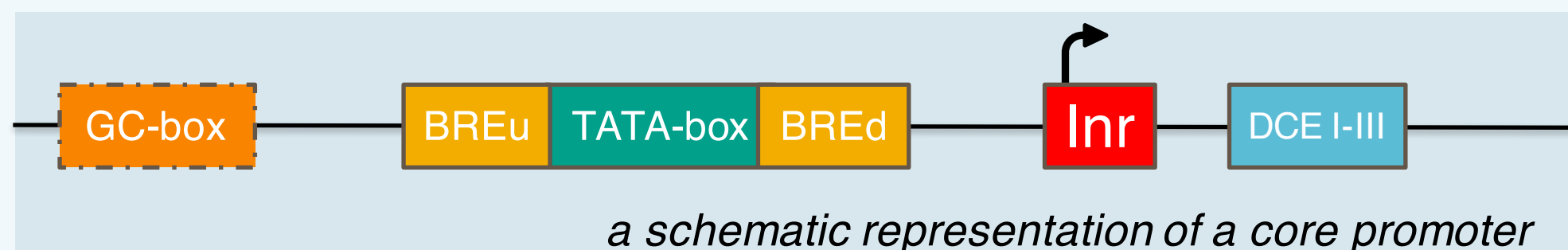


Core promoters

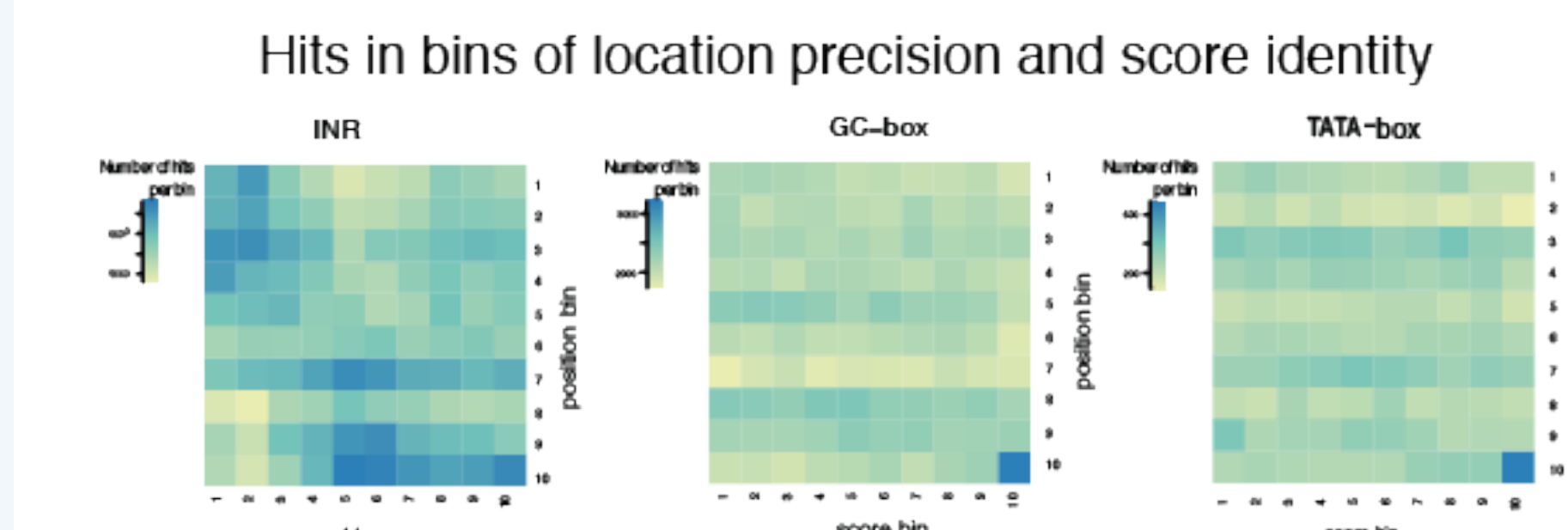
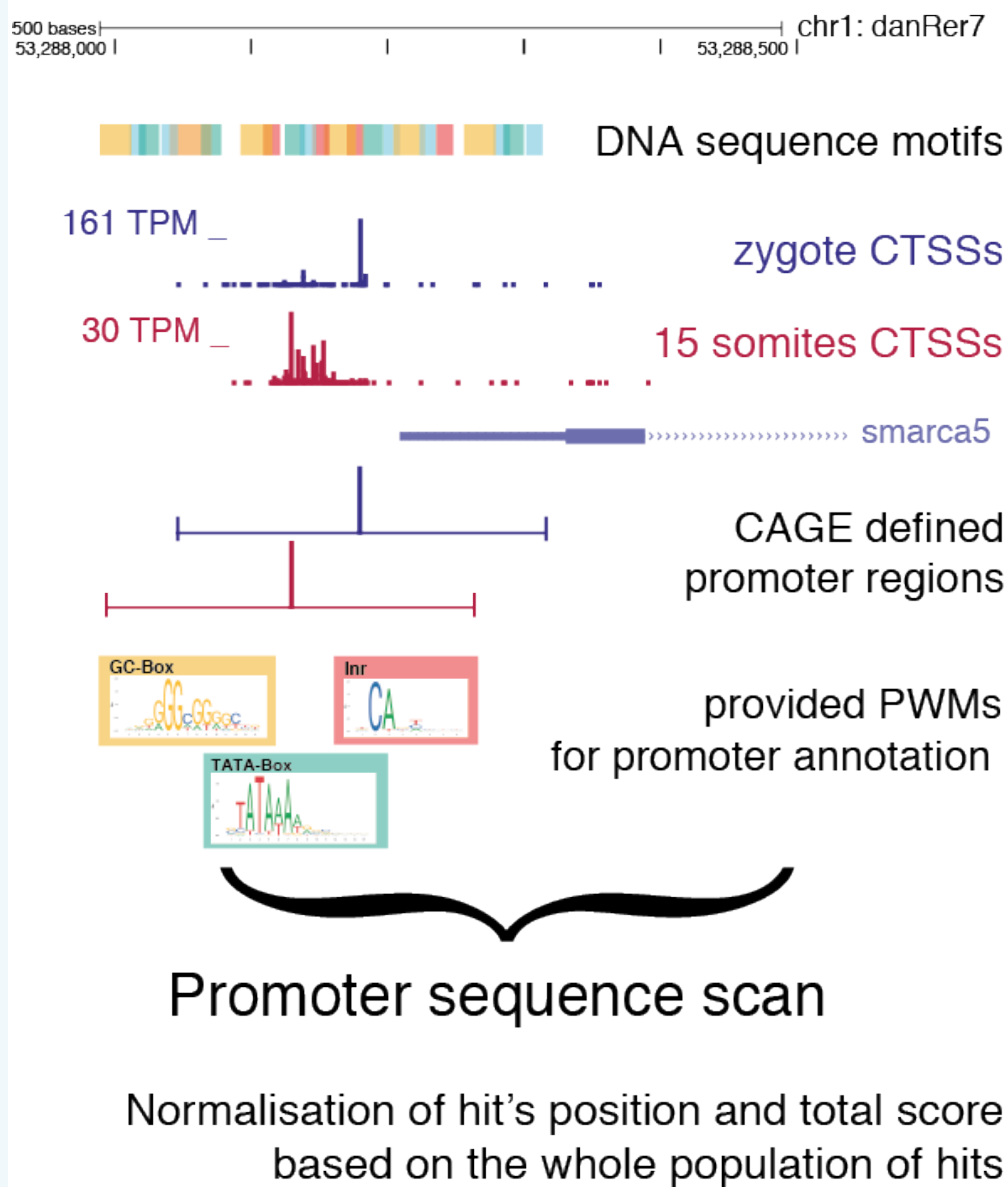


Core promoters are regions of DNA where integration of all regulatory inputs on the RNA pol II transcriptional complex occurs. They are highly diverse in terms of general transcription factors that are required to drive expression.

Distinct core promoter architectures are characteristic of at least three main promoter classes¹. By adding non-canonical core promoter features, we aim to sub-classify promoters and correlate them with their function. To do this, we developed **Promoter Ontology**, a tool that annotates promoters and reports significant promoter features for genes of interest.

Promoter annotation workflow

- An example of feature annotation on a promoter region of gene *smarca5* (two different promoter regions correspond to two promoter architectures)

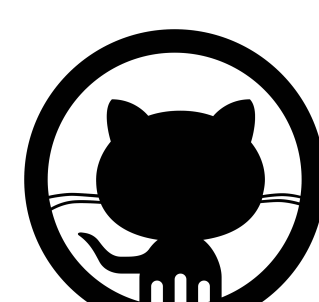


Combined score of hit accuracy

Determination of thresholds for true positive hits based on distribution of all combined scores



PromoterOntology R package, and more about the method can be found at:
<https://github.com/Dunjanik/PromoterOntology>

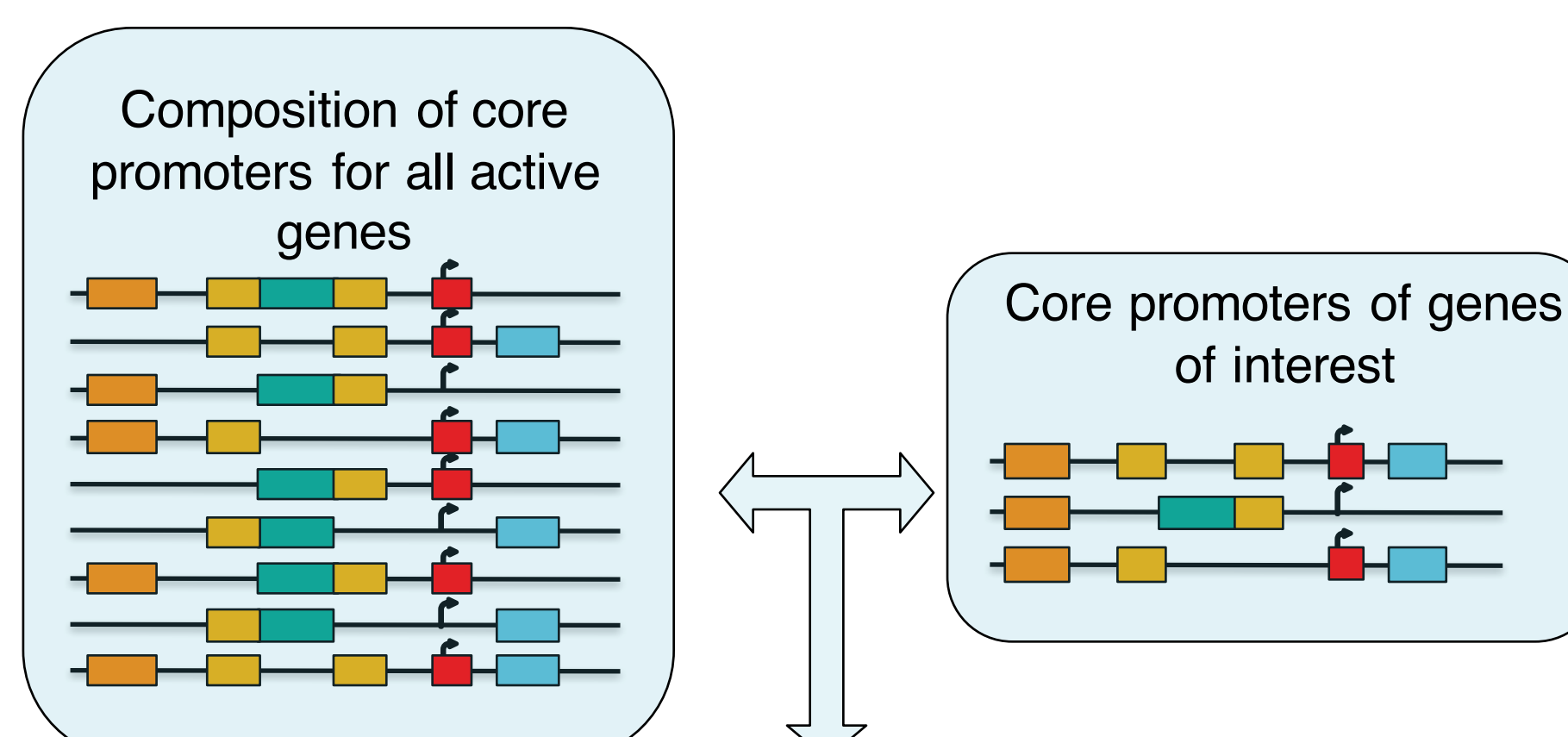


Enrichment analysis

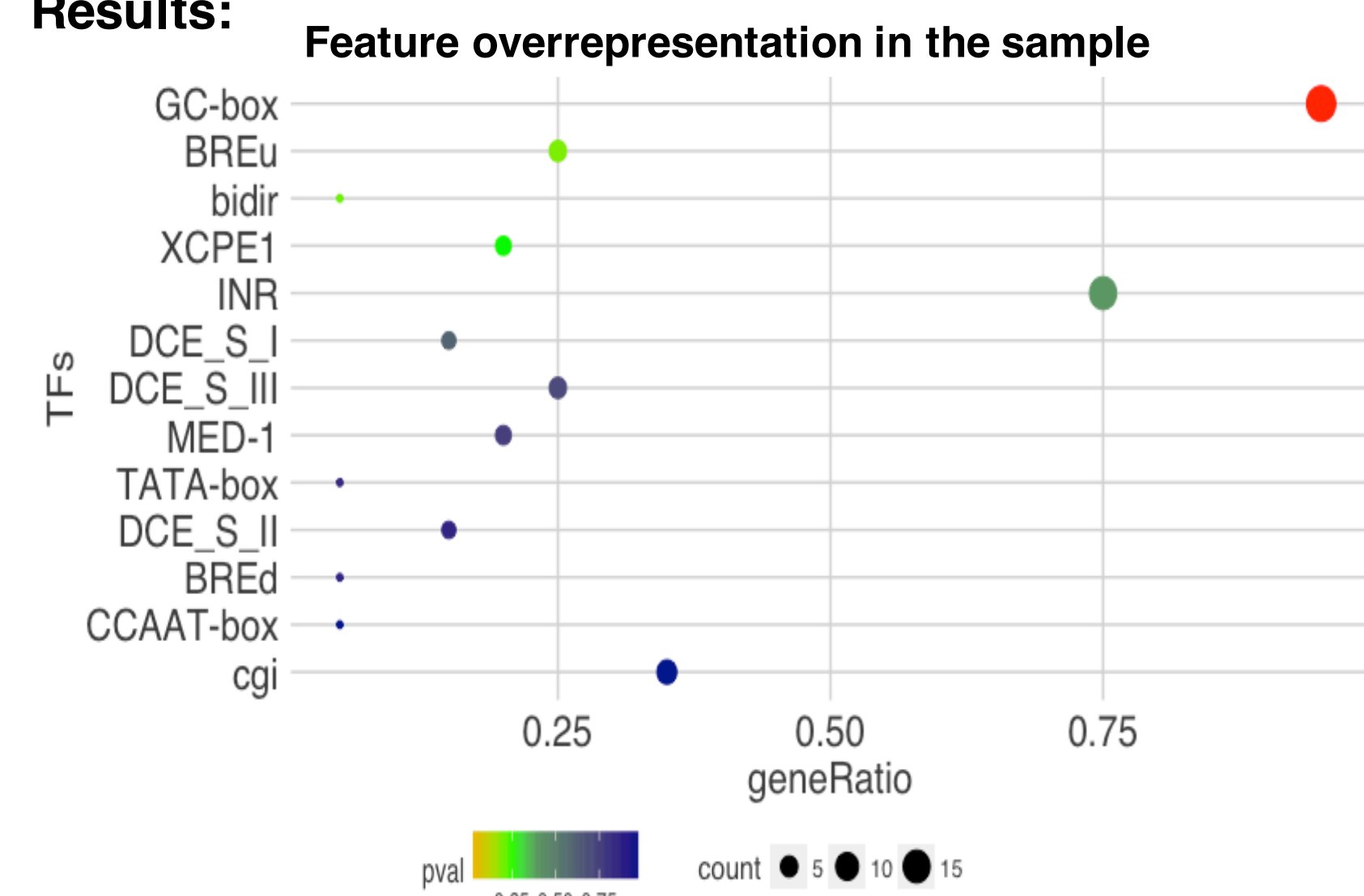
Promoter Ontology performs enrichment analysis of promoter features for a group of promoters (**sample**) compared to all active genes at that stage of development (**background**).

Features tested:

- Canonical and non-canonical core promoter motives
- CpG islands & bidirectionality
- CAGE features (IQ width, expression)



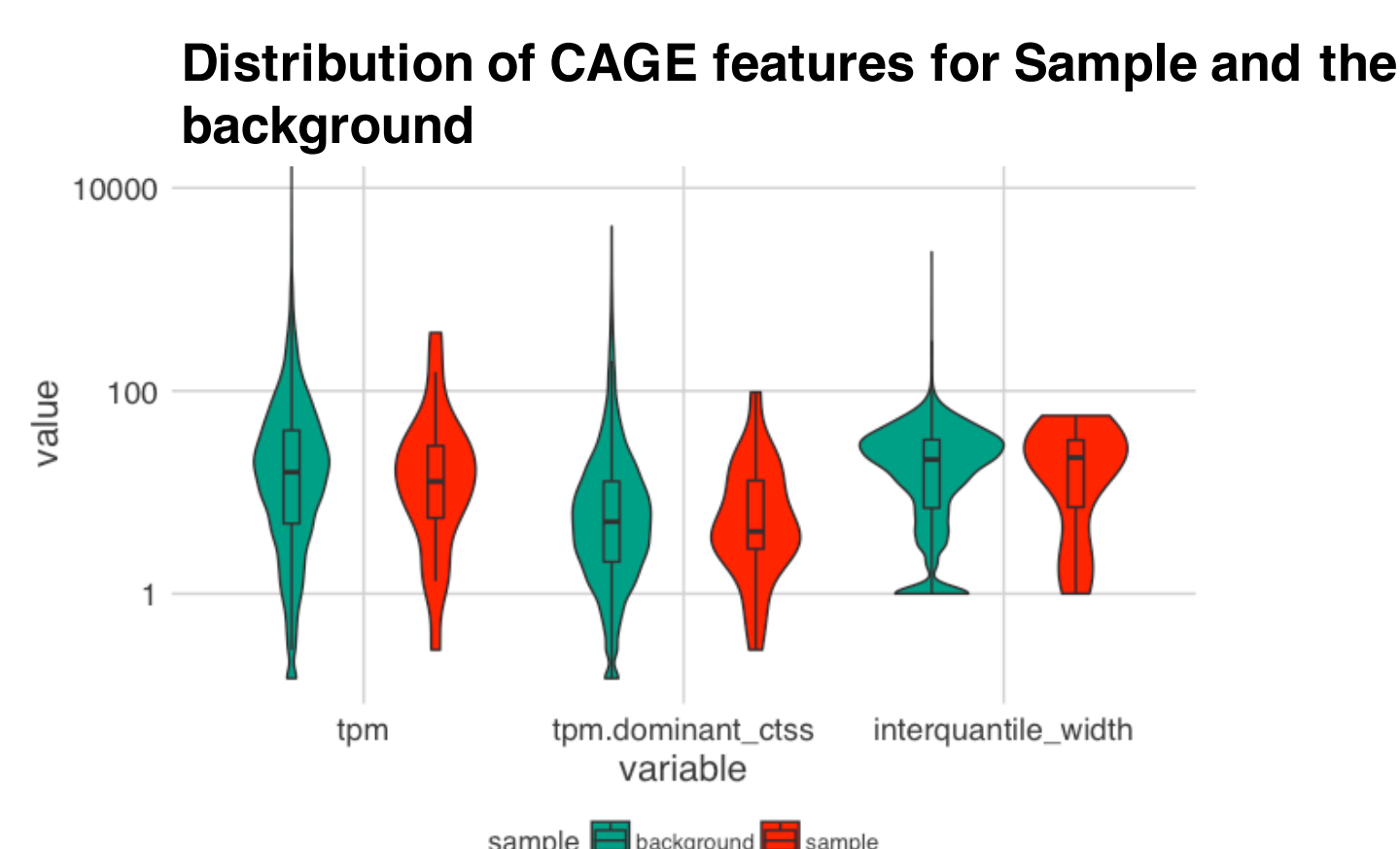
Results:



Visual representation of Promoter Ontology results shows:

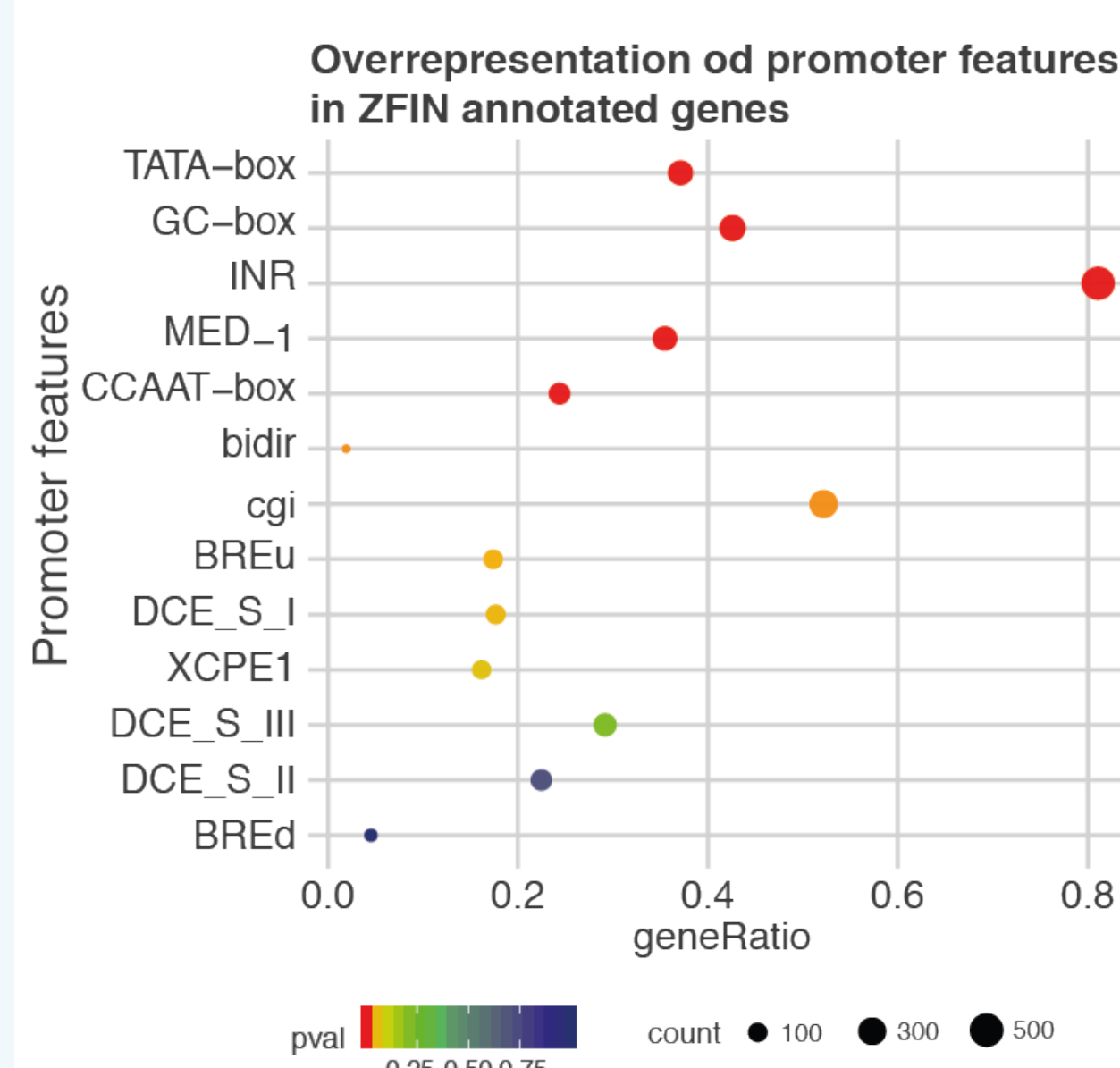
- enrichment significance (colour of the dots),
- number of genes in the sample containing feature of interest (size of the dots)

CAGE features are presented in separate plot that shows distribution of IQ width and expression in the sample and the background.



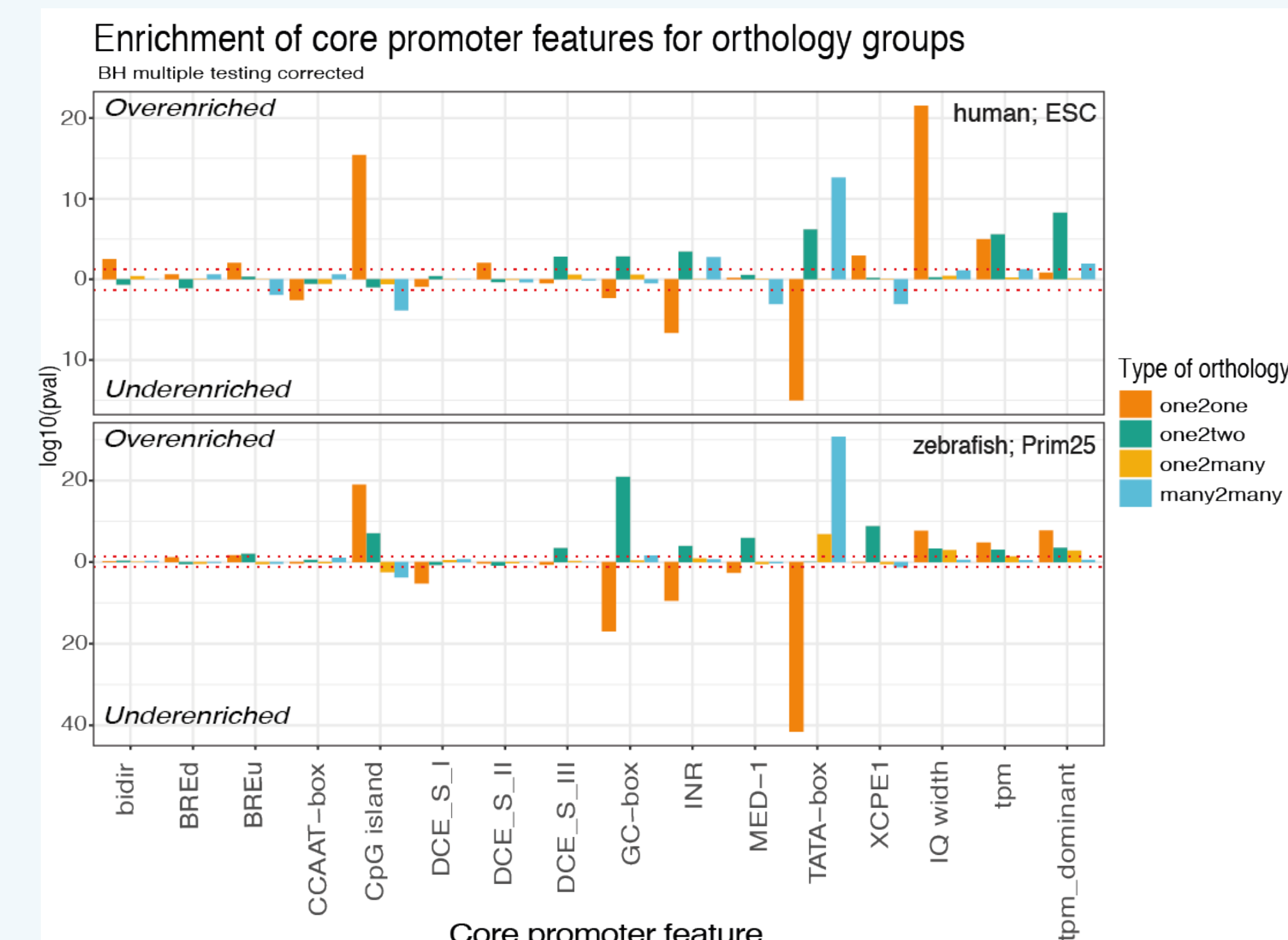
Genes annotated by ZFIN

ZFIN is a central resource of zebrafish research information. Genes annotated by ZFIN represent all genes analysed with mRNA *in situ* hybridisation studies which are mostly genes important for development and neuronal system activity.



A care should be taken when using ZFIN annotations, since this dataset lacks the whole set of TATA-less, GC-less and INR-less genes.

Human – zebrafish orthologs



One2one orthologs:

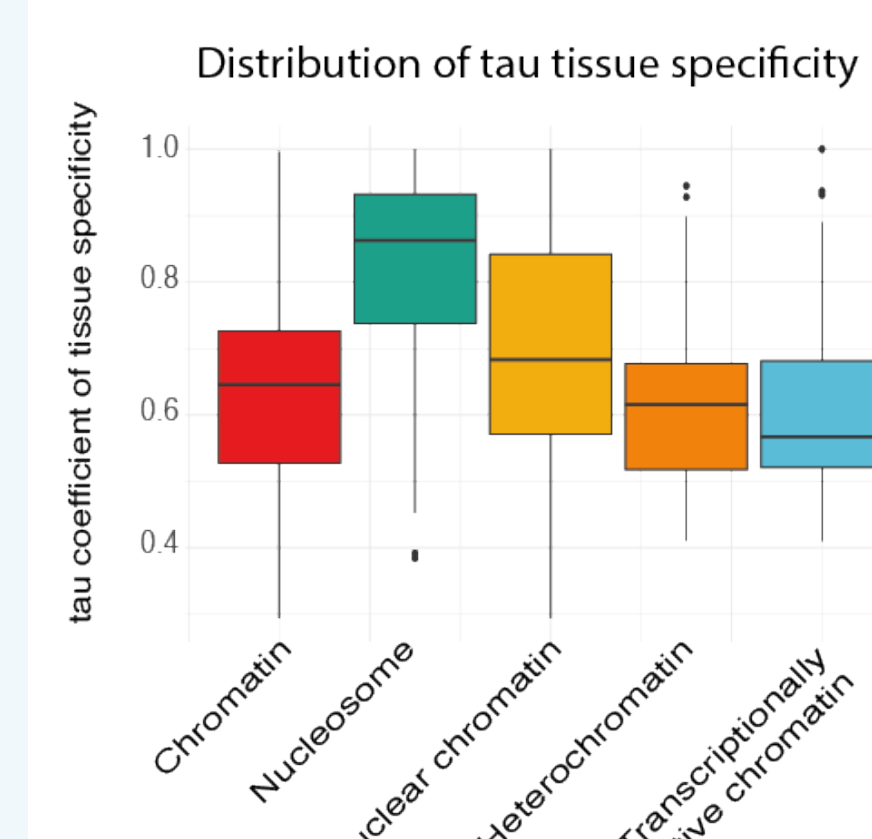
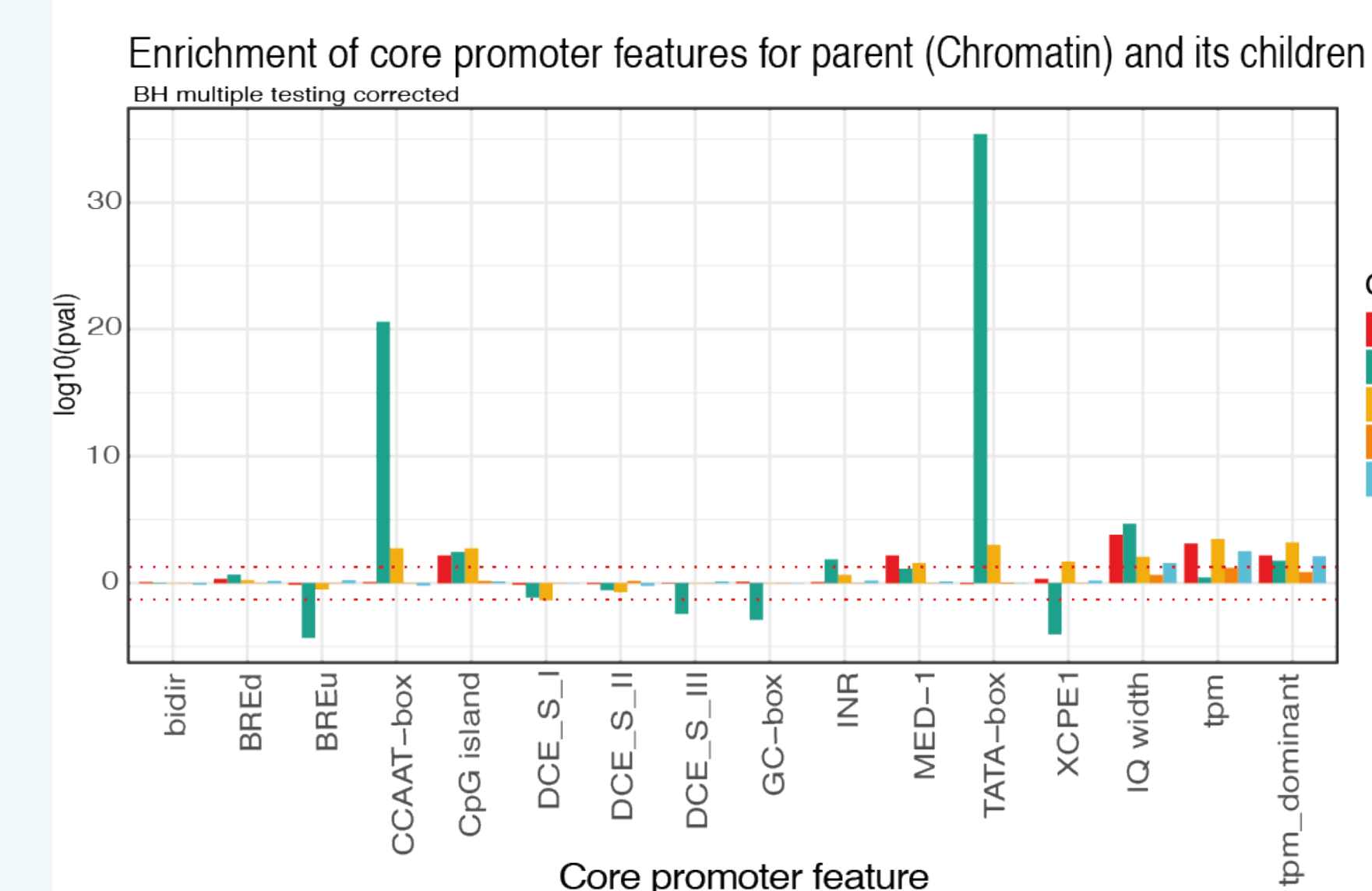
- mainly housekeeping genes
- enriched for CpG islands and strongly depleted of TATA-box.

Many2many orthologs:

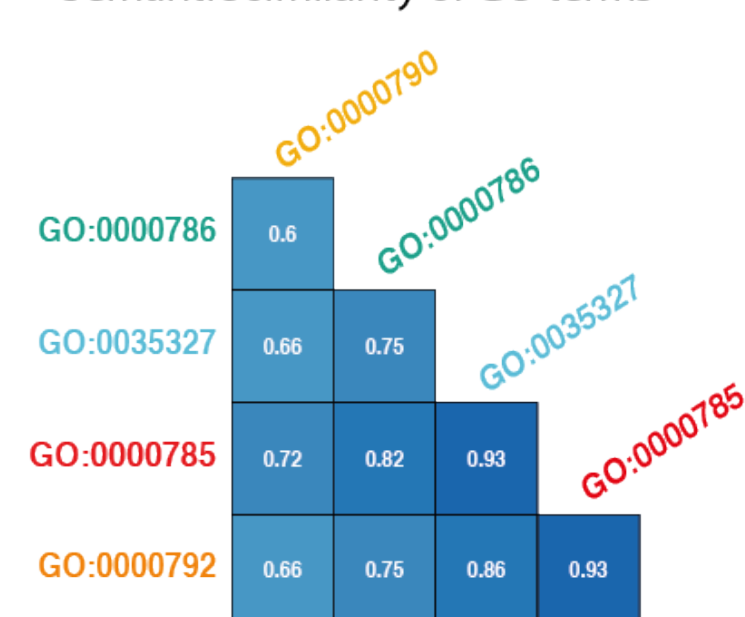
- genes involved in immune processes, and other tissue specific functions.
- enriched for TATA-box and Inr.

Core promoters of GO classes

- Related GO terms are more likely have similar core promoters
- In cases where child terms differ significantly from parent, it is usually that those groups are significantly different in their expression location or semantic similarity of GOs.



Semantic similarity of GO terms



Conclusion

- Promoter Ontology facilitates an easy way to test genes of interest for significant promoter features.
- It provides methods to annotate promoter regions with desired PWMs and other genomic features (CpG islands, bidirectionality...)
- ZFIN group of genes is strongly enriched for TATA-box, GC-box and INR containing genes
- one2one orthologs possess housekeeping-like while many2many tissue specific promoter structure

References:

- Lenhard, B., Sandelin, A., Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics* 13, p 233–245
- Ruzicka, L., Bradford, Y.M., Frazer, K., Howe, D.G., Paddock, H., Ramachandran, S., Singer, A., Toro, S., Van Slyke, C.E., Eagle, A.E., et al. (2015). ZFIN, The zebrafish model organism database: Updates and new directions. *Genesis* 53, 498–509.