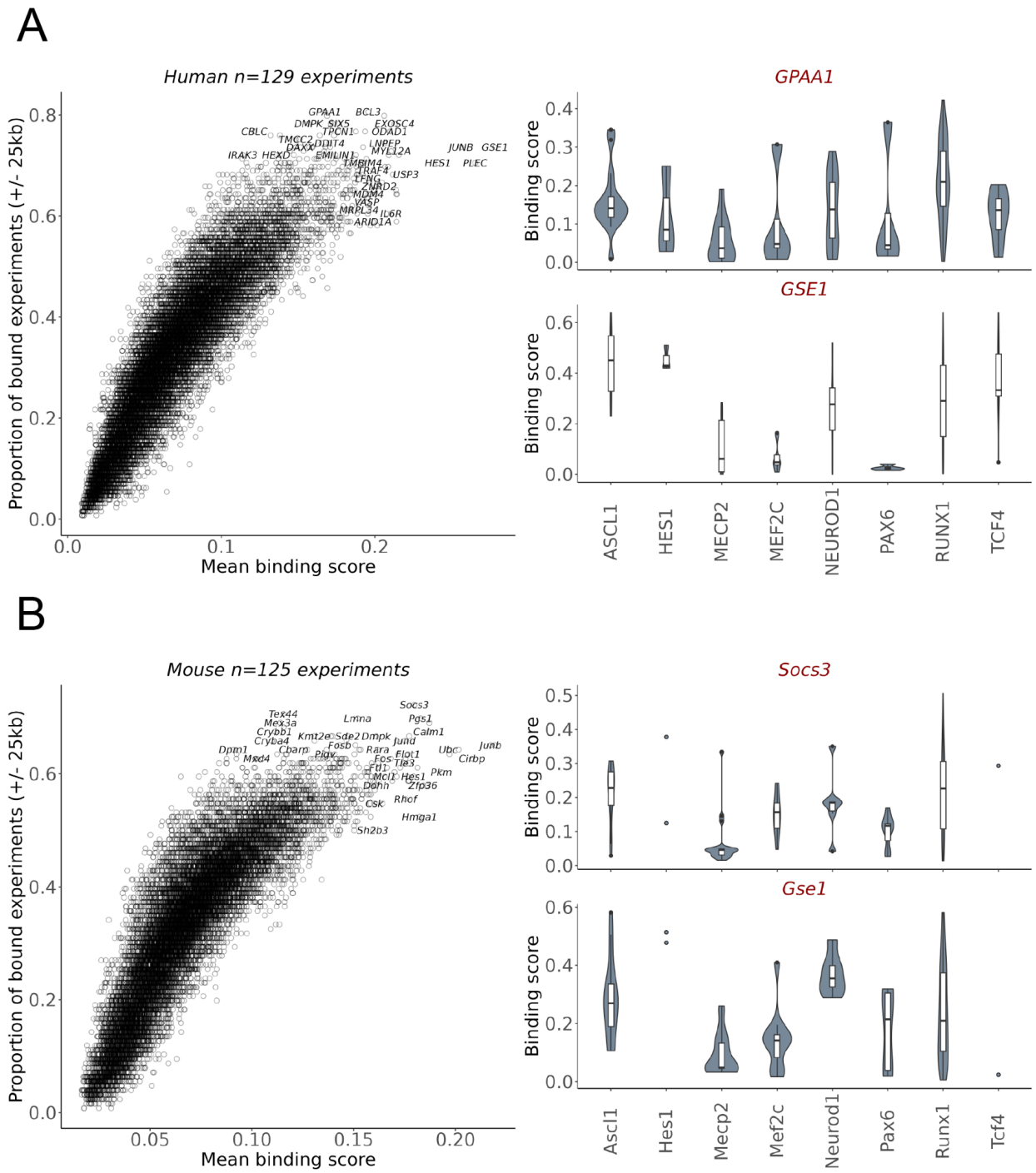


Figure S5. Demonstration of genes with high aggregated scores in the current ChIP-seq corpus. (A) Left panel scatter plot: every point is a gene, where the y-axis shows the proportion of human ChIP-seq experiments that had a peak within 25 kbp of the gene TSS (binary binding assignment), and the x-axis shows the mean of the continuous binding score across all experiments. Right panel: distribution of binding scores for the top gene by proportion bound (GSAA1) and by mean binding score (GSE1). (B) Same as in (A), but for mouse experiments.

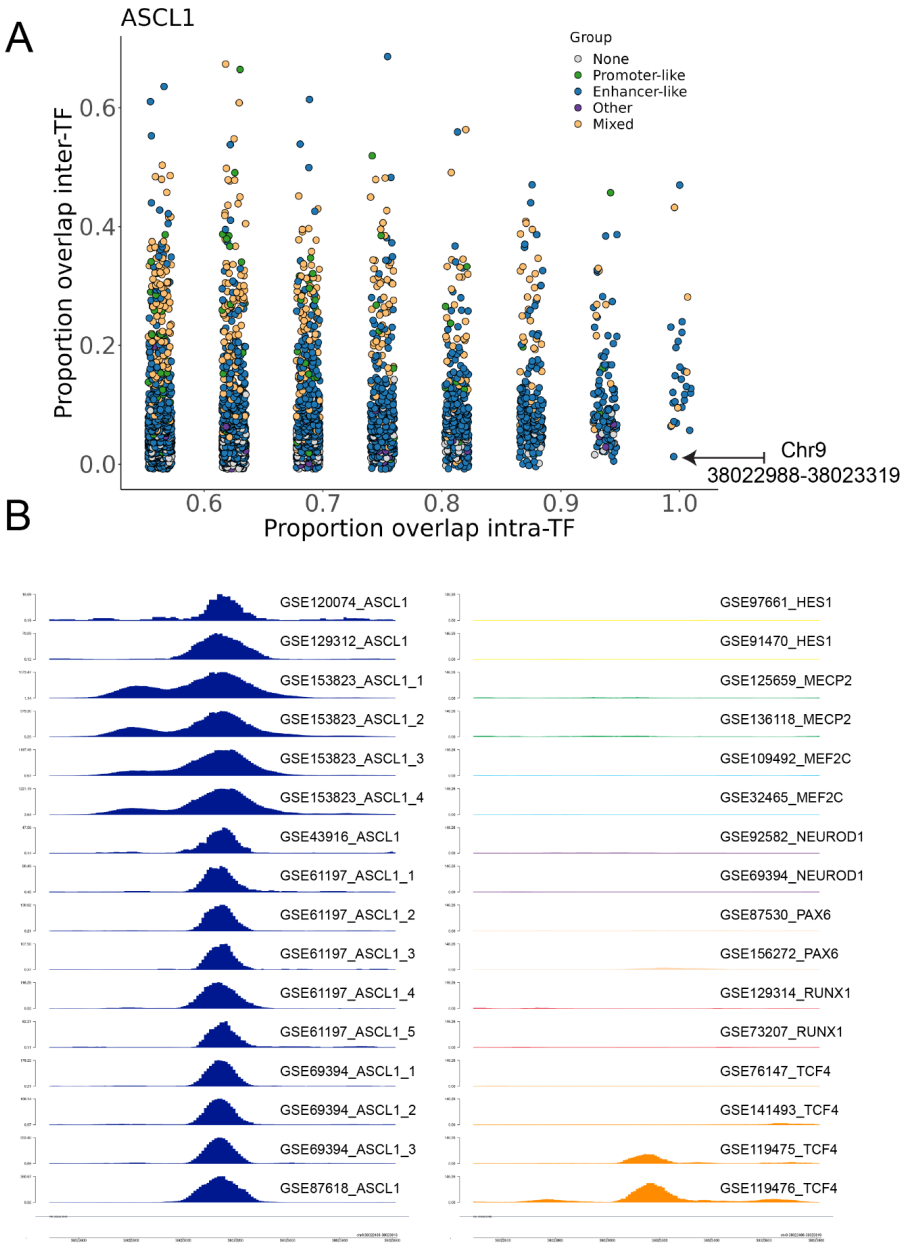


Figure S6. ASCL1 frequently bound loci. (A) Each point represents a human ASCL1-bound region, restricted to regions bound in at least half of ASCL1 experiments. The x-axis is the proportion of human ASCL1 experiments (n=16) in which that region is bound, and the y-axis is this proportion for non-ASCL1 experiments (n=113). Points are colored by ENCODE cCRE definitions, binning each enhancer or promoter subgroup (seen in legend of Figure 3A), and designating non-ELS/PLS elements as 'Other.' Mixed refers to regions straddling two cCRE groups. (B) Trackplot highlighting an intronic region of *SHB*, indicated with an arrow in (A), that had a peak in all ASCL1 experiments and only in 2/113 non-ASCL1 experiments. Left panel shows all ASCL1 experiments, right panel shows two sampled experiments for the other TRs, as well as the two TCF4 experiments in which this region had a peak called. Note that the left panel has variable y-axes limits (reflecting read enrichment) to fit all the peaks, while the right panel y-axes limits are fixed to the median of the maximum ASCL1 intensities to demonstrate the relative lack of signal and to fit the two TCF4 peaks.

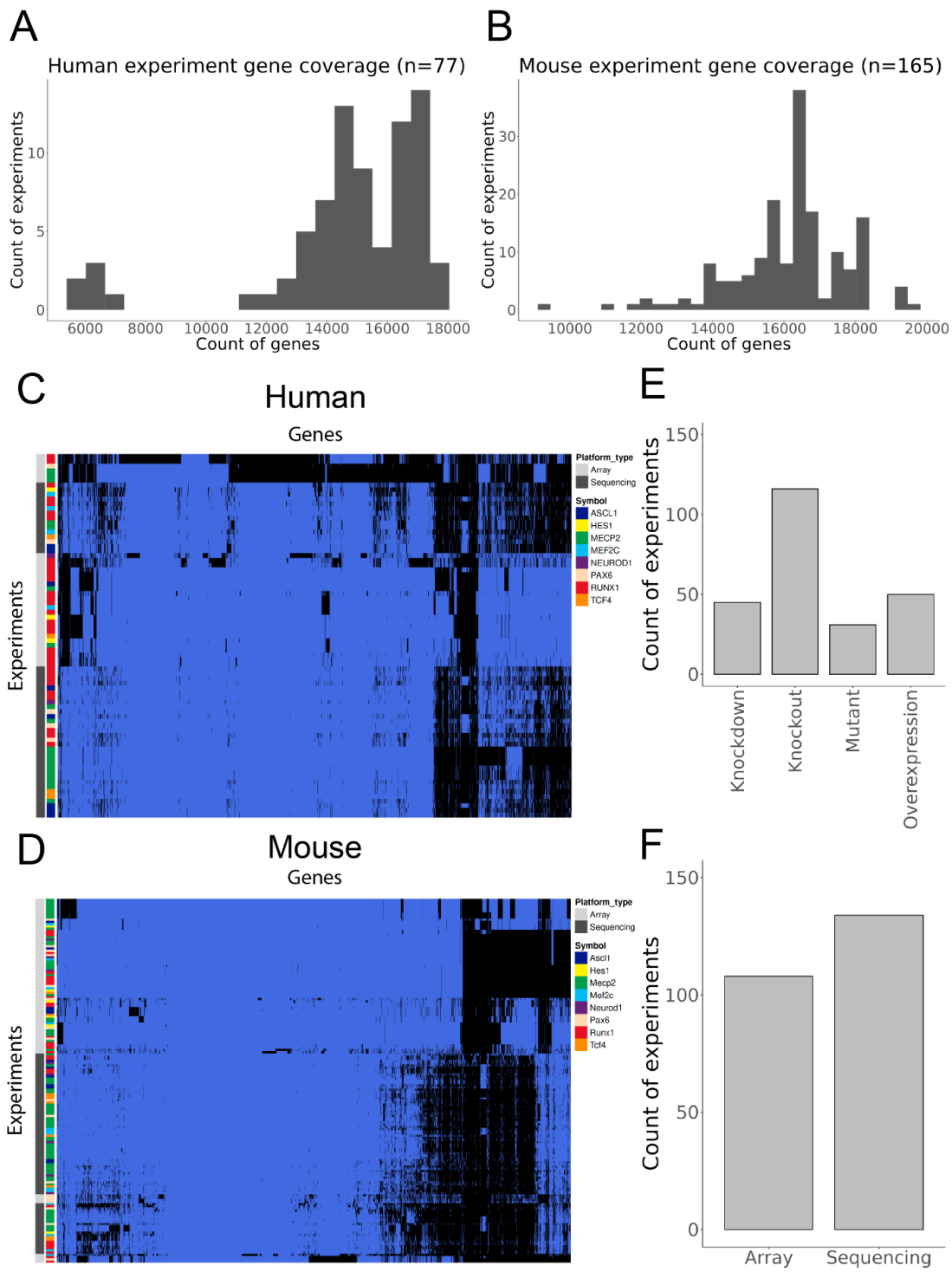


Figure S7. TR perturbation experiment overview. (A-B) Histograms of the count of genes measured in human and mouse perturbation experiments. For microarray platforms, genes that were not measured are not on the platform or were filtered in QC. For RNA-sequencing, genes that were not measured were not appreciably detected in the tested context. (C-D) Binary heatmap of genes measured across human (C) and mouse (D) experiments; blue = measured, black = not measured. (E-F) Barcharts for the count of experiments by perturbation strategy (E) or expression technology (F).

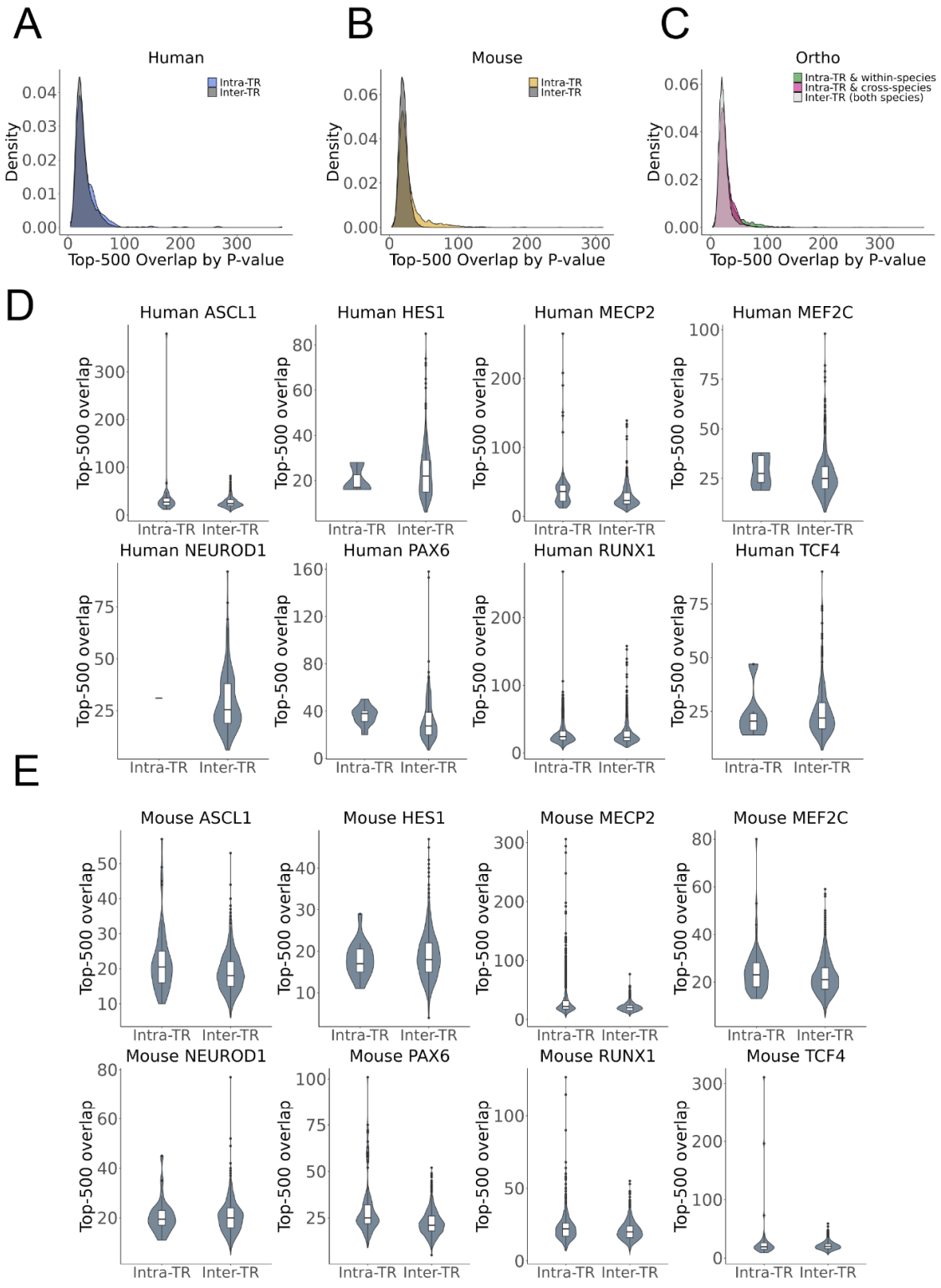


Figure S8. TR perturbation experiment similarity. (A-C) Density plots of the count of overlapping genes between perturbation experiments, selecting for the top 500 mutually-measured genes sorted by p-value from the differential expression analysis. (D) Distribution of the counts from (A) broken down by TR. (E) Distribution of the counts from (B) broken down by TR.

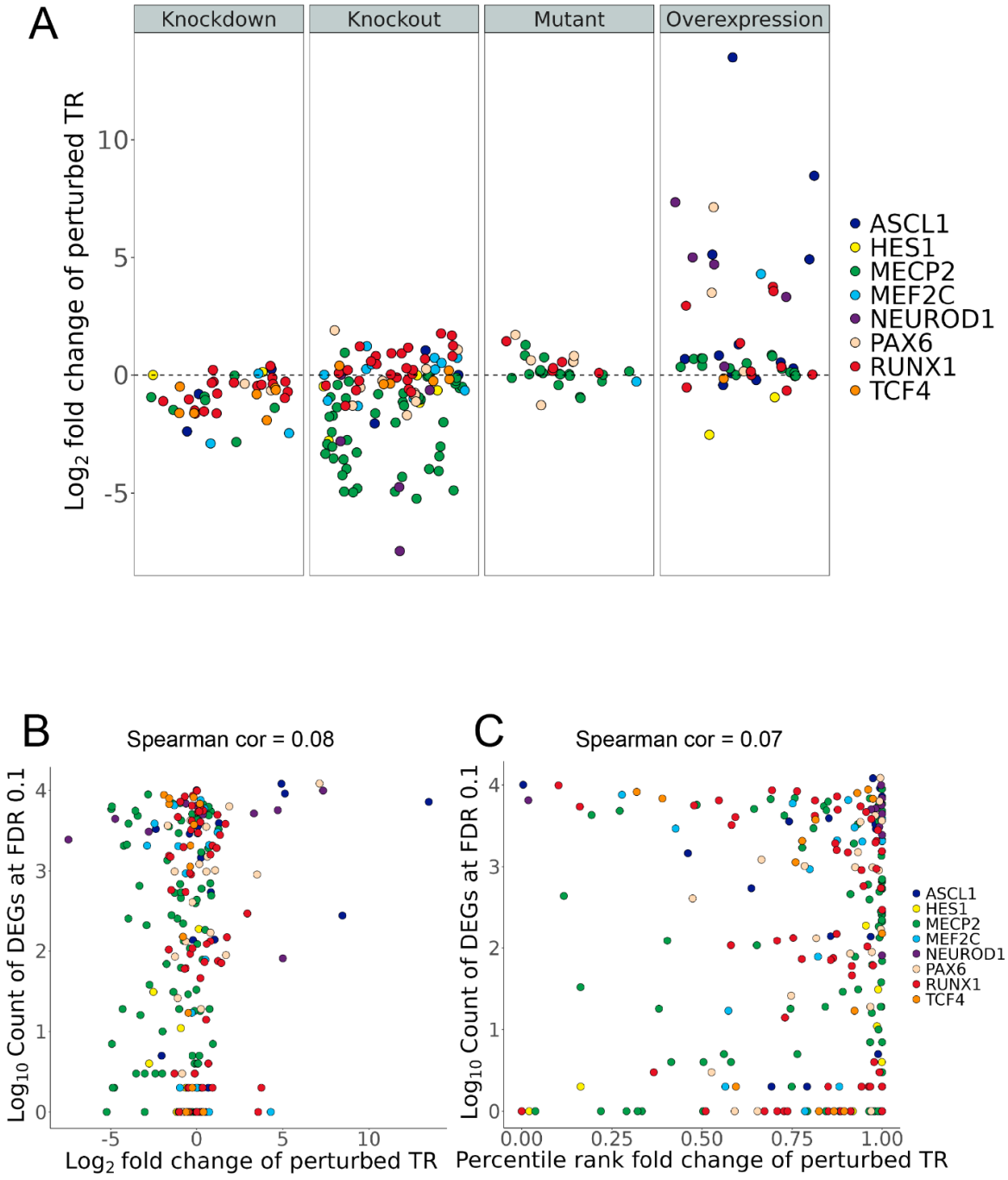


Figure S9. Perturbation effect sizes. (A) Each point represents the $\log_2$ fold change of the perturbed TR in mouse and human perturbation experiments, coloured by TR and grouped by perturbation strategy. (B-C) Relationship between the effect size of the perturbed TR and the count of differentially expressed genes in that experiment. (B) The fold change of the perturbed TR is shown, while (C) shows the percentile rank of the absolute fold change of the perturbed TR (where 1.0 = the gene with the highest absolute fold change in that experiment).

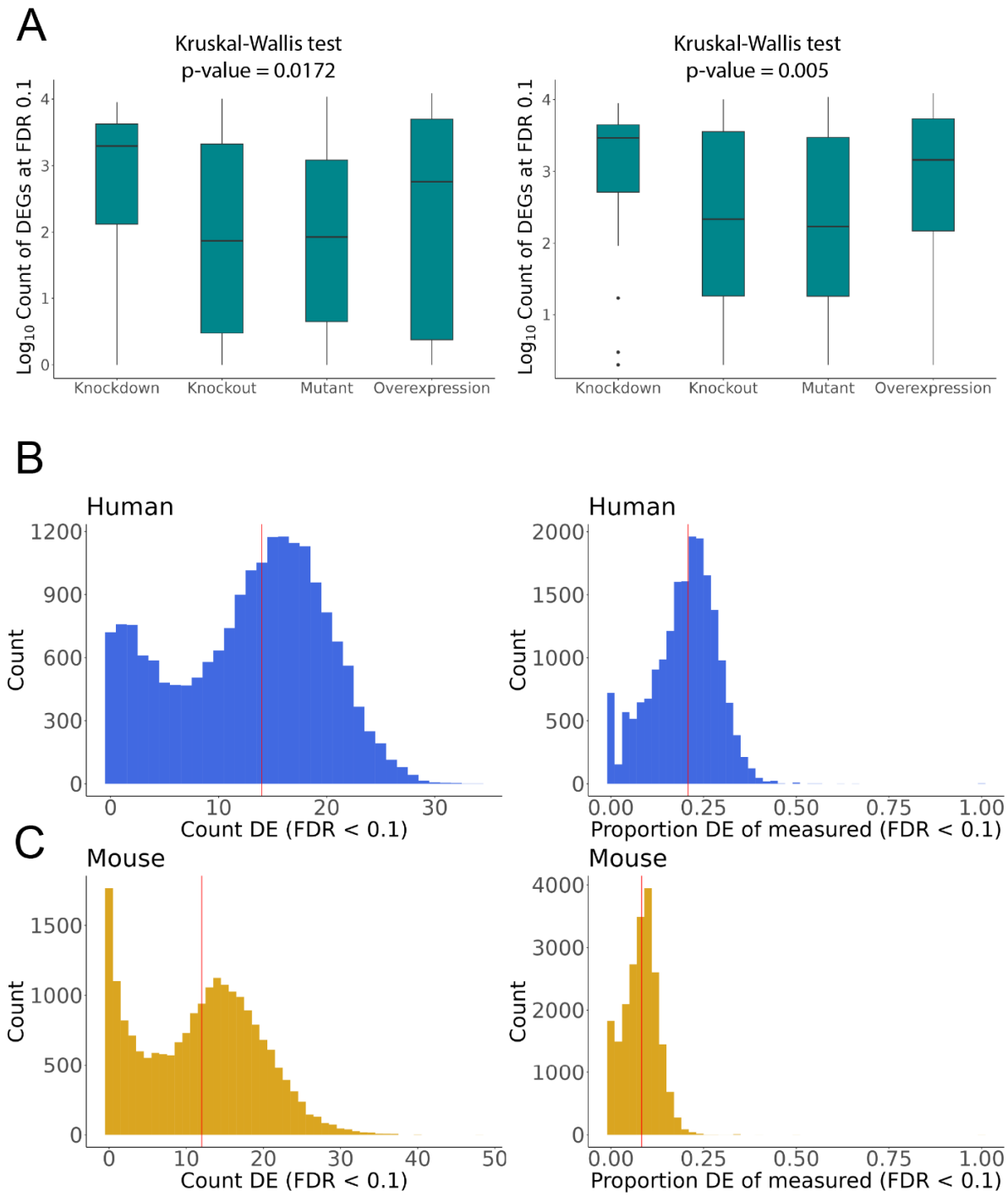


Figure S10. Count of differentially expressed genes (DEG). (A) Count of DEGs by perturbation strategy, either including experiments with no DEGs (left panel) or excluding experiments with no DEGs (right panel). (B) Left panel: histogram of the count of times a gene was measured as DE across human experiments (n=77); Right panel: instead of raw counts, the histogram shows the proportion of DE measurements relative to the number of experiments in which the gene was measured. (C) Same as in (B), except for mouse experiments (n=165).

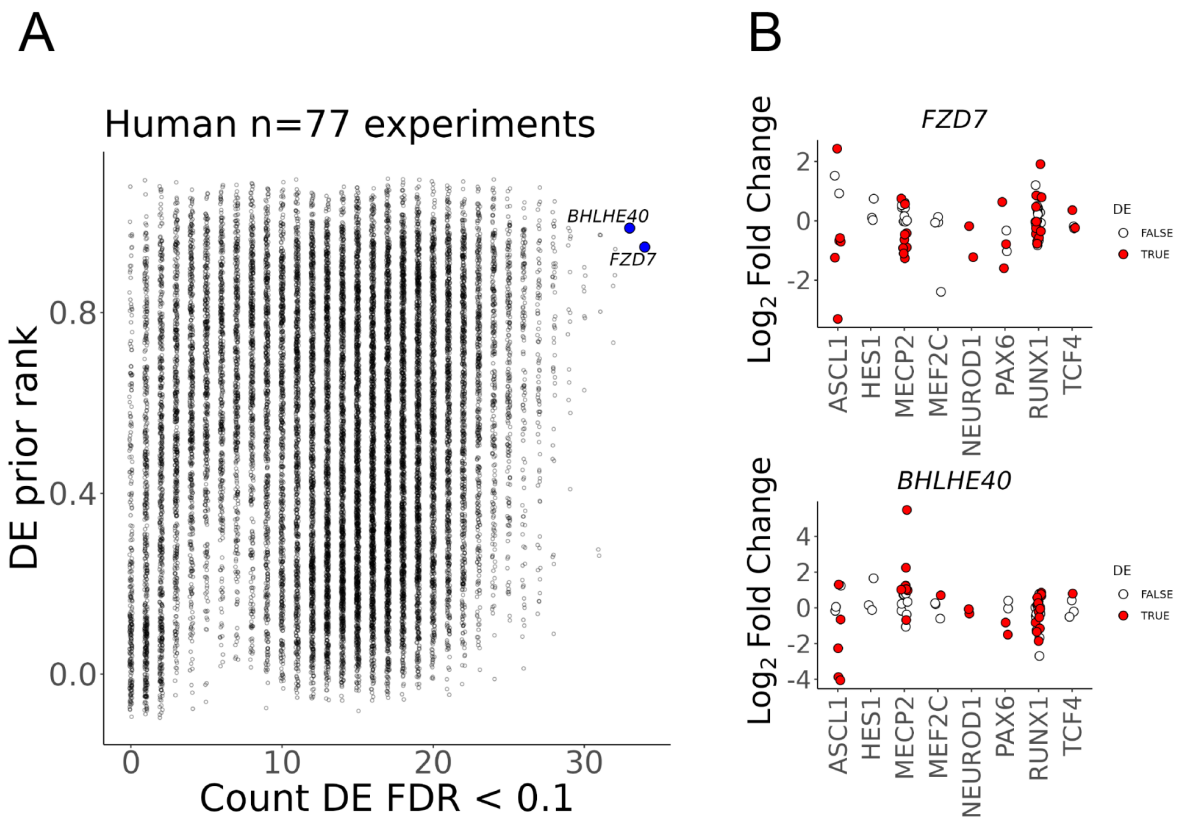


Figure S11. Frequently DE genes. (A) Relationship between the count of times a gene was DE across all human experiments for the eight TRs, and the DE prior ranking, as in Figure 5A. Highlighted are the two genes with the highest DE counts. (B) FC distributions for the genes highlighted in (A), coloured by DE status.

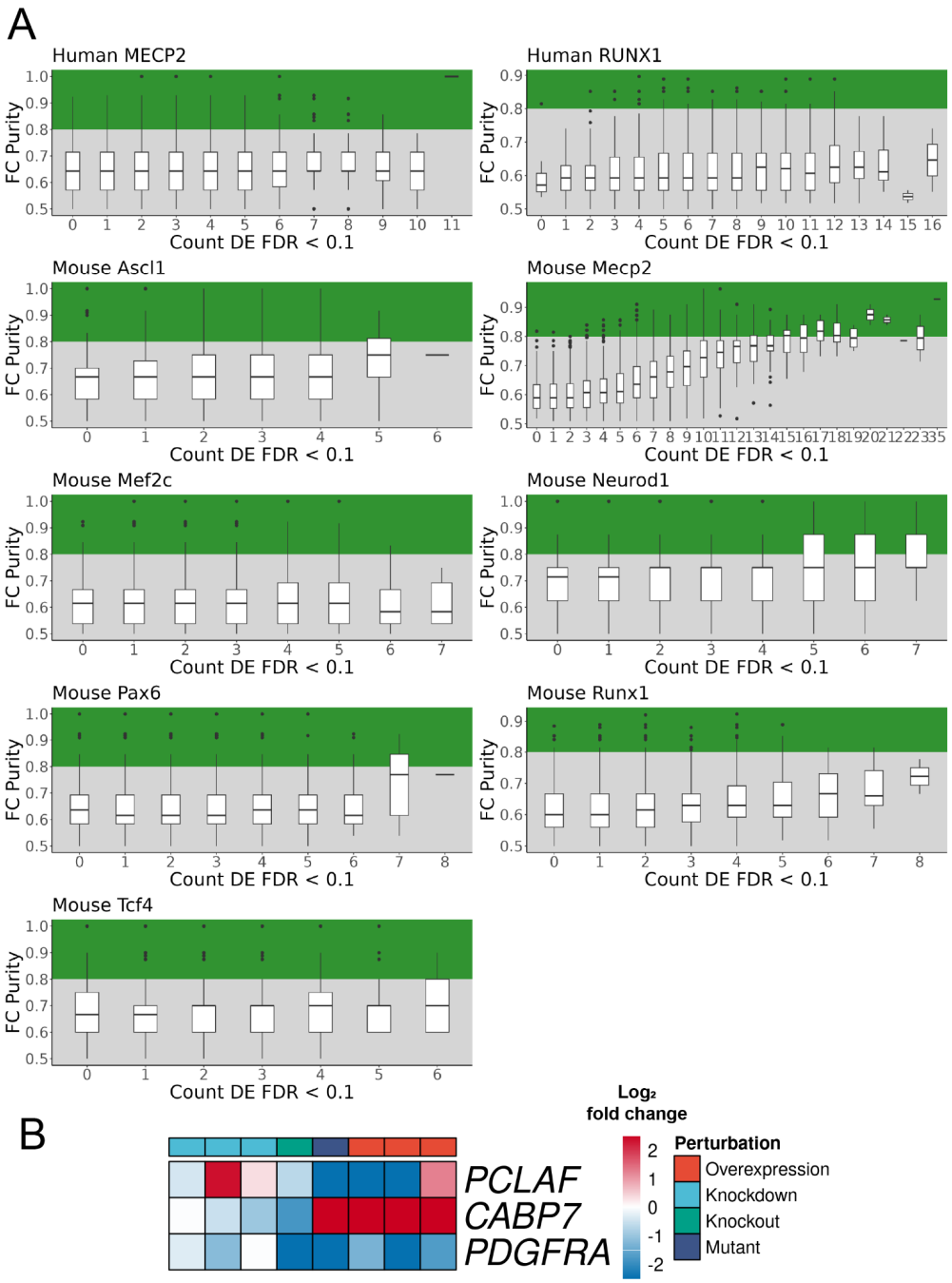


Figure S12. Genes typically show variable changes in FC direction across TR perturbation experiments. *Purity* is a measurement of the consistency of a gene's FC direction within TR gain/loss of function experiments. (A) Demonstrates the TR-specific distributions of gene *Purity* scores, stratified by how often the genes were measured as DE across each set of TR perturbation experiments. Only TRs with at least eight non-mutant experiments were kept to alleviate issues of sample size. Genes with consistent FC direction were arbitrarily defined as having *Purity* > 0.8 (green background). (B) Shows three genes with strong human ASCL1 DE evidence (DE > 4/8 ASCL1 experiments), as in Figure 5D. *PCLAF* has low *Purity* (0.57) due to variation in its FC direction. *CABP7* has perfect *Purity* and *Signed purity* (1.00), while *PDGFRA* has a *Signed purity* of -1.00 as the FC direction is the same for gain/loss of function experiments.

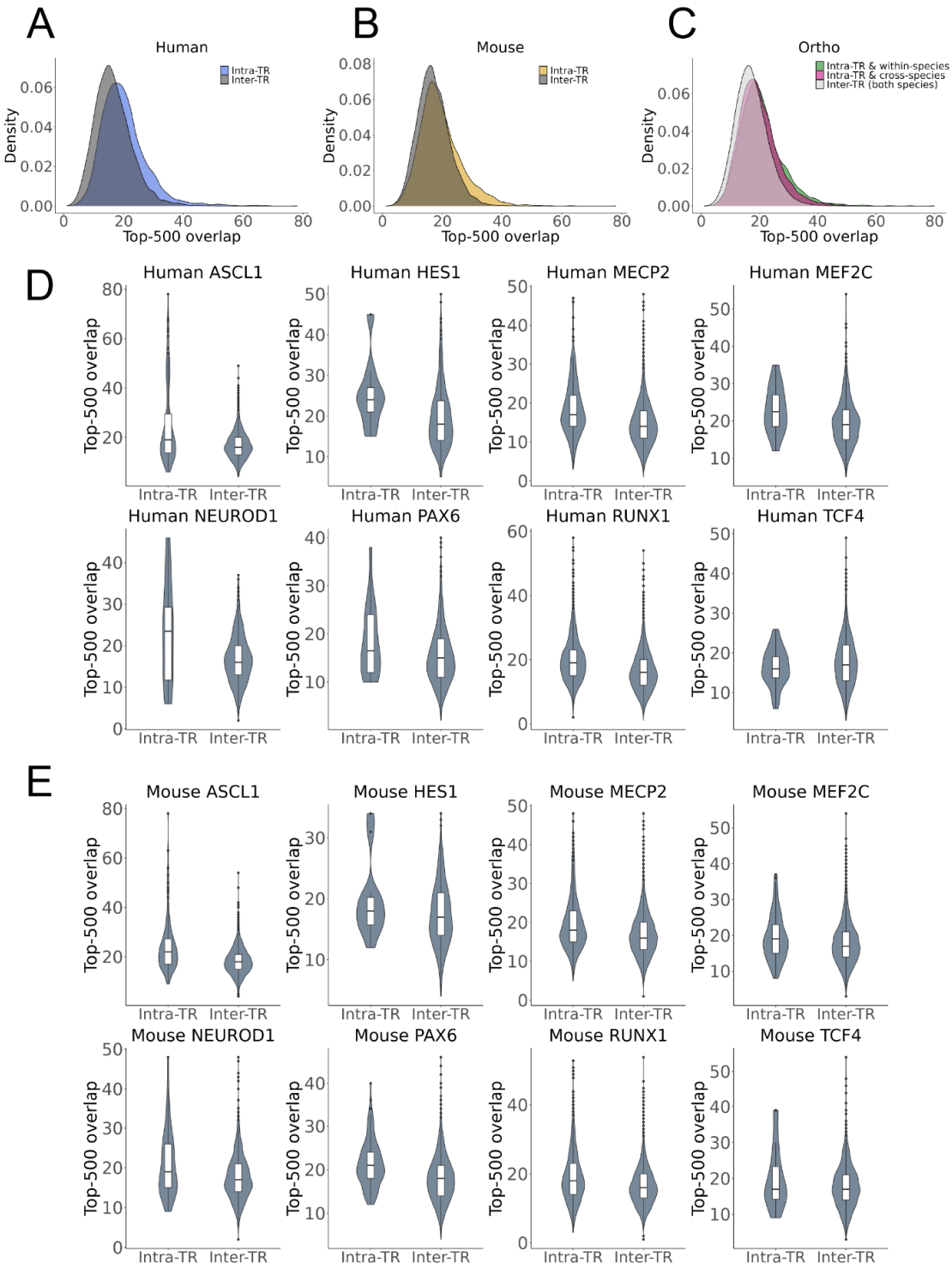


Figure S13. Top scoring gene overlap between ChIP-seq and TR perturbation experiments. (A-C) Density plots of the count of intersecting genes between ChIP-seq and perturbation experiments, selecting for the top 500 genes by binding score (ChIP-seq) or by p-value from the DE analysis (perturbation). (D) Distribution of the counts from (A) broken down by TR. (E) Distribution of the counts from (B) broken down by TR.

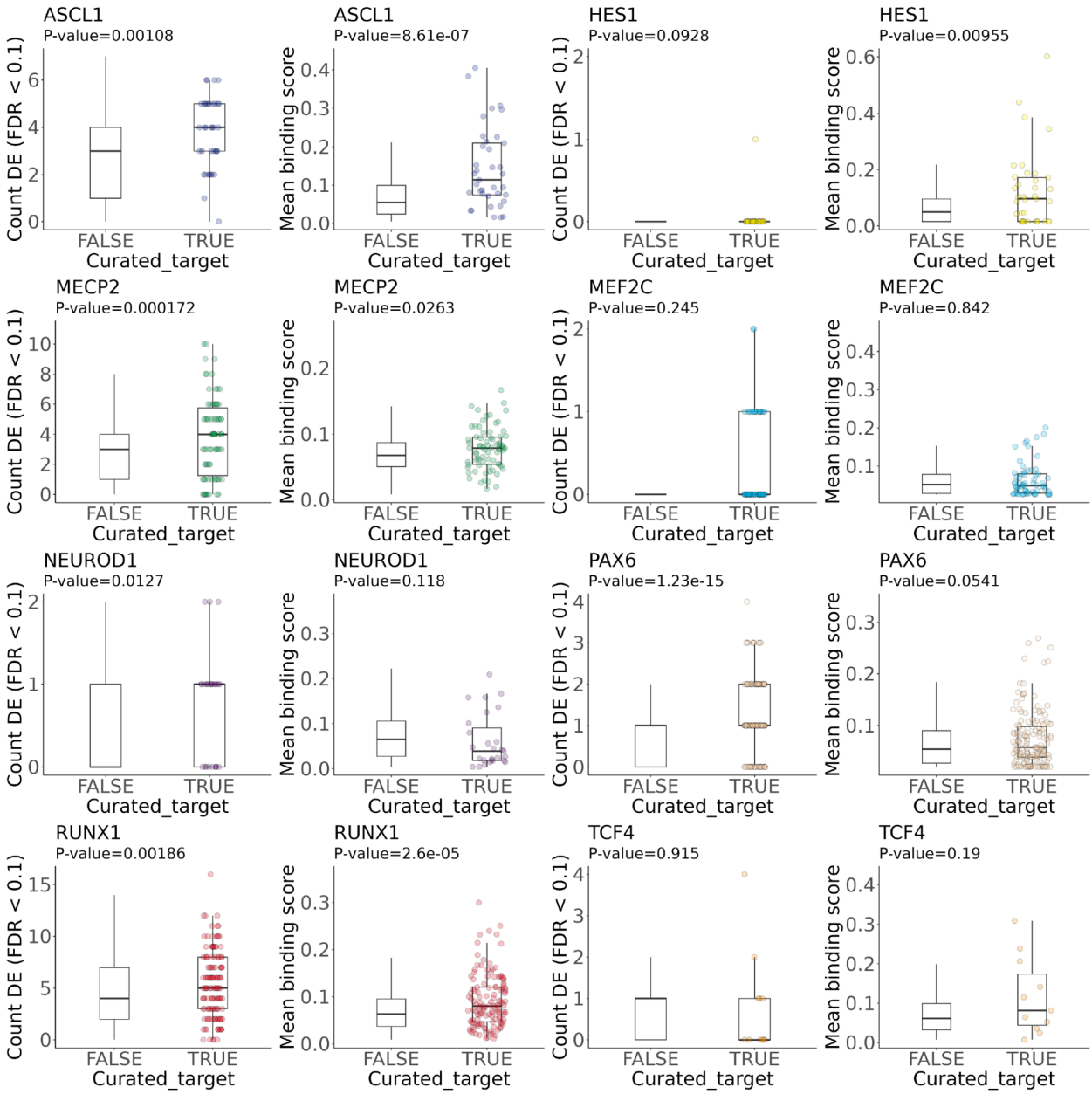


Figure S14. Human aggregate scores by curation status. Boxplots show the aggregated genomic experiment gene scores for gene status in the low-throughput evidence resource. P-values are from the Wilcoxon rank sum test.

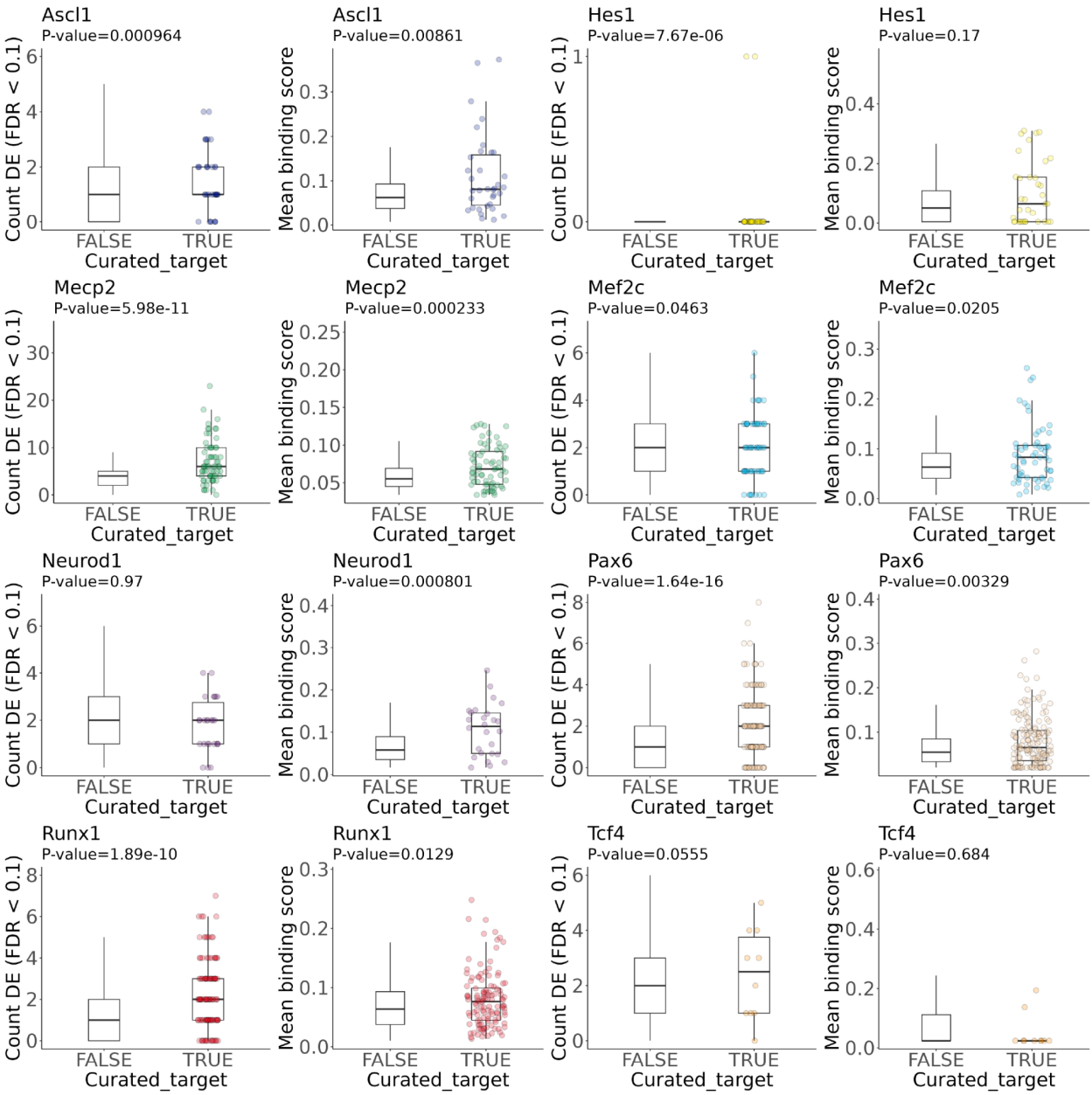


Figure S15. Mouse aggregate scores by curation status. Boxplots show the aggregated genomic experiment gene scores for gene status in the low-throughput evidence resource. P-values are from the Wilcoxon rank sum test.

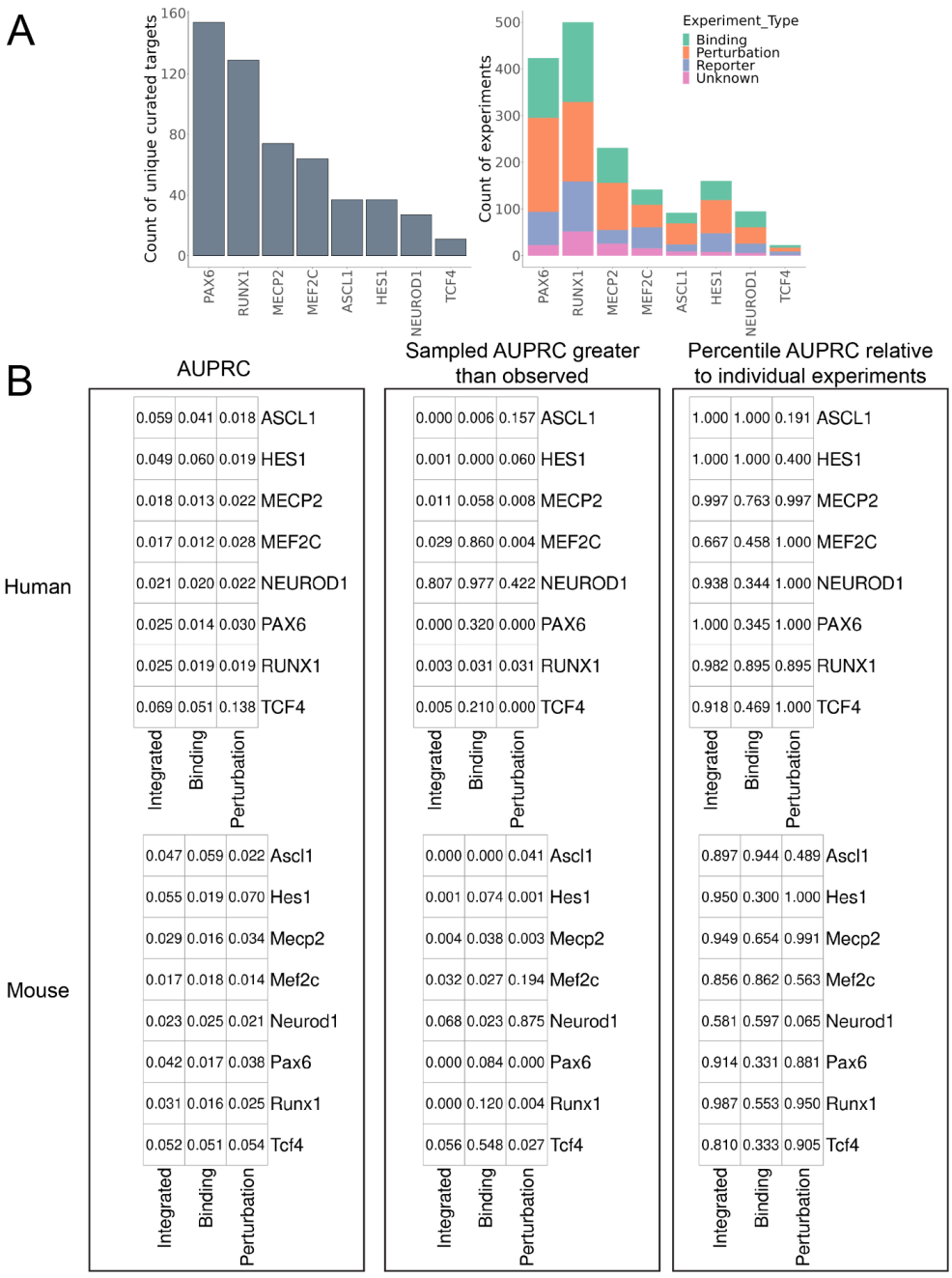


Figure S16. Detailed evaluation metrics for TR target prediction. For evaluation we accepted all low-throughput modes of evidence and included orthologous gene targets conducted in either species. (A) Left panel: Count of unique targets in the low-throughput target resource. Right panel: Count of unique experiments in the low-throughput resource. A single interaction may be supported by multiple experiments. Unknown refers to experiments recorded by external resources like TRRUST. (B) Left panel: AUPRC values are as demonstrated in Figure 6A,B. Mid panel: The proportion of sampled targets whose AUPRC exceeded the observed aggregated AUPRC, as in Figure 6C. 1000 samples, size matched to each TR's curated targets, were drawn from the entire curated resource and evaluated with the integrated ranking. Right panel: The percentile of the aggregated AUPRCs relative to the distribution of AUPRCs generated by treating each contributing experiment as its own ranking, as in Figure 6D.

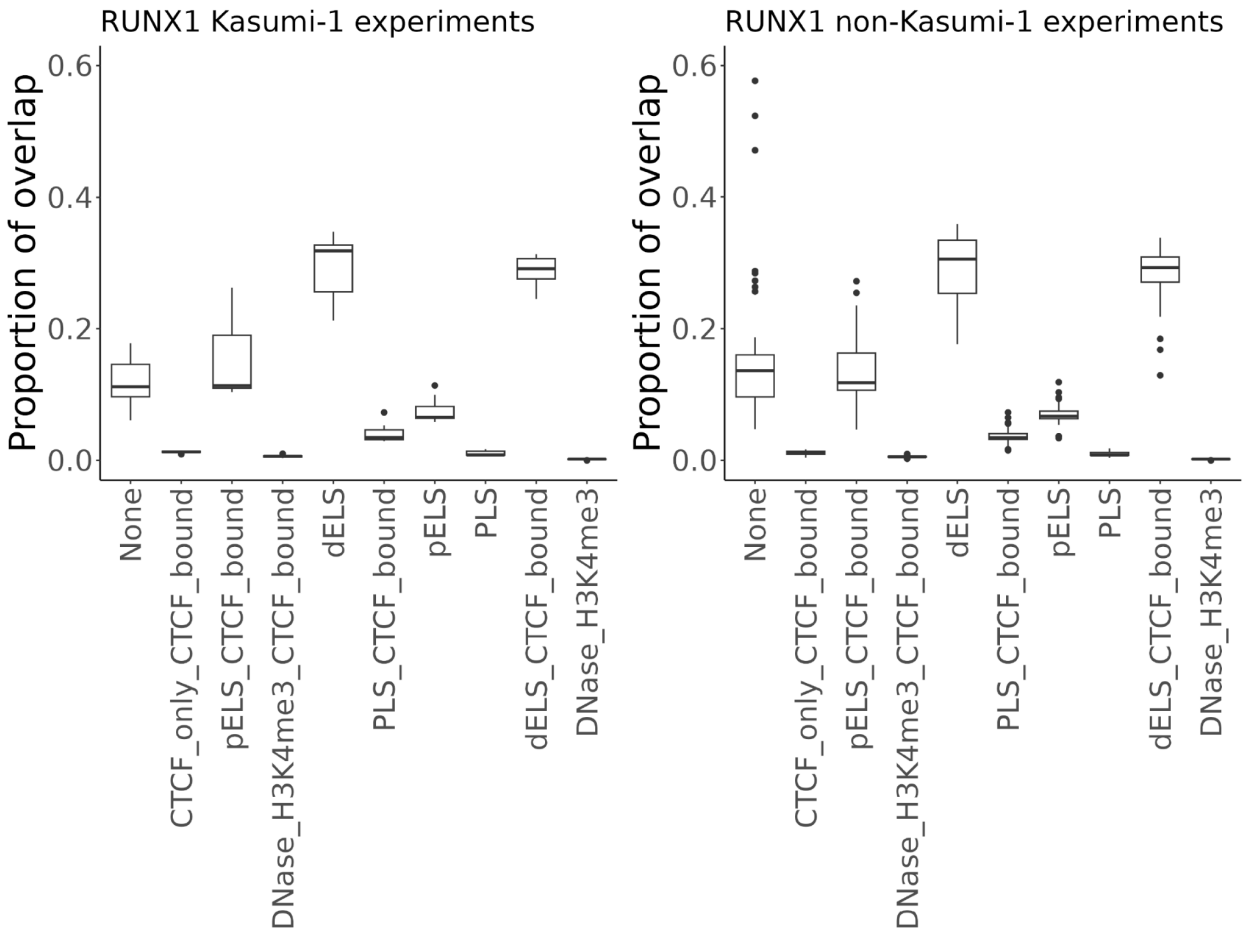


Figure S17. Distribution of cCRE overlap for a well-represented context. Left panel: cCRE overlap for n=11 RUNX1 ChIP-seq experiments in the Kasumi-1 cell line. Right panel: The same distributions for all other n=58 RUNX1 experiments. The Kasumi-1 experiments have a comparable range of overlap compared to all other experiments combined. This suggests technical variability is a factor, but other biological details beyond the cell line may be relevant as some of the Kasumi-1 studies include additional treatments.

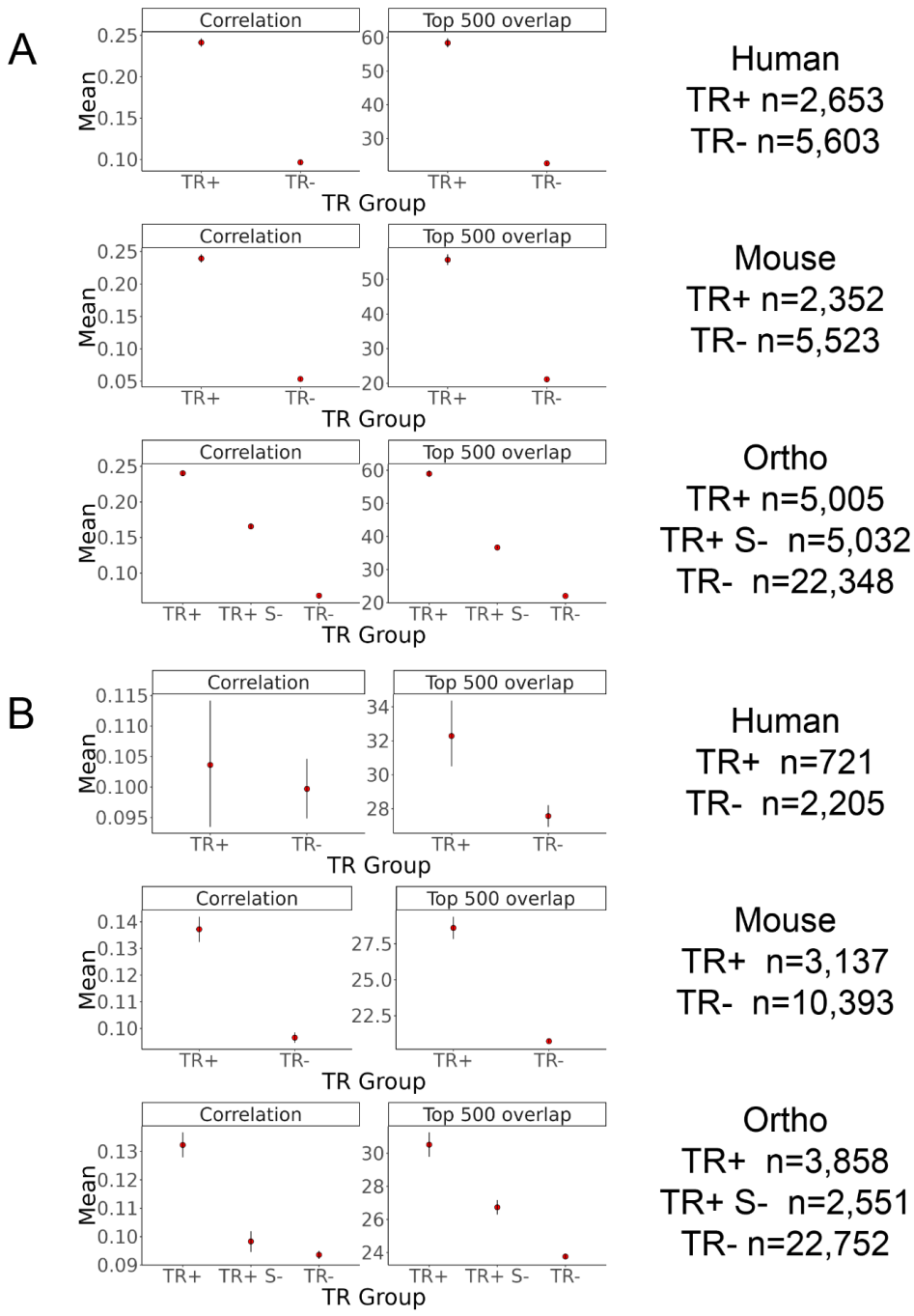


Figure S18. Point estimates and bootstrapped 95% confidence intervals (CIs) of experiment similarity. Relevant to Supplemental Figs S2 and S8. Experiment similarities were calculated as the average absolute fold change between experiments (left-hand panels), or the average top 500 overlap of genes sorted by p-values from the DE analysis (right-hand plots). The corresponding sample sizes (number of comparisons) are given in the right-hand column (A) Gene binding scores compared between ChIP-seq experiments. Comparisons are grouped by experimental pairs targeting the same TR (TR+), the same TR but different species (TR+ S-), or different TRs regardless of species (TR-). (B) The same as in (A) but for perturbation experiments.

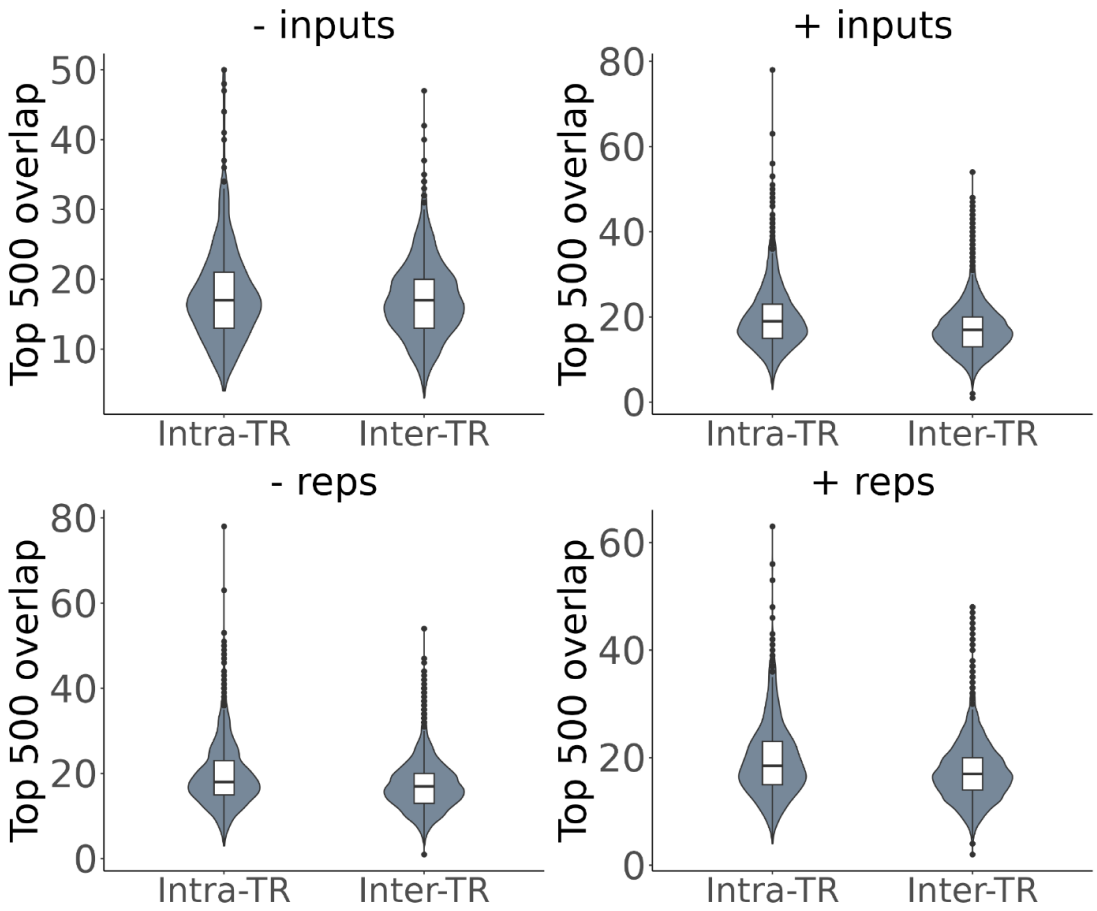


Figure S19. ChIP-seq experimental designs have minimal effect on the overlap of ChIP-seq and perturbation experiments. Relevant to Supplemental Fig. S13. The top 500 overlap of genes between mouse ChIP-seq (sorted by binding scores) and mouse perturbation (sorted by DE analysis p-values) experiments, split by ChIP-seq experiments without input controls (- inputs), with input controls (+ inputs), without replicates (- reps), and with replicates (+ reps). Mouse comparison shown due to prevalence of perturbation data relative to human.

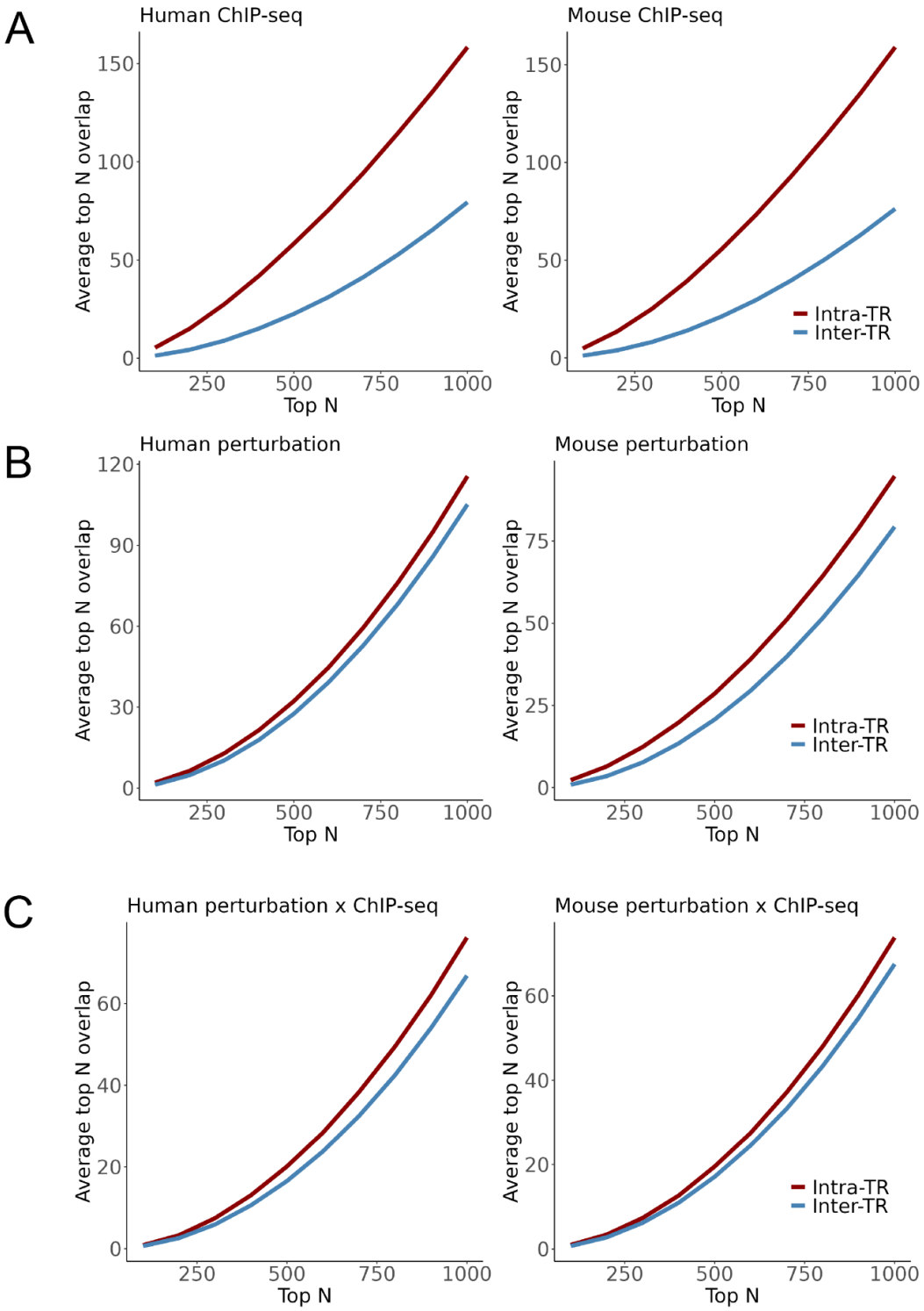


Figure S20. Relative comparisons of top-overlaps are insensitive to choice of threshold. The average count of overlapping genes between experiments as a function of the cutoff of top N genes. Results for human and mouse are in the left and right columns respectively. (A) ChIP-seq experiments, with genes sorted by binding scores. (B) Perturbation experiments, with genes sorted by p-values from the DE analysis. (C) Pairing ChIP-seq and perturbation experiments.

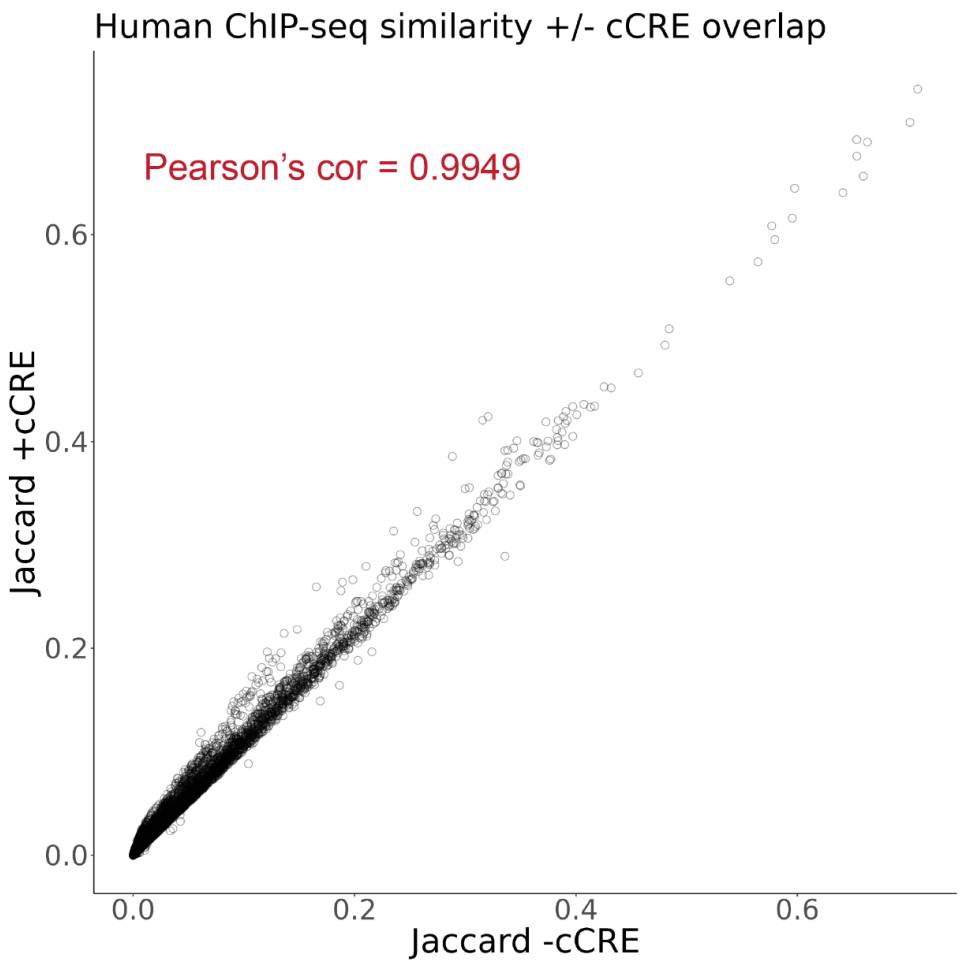


Figure S21. Filtering ChIP-seq experiments for cCRE overlap does not change experiment similarity structure. ChIP-seq experiments were represented as binary vectors representing whether genomic regions were bound or unbound. The Jaccard coefficient was calculated between every pair of binding vectors as a measure of binding region similarity (distinct from the gene binding score similarities used in the rest of this study). This process was repeated, except considering only the bound regions that also overlapped an ENCODE cCRE. Each point in the plot represents an experiment pair whose Jaccard was calculated with or without the cCRE filter.

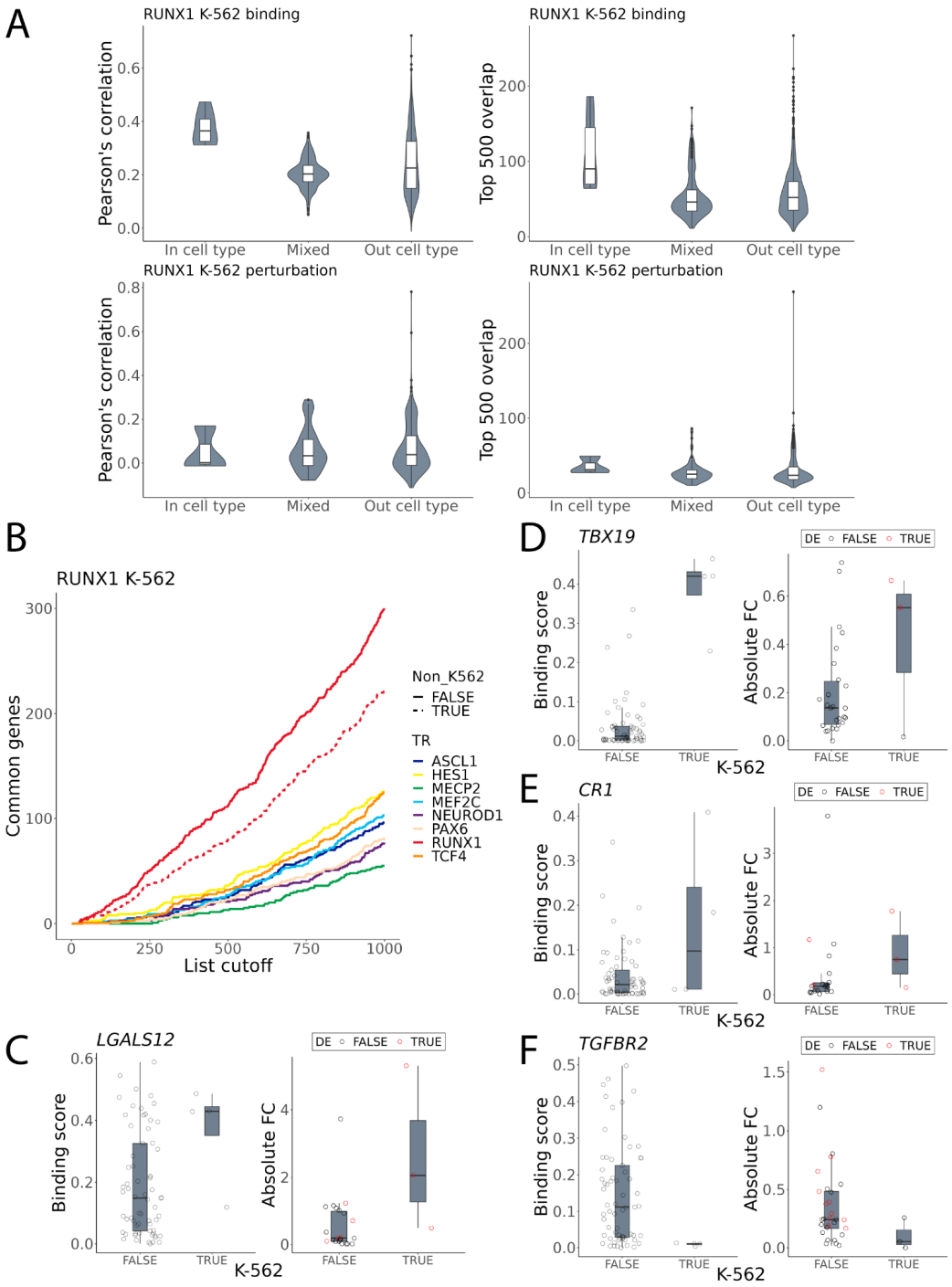


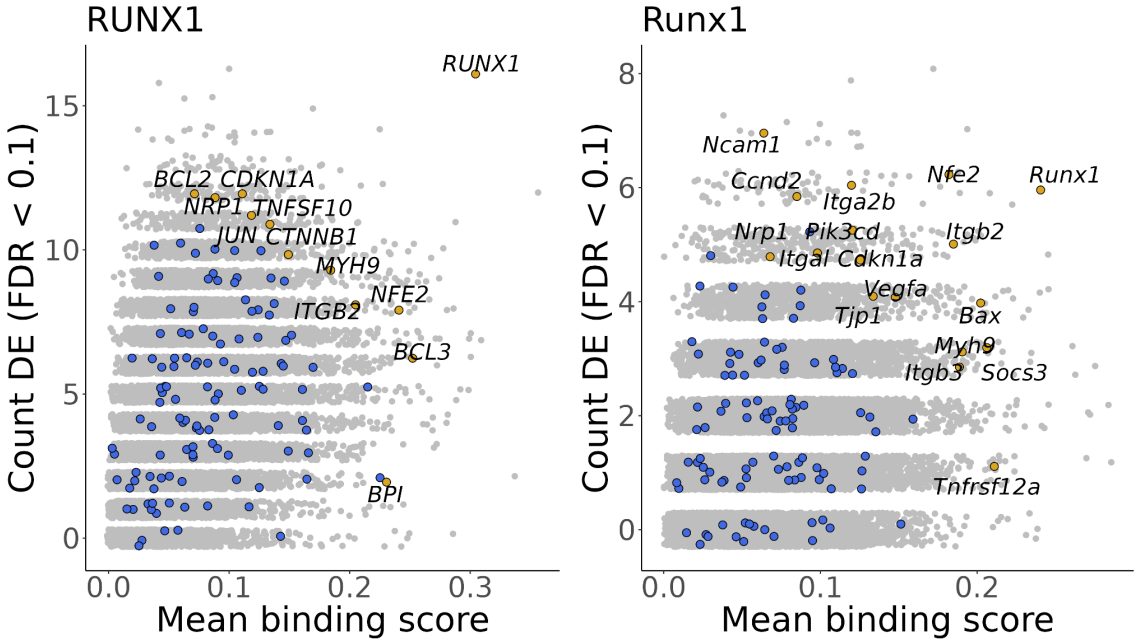
Figure S22. Context-specific analysis of RUNX1 in the K-562 context. (A) Top panel shows similarity for human RUNX1 ChIP-seq experiments (n=4), grouped by pairs in the K-562 cell line, K-562 experiments paired with other cell types (Mixed), or non-K-562 pairs. Perturbation experiments (n=3) are shown in the bottom panel. (B) Count of common genes between the RUNX1 K-562 integrated rankings and other rankings as a function of list cutoff. The dashed line shows the RUNX1 non-K-562 ranking. (C-F) Gene binding scores and absolute fold changes across RUNX1 experiments in or out of the K-562 cell type for (C) the top ranked gene for K-562; (D) the top ranked gene in K-562 relative to non-K562; (E) the top ranked curated gene in K-562 relative to non-K-562; (F) the top ranked gene in non-K-562 relative to K-562.

Overview of TR-targets

In the following section, we provide examples of candidate targets identified for each TR and the corresponding integrated ranking in mouse and human. We draw attention to the curated targets with the highest rankings, as well as examples of genes with strong evidence but limited representation in our search of the literature. Each TR is accompanied by the same style of plot as Fig. 6G, with the left panel showing human targets and mouse in the right panel. Each point is a gene, where blue points represent curated targets, and gold points represent curated targets that also have a high integrated ranking (here arbitrarily selected as those within the top 500).

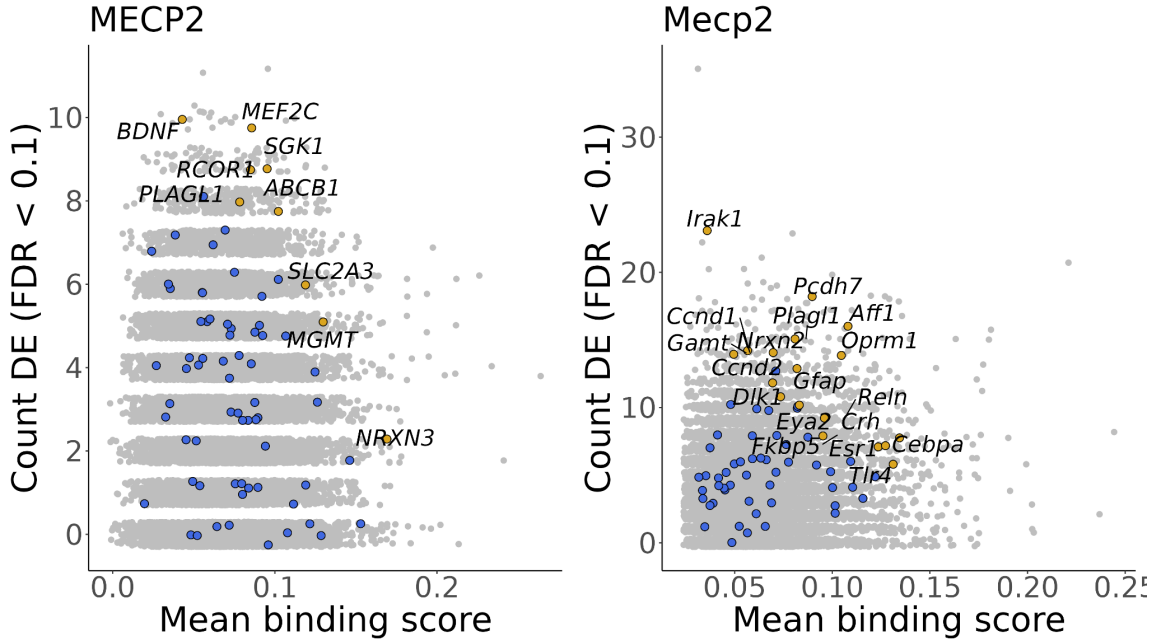
RUNX1 was the best represented TR overall, with many of the data sets from this collection coming from either blood progenitors or blood cancer cell lines. Correspondingly, we found examples of hematopoietic genes with curated evidence and a high integrated ranking, such as the erythroid TF *NFE2* (human: 109th, mouse: 15th; Wang et al., 2010) and immune adhesion subunit *ITGB2* (human: 179th, mouse: 26th; Puig-Kröger et al., 2003). We also found many genes involved in cellular motility and cytoskeletal functionality among the top ranked. Some, like myosin heavy chain 9 (*MYH9;* human: 282nd, mouse: 230th), had curated evidence (Bluteau et al., 2012). Most however did not, such as plectin (*PLEC*; human: 2nd, mouse: 24th), noted to be frequently implicated in cancer (Perez et al., 2021*)*, as well as lamin A/C (*LMNA*; human: 1,387th, mouse: 34th), vimentin (*VIM;* human: 6th, mouse: 17th), and paxillin (*PXN;* human: 129th, mouse: 321st).

We caution that the aggregate rankings did not always prioritize curated interactions, such as the important hematopoietic TF *SPI1* (human: 2,378th, mouse: 1,544th; Huang et al., 2008). *SPI1* had moderately elevated RUNX1 binding in both species, but weaker perturbation evidence lowered its integrated rankings. *SPI1* was also not prioritized in the minimum ranking implementation (human: 1,160th, mouse: 3,107th). We further note that RUNX1 was ranked first as its own target in both species. While autoregulatory functionality has been shown for this TR (Nottingham et al., 2012), the rankings will prioritize each respective TR by virtue of these genes being frequently DE in their perturbation experiments. Nevertheless, the rankings also highlighted orthologous genes that had strong evidence in both species but are rarely studied, such as *MOB3A (*human: 11th, mouse: 104th), believed to be involved in the angiogenic Hippo pathway-YAP/TAZ (Boopathy and Hong 2019; Dutchak et al., 2022). Finally, we highlight that *PHF19* (human: 5th) was recently identified in a crowd-sourced competition as an aggressive marker of multiple myeloma (Mason et al., 2020), suggesting further potential oncogenic activity of RUNX1 misregulation.



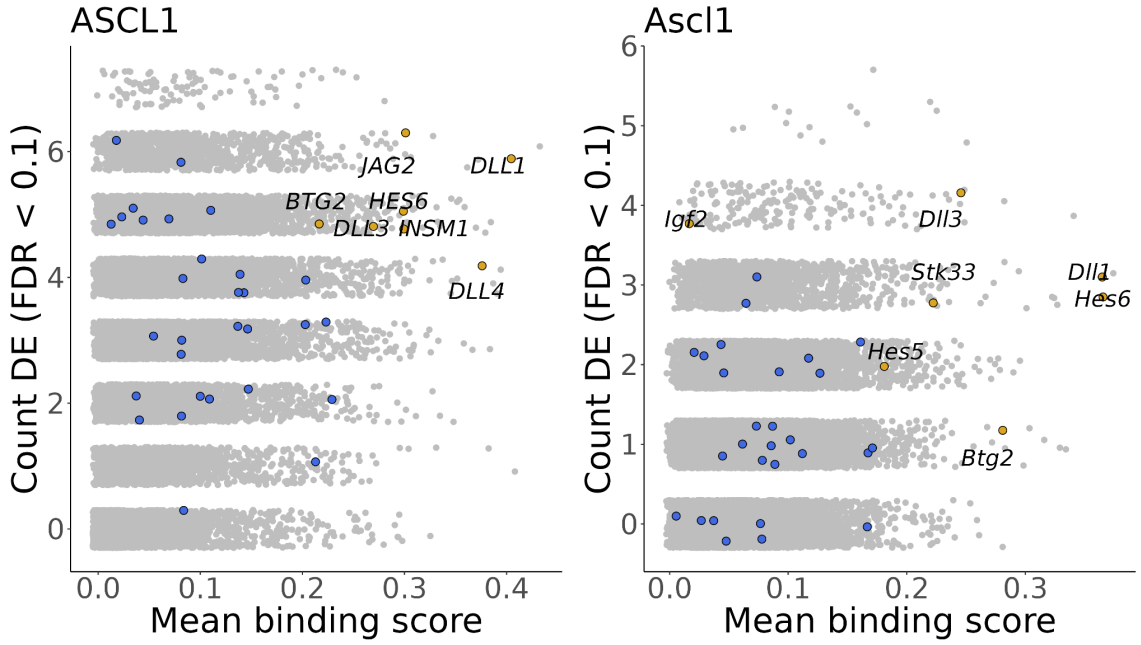
MECP2 was the next most represented TR and had the most brain-relevant experiments. While the TSS-based logic of the binding score metric is likely better suited to the TFs, the rankings still provided relevant examples. *SDK1* (human: 1,877th, mouse: 10th) and *AUTS2* (human: 6,905th, mouse: 15th) both ranked highly in mouse. A recent murine MECP2 study demonstrated that these genes are in a contacting chromosomal domain enriched for methylated cytosines (Clemens et al., 2020), adding a further line of evidence beyond the scope of this study. *SDK1* has been studied for its role in retinal development and cocaine addiction behavior (Scobie et al., 2014; Yamagata and Sanes 2019), while the connection between *MECP2* and *AUTS2* in particular has garnered interest for its involvement in neurodevelopmental and psychiatric outcomes (Pang et al., 2021).

The top curated target by human rankings was *BDNF* (human: 38th, mouse: 3,234th; Abuhatzira et al., 2007), although this was driven almost exclusively by the perturbation evidence. In mouse it was *AFF1* (human: 16,547th, mouse: 26th; Urdinguio et al., 2008), which was well supported by both data types. We also note that the curated target *IRAK1* (Urdinguio et al., 2008), which neighbours the *MECP2* loci, had excellent perturbation evidence in mouse but essentially no binding evidence in our framework. It has been posited that *MECP2* regulates *IRAK1* through both indirect (Urdinguio et al., 2010) and direct (Kishi et al., 2016) means. Examples of candidates without curated evidence include the rarely studied *CDS1* (human: 77th, mouse: 68th), which is involved in lipid droplet growth (D'Souza et al., 2014; Xu et al., 2019). Interestingly, aberrant lipid growth has been documented in murine Rett syndrome models (Vashi et al., 2021). Examples of high ranking candidates that are relevant to brain health but lack curated evidence include *SHANK2* (human: 778th, mouse: 222nd; Zaslavsky et al., 2020), *PLXNA2* (human: 81st, mouse: 308th; Kong et al., 2016; Pijuan et al., 2021), *CNTNAP2* (human: 1,566th, mouse: 81st; Xing et al., 2019), *TENM2* (human: 2,588th, mouse: 257th; Wray et al., 2018), and *TENM3* (human: 6,614th, mouse: 14th; Singh et al., 2019).

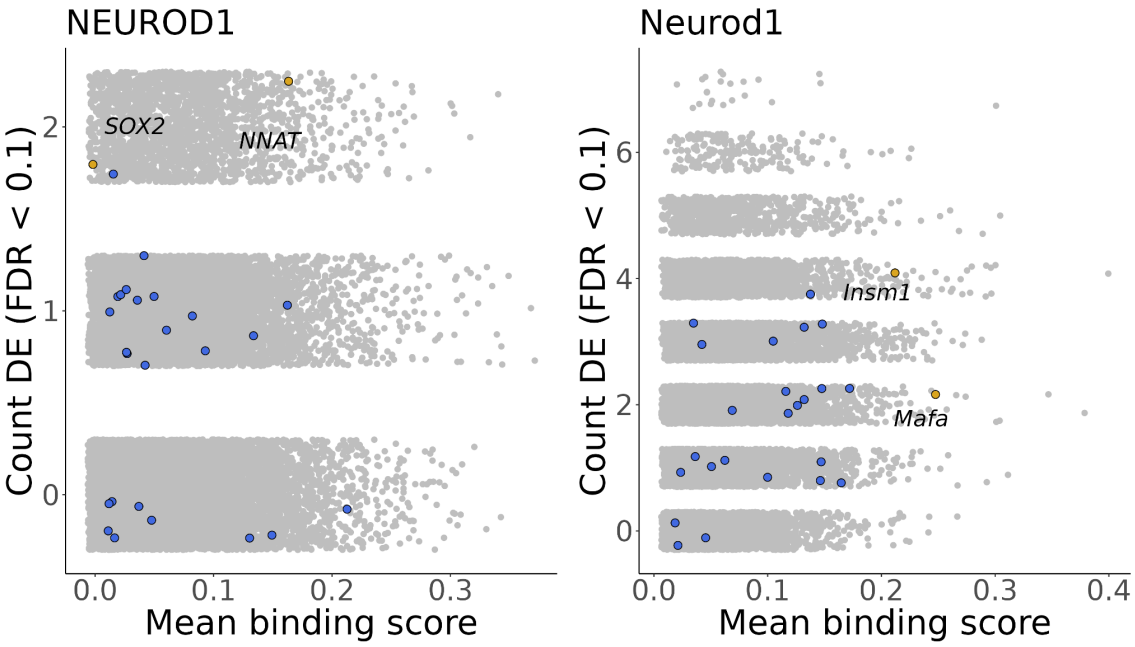


ASCL1 experiments can be broadly characterized as deriving from embryonic systems (some from primary tissues) in mouse, while human experiments were overwhelmingly carried out in cancer cell lines. Highly ranked and previously described targets include Notch pathway effectors *DLL1* (human: 1st, mouse: 7th), *DLL3* (human: 96th, mouse: 9th), *DLL4* (human: 36th, mouse: 1,218th), *HES6* (human: 73rd, mouse: 4th) and *JAG2* (human: 9th, mouse: 4,720th) (Castro et al., 2006; Henke et al., 2009; Johansson et al., 2009; Nelson et al., 2009; Castro et al., 2011; Augustyn et al., 2014; Misra et al., 2014). Notch effector *LFNG* was also highly ranked (human: 13th, mouse: 96th) but was not in the curated resources. Similarly, the cell-cycling *CDC25B* had robust evidence in both data types for both species (human: 5th, mouse: 2nd); this gene was not in the curated resource but previously had been noted as a candidate target (Castro et al., 2006; Castro et al., 2011).

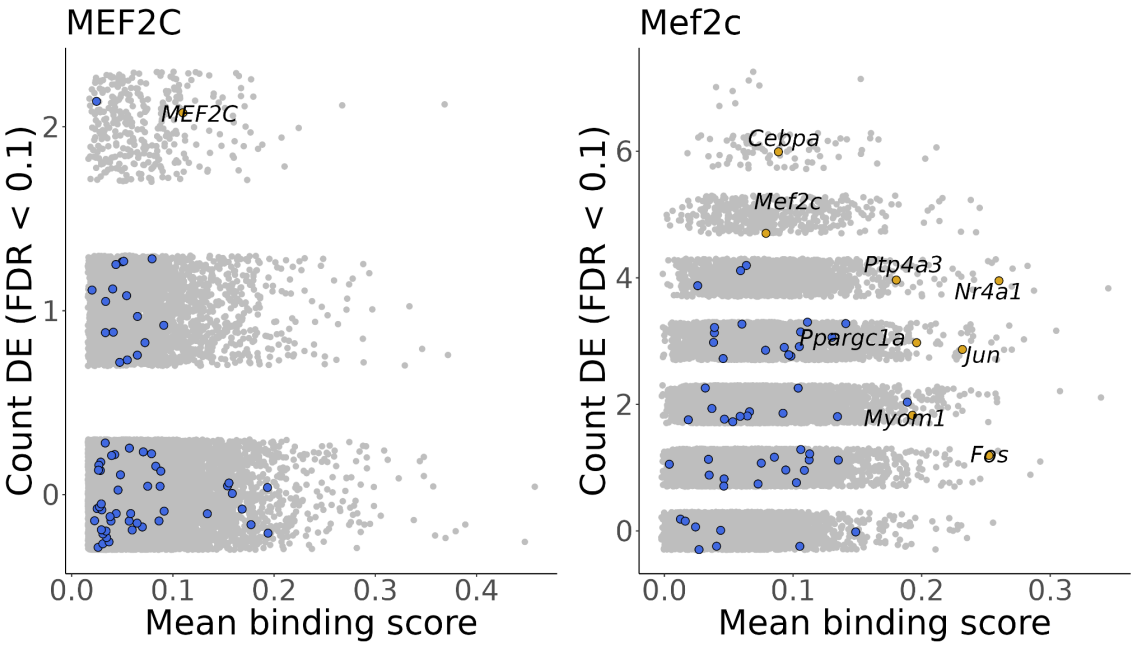
Other examples of candidates lacking curated evidence but have been putatively linked to ASCL1 include the repressive TFs *ID1* (human: 14th, mouse: 27th) and *ID3* (human: 19th, mouse: 14th) (Jorstad et al., 2020), as well as *ZBTB18* (human: 33rd, mouse: 18th) (Wapinski et al., 2013). Our work supports previous connections made between ASCL1 and the transcription repressor *CBFA2T3* (human: 524th, mouse: 3rd) (Alishahi et al., 2009; Aaker et al., 2010), and novelly links ASCL1 to the epilepsy-implicated potassium channel gene *KCNH2* (human: 7th, mouse: 113th) (Bagnall et al., 2016 & 2017).



NEUROD1 experiments were typically conducted in embryonic stem cells, pancreatic tissues, or cancer cell lines. This TR had the second fewest targets in the curated resource, and only a handful of these also had high integrated ranks in one species alone, such as *ISNM1* (human: 5,556th, mouse: 90th; Breslin et al., 2003). This gene is active in both neural and endocrine systems, and we found other examples of candidate targets that implicate NEUROD1's multi-tissue activity. *SRRM4* (human: 137th, mouse: 36th) encodes a crucial neurodevelopmental splicing factor recently implicated as contributing to dysregulated splicing in islets cells in a murine diabetic model (Quesnel-Vallières et al., 2015; Wilhelmi et al., 2021). Congenital disorders in Golgi functionality gene *COG1* (human: 312th, mouse: 53rd) are associated with hypoglycemia and neurological impairment (Climer et al., 2018; Zhang et al., 2018; Huang et al., 2021). *CXXC4* (human: 5,465th, mouse: 4th) has evidence supporting roles in beta cell differentiation and neuronal stem cell proliferation (Ahn et al., 2004; Guan et al., 2020), while *STXBP1* (human: 1,478th, mouse: 146th) has been linked to encephalopathies and enteroendocrine functionality (Stamberger et al., 2016; Campbell et al., 2020).

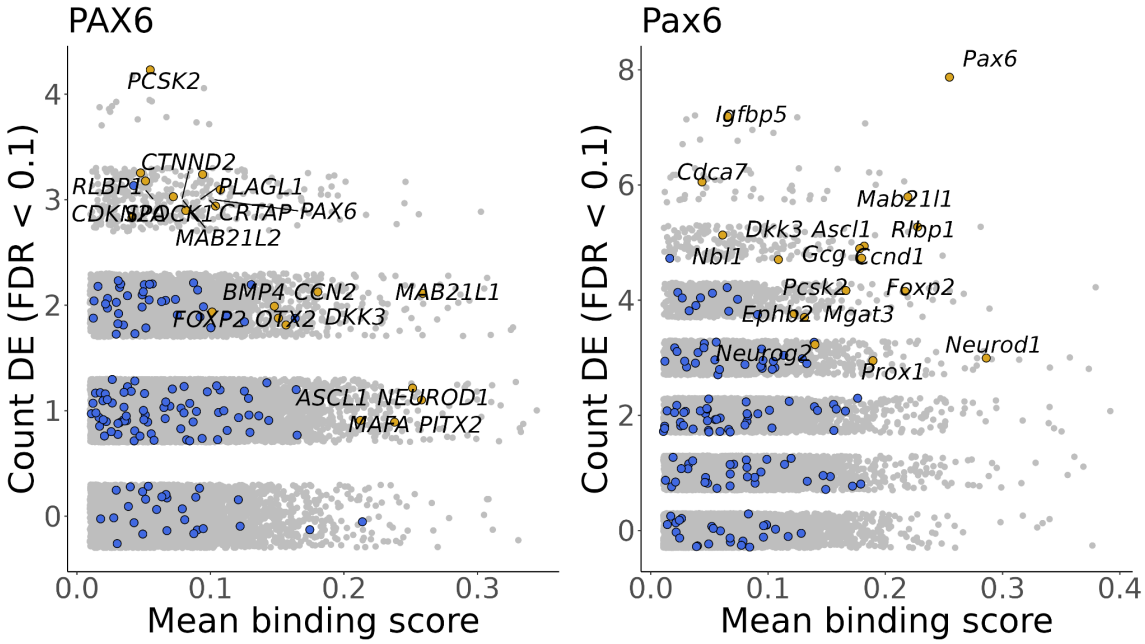


MEF2C heavily skewed towards mouse for both data types, with none of the human experiments conducted in neuronal contexts. Only a single curated target in human had a high integrated ranking, which was *MEF2C* itself (human: 3rd, mouse: 276th; Escher et al., 2011), while the top curated target in mouse was *NR4A1* (human: 1,458th, mouse: 10th; Nagel et al., 2008). *HDAC5* (human: 165th, mouse: 92nd), *HDAC9* (human: 59th, mouse: 95th), and *MEF2D* (human: 26th, mouse: 165th) were ranked highly in both species. Prior work has well established the interplay between HDACs and MEF2 TFs across multiple systems (Haberland et al., 2009), adding further support that our framework can prioritize known biology. Kruppel-like factors *KLF6* (human: 22nd, mouse: 5th) and *KLF2* (human: 135th, mouse: 41st) were similarly highly ranked in both species, while *KLF4* was high in mouse (human: 3,640th, mouse: 73rd). While absent from the curated resources, prior studies have nominated these interactions (Salma and McDermott 2012; Xu et al., 2015; Lu et al., 2021). Finally, we highlight *ARID1A* (human: 2nd, mouse: 281st) which encodes a DNA-binding SWI/SNF chromatin remodeler that was recently demonstrated to physically interact with MEF2C and regulate embryonic cardiac versus neuronal differentiation (Liu et al., 2020). Our work further suggests that MEF2C regulates expression of *ARID1A* in addition to interacting with its protein product.



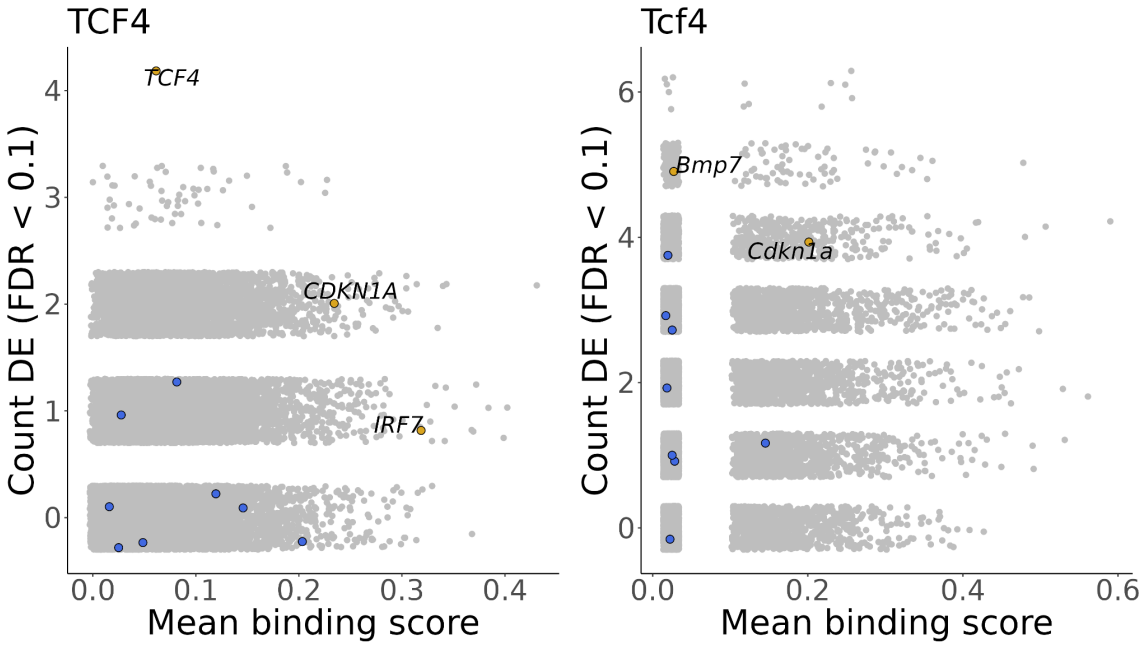
PAX6 also had more data in mouse than human, with experiments generally derived from retinal, cortical, or pancreatic tissues. After *PAX6* itself (Sun et al., 2015), high ranking curated targets included *MAB21L1* (human: 6th, mouse: 5th; Sun et al., 2015), which encodes a nucleotidyltransferase noted for its “exceptional” conservation and role in eye development (de Oliveira Mann et al., 2016). *MAB21L2* meanwhile was a curated target (Wolf et al., 2009) that ranked highly in human (68th) but not mouse (9,157th). Retinaldehyde binding protein 1 (*RLBP1*; human: 119th, mouse: 7th; Boppana et al., 2012) was another highly ranked and curated target with defined functionality in the eye.

Examples of top human candidates absent from the curation resource includes *EPHA3* (human: 21st, mouse: 4,613th), which encodes an ephrin receptor kinase involved in retinotectal mapping (Lemke and Reber 2005), and *DNAJB6* (human: 1st, mouse: 2,342nd), which produces a chaperone found to play a critical role in preventing neuronal protein aggregation (Thiruvalluvan et al., 2020). The lead mouse target was the biliverdin reductase-encoding *BLVRA* (human: 14,453rd, mouse: 2nd), with contributions from pancreatic as well as neuronal experiments. *BLVRA* was also the most significant gene in the mouse binding specificity analysis (Fig. S4B). Other uncurated candidates include the neuronal marker *MAP2* (human: 143rd; mouse: 71st), the poorly characterized *ZNF608* (human: 50th, mouse: NA), and *ABHD4* (human: 685th, mouse: 65th), whose lysophospholipase product was described as protective against the inappropriate delamination and migration of *Pax6*-expressing murine radial glial progenitor cells (László and Lele et al., 2020).



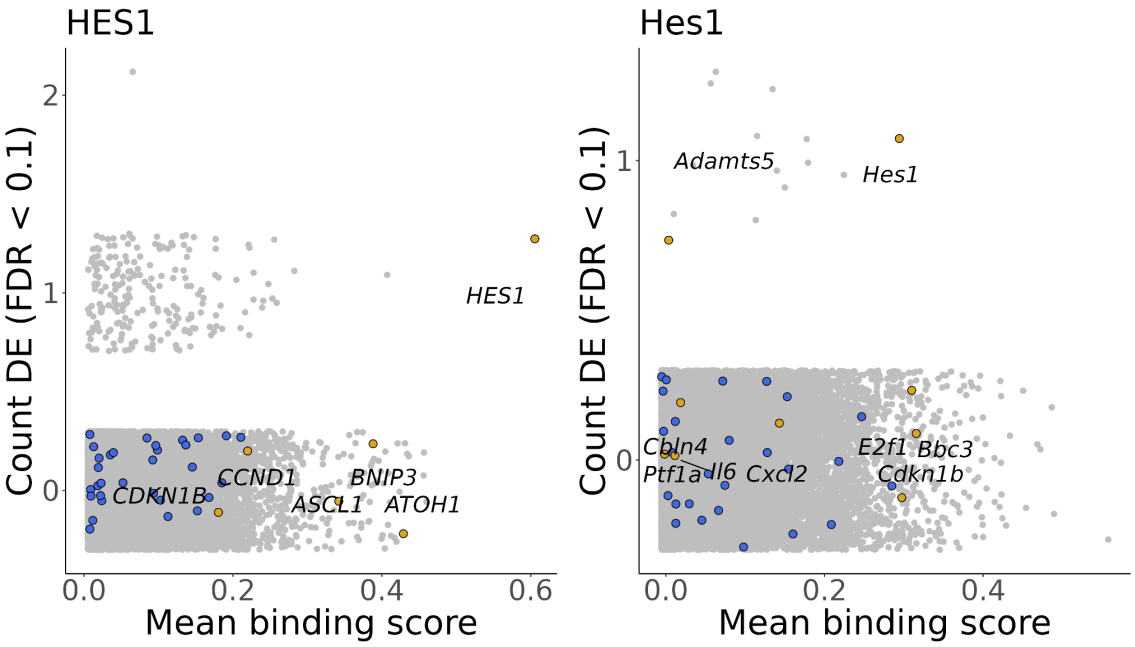
TCF4 perturbation experiments for both species covered a range of neuronal contexts, as well as experiments in blood and kidney systems. Only a single mouse Tcf4 ChIP-seq experiment was identified (a neural stem cell line; Moen et al., 2017), while the human ChIP-seq experiments tended to be from blood cancer and neuroblastoma cell lines. TCF4 was the most difficult TR to curate given the widespread prevalence of the *TCF4* symbol being used for *TCF7L2* (also known as T-Cell Factor 4), a TF commonly studied in Wnt signalling and colon cancer. A warning of this potential for confusion is also found in Papes et al., 2022. We suspected that this naming "contamination" affected other literature curated resources, and after screening all candidates only eleven *TCF4* targets were recovered.

TCF4 (human: 8th, mouse: 943rd) has been shown to have autoregulatory functionality in murine dendritic cell differentiation (Grajkowska et al., 2017). However, binding was low in human and not detected in the single mouse Tcf4 ChIP-seq data set, unlike for the curated target and cyclin-dependent kinase inhibitor encoding *CDKN1A* (human: 21st, mouse: 369th; Deliri et al., 2011). Lead human candidates include *DEPP1* (human: 1st, mouse; 10,999th), an autophagy regulator gene, while the microfibril-associated *MFAP4* (human: 3rd, mouse: 19th) ranked highly in both species, as did the neurodevelopmental TF gene *ZBTB18* (human: 4th, mouse: 35th). Noting that the binding was driven by a single experiment, mouse was led by *Gpsm2* (human: 5,569th, mouse: 1st), affiliated with a human hearing phenotype (Doherty et al., 2012), and the melanocortin receptor *Mc4r* (human: 15,179th, mouse: 2nd), commonly studied in appetite control (Asai et al., 2013). Finally, *TLCD1* (human: 27th, mouse: 3rd) was a top candidate in both species, with the sparse literature for this gene pointing to its role in membrane fluidity regulation (Ruiz et al., 2018).



HES1 ChIP-seq data was scarce and the perturbation experiments had relatively few DE genes (only 223 genes were DE at least once across all 9 experiments in both species; Fig. 4B), leading to less consistent aggregated evidence. Lymphotoxin beta *LTB* (human: 8th, mouse: 3,213th) was the only gene to be DE more than once in the human HES1 experiments, with binding signal in the K562 but not HepG2 or MCF7 ChIP-seq experiments. *HES1* (human: 1st, mouse: 1st) was the top curated target in both species, and has been characterized as having auto-repressive functionality (Havrda et al., 2008). *ATOH1* (human: 50th, mouse: 4,923th) was a curated target that was also the most significant gene in the HES1 specific binding analysis (Fig. 2B), but the perturbation evidence was negligible.

Of the human genes with one DE count, the mitochondrial lipid transporter-encoding *STARD7* (human: 2nd, mouse: 7,311th) had the strongest binding score, while TF gene *E2f5* (human: 1,085th, mouse: 4th) had the strongest bind score among mouse genes with one DE count. The neurodegenerative-associated Prolifin (*Pfn1*; human: 134th, mouse: 5th; Yang et al., 2016) had the strongest binding signal in mouse – while it was not DE in the mouse studies, *PFN1* was DE once in human. Finally, *BAHCC1* (human: 34th, mouse: 327th), which encodes a repressive histone mark reader (Fan et al., 2020), was highly bound in both species, as was the cell-cycle-associated *FBXO31* (human: 48th, mouse: 129th).



Supplemental citations

Aaker JD, Patineau AL, Yang HJ, et al. Interaction of MTG family proteins with NEUROG2 and ASCL1 in the developing nervous system. *Neurosci Lett*. 2010;474(1):46-51. doi:10.1016/j.neulet.2010.03.004

Abuhatzira L, Makedonski K, Kaufman Y, Razin A, Shemer R. MeCP2 deficiency in the brain decreases BDNF levels by REST/CoREST-mediated repression and increases TRKB production. *Epigenetics*. 2007;2(4):214-222. doi:10.4161/epi.2.4.5212

Ahn JI, Lee KH, Shin DM, et al. Comprehensive transcriptome analysis of differentiation of embryonic stem cells into midbrain and hindbrain neurons. *Dev Biol*. 2004;265(2):491-501. doi:10.1016/j.ydbio.2003.09.041

Alishahi A, Koyano-Nakagawa N, Nakagawa Y. Regional expression of MTG genes in the developing mouse central nervous system. *Dev Dyn*. 2009;238(8):2095-2102. doi:10.1002/dvdy.22021

Asai M, Ramachandrappa S, Joachim M, et al. Loss of function of the melanocortin 2 receptor accessory protein 2 is associated with mammalian obesity. *Science*. 2013;341(6143):275-278. doi:10.1126/science.1233000

Augustyn A, Borromeo M, Wang T, et al. ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. *Proc Natl Acad Sci U S A*. 2014;111(41):14788-14793. doi:10.1073/pnas.1410419111

Bagnall RD, Crompton DE, Petrovski S, et al. Exome-based analysis of cardiac arrhythmia, respiratory control, and epilepsy genes in sudden unexpected death in epilepsy: Exome Sequencing Analysis of SUDEP. *Ann Neurol*. 2016;79(4):522-534. doi:10.1002/ana.24596

Bagnall RD, Crompton DE, Semsarian C. Genetic Basis of Sudden Unexpected Death in Epilepsy. *Front Neurol*. 2017;8:348. doi:10.3389/fneur.2017.00348

Bai X, Shi S, Ai B, et al. ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res*. 2020;48(D1):D51-D57. doi:10.1093/nar/gkz973

Bluteau D, Glembotsky AC, Raimbault A, et al. Dysmegakaryopoiesis of FPD/AML pedigrees with constitutional RUNX1 mutations is linked to myosin II deregulated expression. *Blood*. 2012;120(13):2708-2718. doi:10.1182/blood-2012-04-422337

Boopathy GTK, Hong W. Role of Hippo Pathway-YAP/TAZ Signaling in Angiogenesis. *Front Cell Dev Biol*. 2019;7:49. doi:10.3389/fcell.2019.00049

Boppana S, Scheglov A, Geffers R, Tarabykin V. Cellular retinaldehyde-binding protein (CRALBP) is a direct downstream target of transcription factor Pax6. *Biochim Biophys Acta*. 2012;1820(2):151-156. doi:10.1016/j.bbagen.2011.09.015

Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*. 2012;13:405. doi:10.1186/1471-2164-13-405

Breslin MB, Zhu M, Lan MS. NeuroD1/E47 regulates the E-box element of a novel zinc finger transcription factor, IA-1, in developing nervous system. *J Biol Chem*. 2003;278(40):38991-38997. doi:10.1074/jbc.M306795200

Campbell JR, Martchenko A, Sweeney ME, et al. Essential Role of Syntaxin-Binding Protein-1 in the Regulation of Glucagon-Like Peptide-1 Secretion. *Endocrinology*. 2020;161(5):bqaa039. doi:10.1210/endocr/bqaa039

Carrasco Pro S, Dafonte Imedio A, Santoso CS, et al. Global landscape of mouse and human cytokine transcriptional regulation. *Nucleic Acids Res*. 2018;46(18):9321-9337. doi:10.1093/nar/gky787

Castro DS, Martynoga B, Parras C, et al. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev*. 2011;25(9):930-945. doi:10.1101/gad.627811

Castro DS, Skowronska-Krawczyk D, Armant O, et al. Proneural bHLH and Brn Proteins Coregulate a Neurogenic Program through Cooperative Binding to a Conserved DNA Motif. *Developmental Cell*. 2006;11(6):831-844. doi:10.1016/j.devcel.2006.10.006

Chu ECP, Morin A, Chang THC, et al. Experiment level curation of transcriptional regulatory interactions in neurodevelopment. *PLOS Computational Biology*. 2021;17(10):e1009484. doi:10.1371/journal.pcbi.1009484

Clemens AW, Wu DY, Moore JR, Christian DL, Zhao G, Gabel HW. MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. *Molecular Cell*. 2020;77(2):279-293.e8. doi:10.1016/j.molcel.2019.10.033

Climer LK, Hendrix RD, Lupashin VV. Conserved Oligomeric Golgi and Neuronal Vesicular Trafficking. *Handb Exp Pharmacol*. 2018;245:227-247. doi:10.1007/164_2017_65

D'Souza K, Kim YJ, Balla T, Epand RM. Distinct properties of the two isoforms of CDP-diacylglycerol synthase. *Biochemistry*. 2014;53(47):7358-7367. doi:10.1021/bi501250m

de Oliveira Mann CC, Kiefersauer R, Witte G, Hopfner KP. Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein MAB21L1. *Sci Rep*. 2016;6(1):27498. doi:10.1038/srep27498

Deliri H, Meller N, Kadakkal A, et al. Increased 12/15-lipoxygenase enhances cell growth, fibronectin deposition, and neointimal formation in response to carotid injury. *Arterioscler Thromb Vasc Biol*. 2011;31(1):110-116. doi:10.1161/ATVBAHA.110.212068

Doherty D, Chudley AE, Coghlan G, et al. GPSM2 mutations cause the brain malformations and hearing loss in Chudley-McCullough syndrome. *Am J Hum Genet*. 2012;90(6):1088-1093. doi:10.1016/j.ajhg.2012.04.008

Dutchak K, Garnett S, Nicoll M, de Bruyns A, Dankort D. MOB3A Bypasses BRAF and RAS Oncogene-Induced Senescence by Engaging the Hippo Pathway. *Mol Cancer Res*. 2022;20(5):770-781. doi:10.1158/1541-7786.MCR-21-0767

Escher P, Schorderet DF, Cottet S. Altered expression of the transcription factor Mef2c during retinal degeneration in Rpe65-/- mice. *Invest Ophthalmol Vis Sci*. 2011;52(8):5933-5940. doi:10.1167/iovs.10-6978

Essaghir A, Toffalini F, Knoops L, Kallin A, van Helden J, Demoulin JB. Transcription factor regulation can be accurately predicted from the presence of target gene signatures in microarray gene expression data. *Nucleic Acids Res*. 2010;38(11):e120. doi:10.1093/nar/gkq149

Fan H, Lu J, Guo Y, et al. BAHCC1 binds H3K27me3 via a conserved BAH module to mediate gene silencing and oncogenesis. *Nat Genet*. 2020;52(12):1384-1396. doi:10.1038/s41588-020-00729-3

Grajkowska LT, Ceribelli M, Lau CM, et al. Isoform-Specific Expression and Feedback Regulation of E Protein TCF4 Control Dendritic Cell Lineage Specification. *Immunity*. 2017;46(1):65-77. doi:10.1016/j.immuni.2016.11.006

Guan B, Zhan Z, Wang L, Wang L, Liu L. CXXC4 mediates glucose-induced β-cell proliferation. *Acta Diabetol*. 2020;57(9):1101-1109. doi:10.1007/s00592-020-01525-5

Han H, Cho JW, Lee S, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*. 2018;46(D1):D380-D386. doi:10.1093/nar/gkx1013

Havrda MC, Harris BT, Mantani A, et al. Id2 is required for specification of dopaminergic neurons during adult olfactory neurogenesis. *J Neurosci*. 2008;28(52):14074-14086. doi:10.1523/JNEUROSCI.3188-08.2008

Henke RM, Meredith DM, Borromeo MD, Savage TK, Johnson JE. Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. *Dev Biol*. 2009;328(2):529-540. doi:10.1016/j.ydbio.2009.01.007

Huang G, Zhang P, Hirai H, et al. PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat Genet*. 2008;40(1):51-60. doi:10.1038/ng.2007.7

Huang Y, Dai H, Yang G, Zhang L, Xue S, Zhu M. Component of oligomeric Golgi complex 1 deficiency leads to hypoglycemia: a case report and literature review. *BMC Pediatr*. 2021;21(1):442. doi:10.1186/s12887-021-02922-7

Johansson TA, Westin G, Skogseid B. Identification of Achaete-scute complex-like 1 (ASCL1) target genes and evaluation of DKK1 and TPH1 expression in pancreatic endocrine tumours. *BMC Cancer*. 2009;9:321. doi:10.1186/1471-2407-9-321

Jorstad NL, Wilken MS, Todd L, et al. STAT Signaling Modifies Ascl1 Chromatin Binding and Limits Neural Regeneration from Muller Glia in Adult Mouse Retina. *Cell Reports*. 2020;30(7):2195-2208.e5. doi:10.1016/j.celrep.2020.01.075

Kim JH, Lee S, Rho JK, Choe SY. AML1, the target of chromosomal rearrangements in human leukemia, regulates the expression of human complement receptor type 1 (CR1) gene. *The International Journal of Biochemistry & Cell Biology.* 1999;31(9):933-940. doi:10.1016/S1357-2725(99)00048-5

Kishi N, MacDonald JL, Ye J, Molyneaux BJ, Azim E, Macklis JD. Reduction of aberrant NF-κB signalling ameliorates Rett syndrome phenotypes in Mecp2-null mice. *Nat Commun*. 2016;7(1):10520. doi:10.1038/ncomms10520

Kong Y, Janssen BJC, Malinauskas T, et al. Structural Basis for Plexin Activation and Regulation. *Neuron*. 2016;91(3):548-560. doi:10.1016/j.neuron.2016.06.018

Lemke G, Reber M. RETINOTECTAL MAPPING: New Insights from Molecular Genetics. *Annu Rev Cell Dev Biol*. 2005;21(1):551-580. doi:10.1146/annurev.cellbio.20.022403.093702

Lesurf R, Cotto KC, Wang G, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res*. 2016;44(D1):D126-132. doi:10.1093/nar/gkv1203

Lu YW, Martino N, Gerlach BD, et al. MEF2 (Myocyte Enhancer Factor 2) Is Essential for Endothelial Homeostasis and the Atheroprotective Gene Expression Program. *ATVB*. 2021;41(3):1105-1123. doi:10.1161/ATVBAHA.120.314978

Lynn DJ, Winsor GL, Chan C, et al. InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol Syst Biol*. 2008;4:218. doi:10.1038/msb.2008.55

Martin CC, Svitek CA, Oeser JK, Henderson E, Stein R, O'Brien RM. Upstream stimulatory factor (USF) and neurogenic differentiation/beta-cell E box transactivator 2 (NeuroD/BETA2) contribute to islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IGRP) gene expression. Biochem J. 2003;371(Pt 3):675-686. doi:10.1042/BJ20021585

Mason MJ, Schinke C, Eng CLP, et al. Multiple Myeloma DREAM Challenge reveals epigenetic regulator PHF19 as marker of aggressive disease. *Leukemia*. 2020;34(7):1866-1874. doi:10.1038/s41375-020-0742-z

Misra K, Luo H, Li S, Matise M, Xiang M. Asymmetric activation of Dll4-Notch signaling by Foxn4 and proneural factors activates BMP/TGFβ signaling to specify V2b interneurons in the spinal cord. *Development*. 2014;141(1):187-198. doi:10.1242/dev.092536

Moen MJ, Adams HHH, Brandsma JH, et al. An interaction network of mental disorder proteins in neural stem cells. *Transl Psychiatry*. 2017;7(4):e1082. doi:10.1038/tp.2017.52

Nagel S, Meyer C, Quentmeier H, Kaufmann M, Drexler HG, MacLeod R a. F. MEF2C is activated by multiple mechanisms in a subset of T-acute lymphoblastic leukemia cell lines. *Leukemia*. 2008;22(3):600-607. doi:10.1038/sj.leu.2405067

Nelson BR, Hartman BH, Ray CA, Hayashi T, Bermingham-McDonogh O, Reh TA. Acheate-scute like 1 (Ascl1) is required for normal delta-like (Dll) gene expression and notch signaling during retinal development. *Dev Dyn*. 2009;238(9):2163-2178. doi:10.1002/dvdy.21848

Nottingham WT, Jarratt A, Burgess M, et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood*. 2007;110(13):4188-4197. doi:10.1182/blood-2007-07-100883

Pang W, Yi X, Li L, Liu L, Xiang W, Xiao L. Untangle the Multi-Facet Functions of Auts2 as an Entry Point to Understand Neurodevelopmental Disorders. *Front Psychiatry*. 2021;12:580433. doi:10.3389/fpsyt.2021.580433

Papes F, Camargo AP, de Souza JS, et al. Transcription Factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content. *Nat Commun*. 2022;13(1):2387. doi:10.1038/s41467-022-29942-w

Perez SM, Brinton LT, Kelly KA. Plectin in Cancer: From Biomarker to Therapeutic Target. *Cells*. 2021;10(9):2246. doi:10.3390/cells10092246

Pijuan J, Ortigoza-Escobar JD, Ortiz J, et al. *PLXNA2* and *LRRC40* as candidate genes in autism spectrum disorder. *Autism Research*. 2021;14(6):1088-1100. doi:10.1002/aur.2502

Puig-Kröger A, Sanchez-Elsner T, Ruiz N, et al. RUNX/AML and C/EBP factors regulate CD11a integrin expression in myeloid cells through overlapping regulatory elements. *Blood*. 2003;102(9):3252-3261. doi:10.1182/blood-2003-02-0618

Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev*. 2015;29(7):746-759. doi:10.1101/gad.256115.114

Ruiz M, Bodhicharla R, Svensk E, et al. Membrane fluidity is regulated by the C. elegans transmembrane protein FLD-1 and its human homologs TLCD1/2. *Elife*. 2018;7:e40686. doi:10.7554/eLife.40686

Salma J, McDermott JC. Suppression of a MEF2-KLF6 survival pathway by PKA signaling promotes apoptosis in embryonic hippocampal neurons. *J Neurosci*. 2012;32(8):2790-2803. doi:10.1523/JNEUROSCI.3609-11.2012

Scobie KN, Damez-Werno D, Sun H, et al. Essential role of poly(ADP-ribosyl)ation in cocaine action. *Proc Natl Acad Sci U S A*. 2014;111(5):2005-2010. doi:10.1073/pnas.1319703111

Singh B, Srivastava P, Phadke SR. Sequence variations in TENM3 gene causing eye anomalies with intellectual disability: Expanding the phenotypic spectrum. *European Journal of Medical Genetics*. 2019;62(1):61-64. doi:10.1016/j.ejmg.2018.05.004

Stamberger H, Nikanorova M, Willemsen MH, et al. *STXBP1* encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology*. 2016;86(10):954-962. doi:10.1212/WNL.0000000000002457

Sun J, Rockowitz S, Xie Q, Ashery-Padan R, Zheng D, Cvekl A. Identification of in vivo DNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. *Nucleic Acids Res*. 2015;43(14):6827-6846. doi:10.1093/nar/gkv589

Thiruvalluvan A, de Mattos EP, Brunsting JF, et al. DNAJB6, a Key Factor in Neuronal Sensitivity to Amyloidogenesis. *Molecular Cell*. 2020;78(2):346-358.e9. doi:10.1016/j.molcel.2020.02.022

Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419. doi:10.1126/science.1260419

Urdinguio RG, Fernández AF, Lopez-Nieva P, et al. Disrupted microRNA expression caused by Mecp2 loss in a mouse model of Rett syndrome. *Epigenetics*. 2010;5(7):656-663. doi:10.4161/epi.5.7.13055

Urdinguio RG, Lopez-Serra L, Lopez-Nieva P, et al. Mecp2-null mice provide new neuronal targets for Rett syndrome. *PLoS One*. 2008;3(11):e3669. doi:10.1371/journal.pone.0003669

Vashi N, Ackerley C, Post M, Justice MJ. Aberrant lung lipids cause respiratory impairment in a *Mecp2* -deficient mouse model of Rett syndrome. *Human Molecular Genetics*. 2021;30(22):2161-2176. doi:10.1093/hmg/ddab182

Wang W, Schwemmers S, Hexner EO, Pahl HL. AML1 is overexpressed in patients with myeloproliferative neoplasms and mediates JAK2V617F-independent overexpression of NF-E2. *Blood*. 2010;116(2):254-266. doi:10.1182/blood-2009-11-254664

Wapinski OL, Vierbuchen T, Qu K, et al. Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. *Cell*. 2013;155(3):621-635. doi:10.1016/j.cell.2013.09.028

Wilhelmi I, Neumann A, Jähnert M, Ouni M, Schürmann A. Enriched Alternative Splicing in Islets of Diabetes-Susceptible Mice. *IJMS*. 2021;22(16):8597. doi:10.3390/ijms22168597

Wolf LV, Yang Y, Wang J, et al. Identification of pax6-dependent gene regulatory networks in the mouse lens. *PLoS One*. 2009;4(1):e4159. doi:10.1371/journal.pone.0004159

Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

Xing X, Zhang J, Wu K, et al. Suppression of Akt-mTOR pathway rescued the social behavior in Cntnap2-deficient mice. *Sci Rep*. 2019;9(1):3041. doi:10.1038/s41598-019-39434-5

Xu Y, Mak HY, Lukmantara I, et al. CDP-DAG synthase 1 and 2 regulate lipid droplet growth through distinct mechanisms. *J Biol Chem*. 2019;294(45):16740-16755. doi:10.1074/jbc.RA119.009992

Xu Z, Yoshida T, Wu L, Maiti D, Cebotaru L, Duh EJ. Transcription factor MEF2C suppresses endothelial cell inflammation via regulation of NF-κB and KLF2. *J Cell Physiol*. 2015;230(6):1310-1320. doi:10.1002/jcp.24870

Yamagata M, Sanes JR. Expression and Roles of the Immunoglobulin Superfamily Recognition Molecule Sidekick1 in Mouse Retina. *Front Mol Neurosci*. 2019;11:485. doi:10.3389/fnmol.2018.00485

Yang C, Danielson EW, Qiao T, et al. Mutant PFN1 causes ALS phenotypes and progressive motor neuron degeneration in mice by a gain of toxicity. *Proc Natl Acad Sci USA*. 2016;113(41). doi:10.1073/pnas.1605964113

Yusuf D, Butland SL, Swanson MI, et al. The transcription factor encyclopedia. *Genome Biol*. 2012;13(3):R24. doi:10.1186/gb-2012-13-3-r24

Zaslavsky K, Zhang WB, McCready FP, et al. SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. *Nature Neuroscience*. 2019;22(4):556. doi:10.1038/s41593-019-0365-8

Zhang TY, Keown CL, Wen X, et al. Environmental enrichment increases transcriptional and epigenetic differentiation between mouse dorsal and ventral dentate gyrus. *Nature Communications*. 2018;9(1):298. doi:10.1038/s41467-017-02748-x