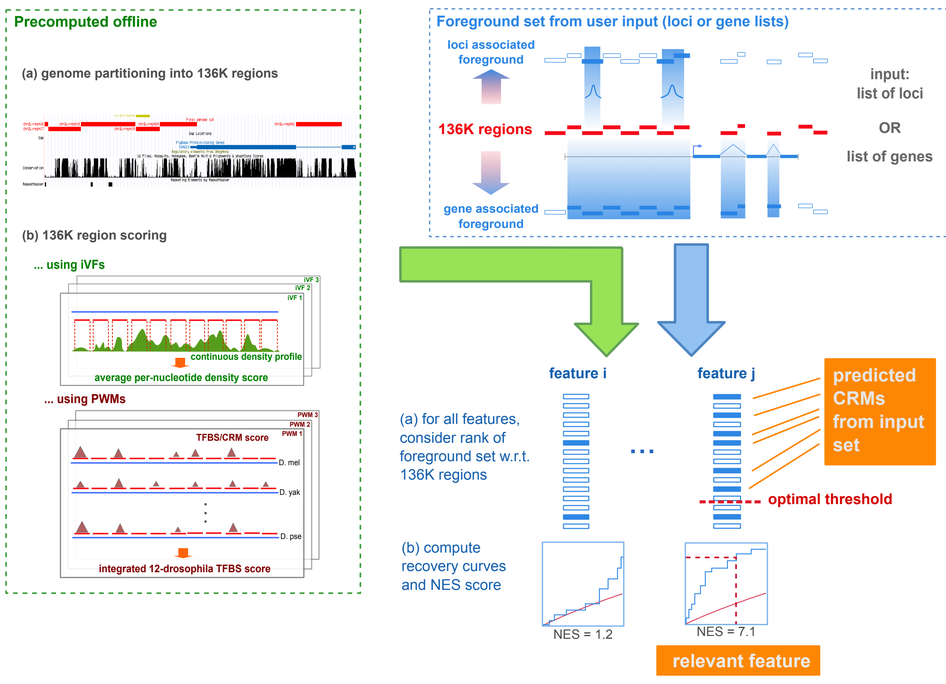
Integrative CRM prediction approaches uncover new direct target genes of retinal determination factors in fly and of cancer-related transcription factors in human.

Bram van de Sande, Marina Naval Sanchez, Delphine Potier, Rekin's Janky, Zeynep Kalender, Annelien Verfaillie, Valerie Christiaens, Lotte Haagen, Carl Herrmann[$], and Stein Aerts

Laboratory of Computational Biology, Department for Human Genetics, University of Leuven, Belgium & [$] Université de la Méditerranée & TAGC – Inserm U928, France

i-cisTarget: motifs and in vivo Tracks



Motifs

6383 Position Weight Matrices (PWM)

JASPAR, TRANSFAC, FlyFactorSurvey, Tiffin Database, O. Elemento et al. conserved motifs, Stark et al. conserved motifs, SelexConsensus, YeTFaSCo, hDPI



In vivo tracks

109 TF ChIP tracks

modEncode (42), BDTNP (41), E. Furlong lab (26)

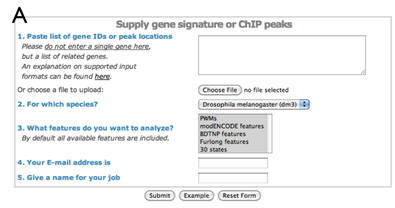
216 non-TF ChIP & DHS tracks

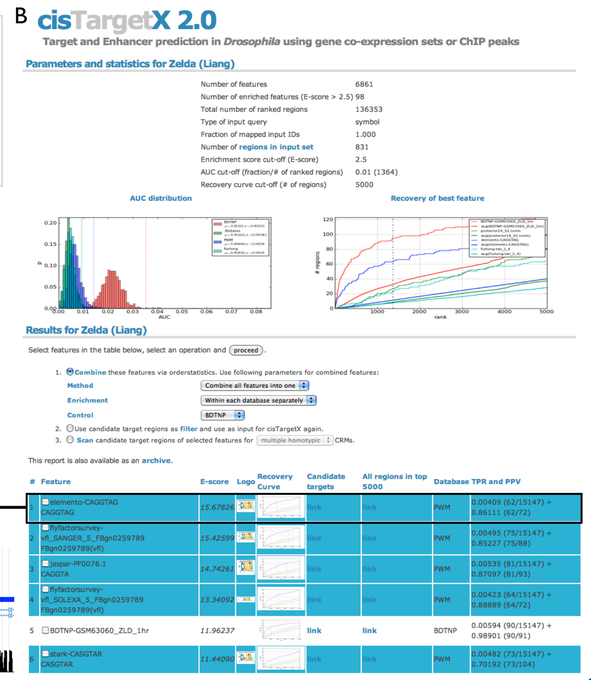
modEncode (208), BDTNP (7), E. Furlong lab (1)

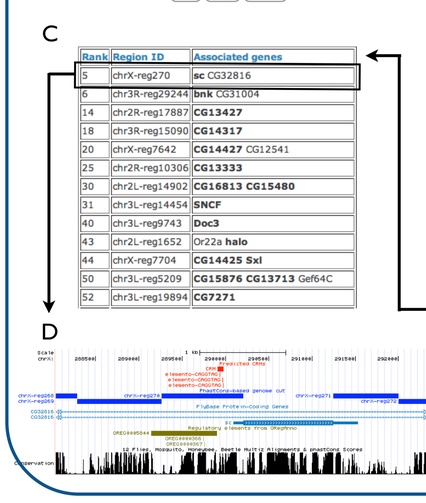
211 histone modification tracks

modEncode (205), BDTNP (0), E. Furlong lab (6)

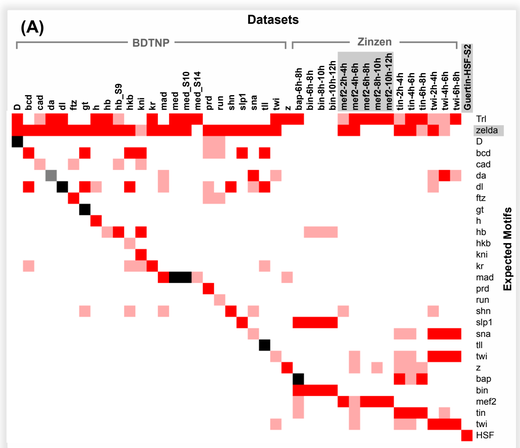
Web application: example Zelda LOF data

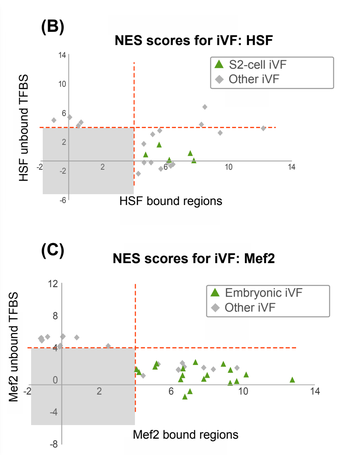






Validation on ChIP peak sets and co-expressed gene sets

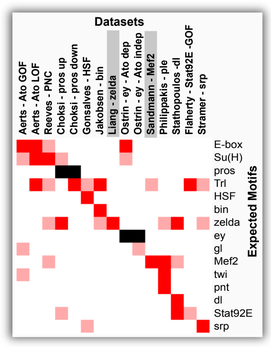


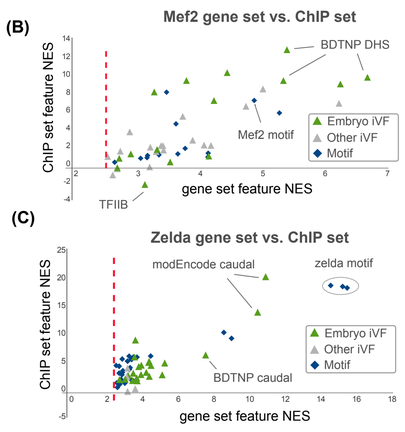


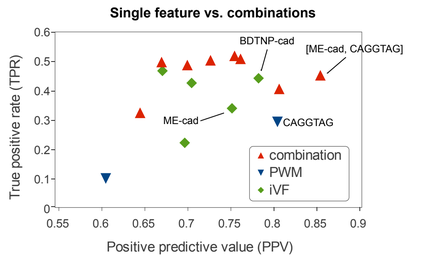
In the **heatmaps**, a red square indicates that the expected motif (on the right) was found among the top 3 motifs. A pink square indicates that the expected motif was found to be significantly enriched. A black square indicates that the expected motif was not found enriched.

The **scatterplots on the left** show that in vivo tracks (IVT) are found enriched in truly bound regions, while not in unbound regions. The bound regions are derived from ChIP peaks, and the unbound regions are genomic regions with high-scoring PWM matches for the TF under study, but without a ChIP peak.

The **scatterplots on the right** show the correlation between the enrichment score for in vivo tracks and motifs, obtained by analyzing a gene set (e.g., Zelda LOF) and a related ChIP peak set (e.g., Zelda ChIP). Agreement between these indicates that the features we discover on the gene set are indicative of true regulatory regions.

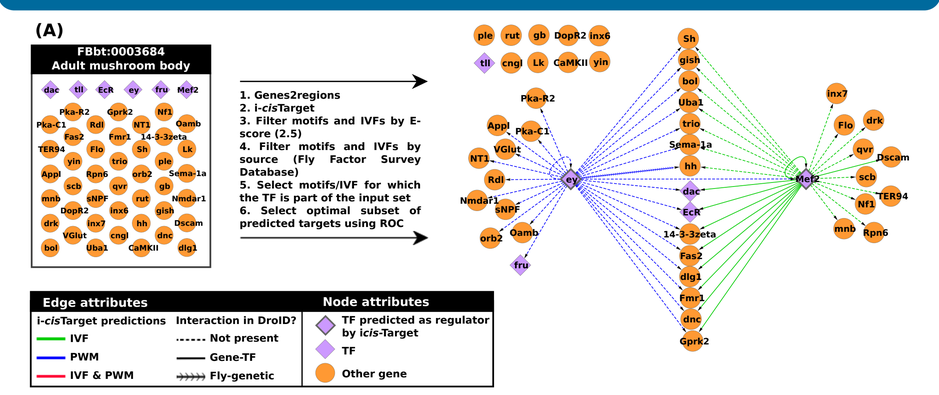




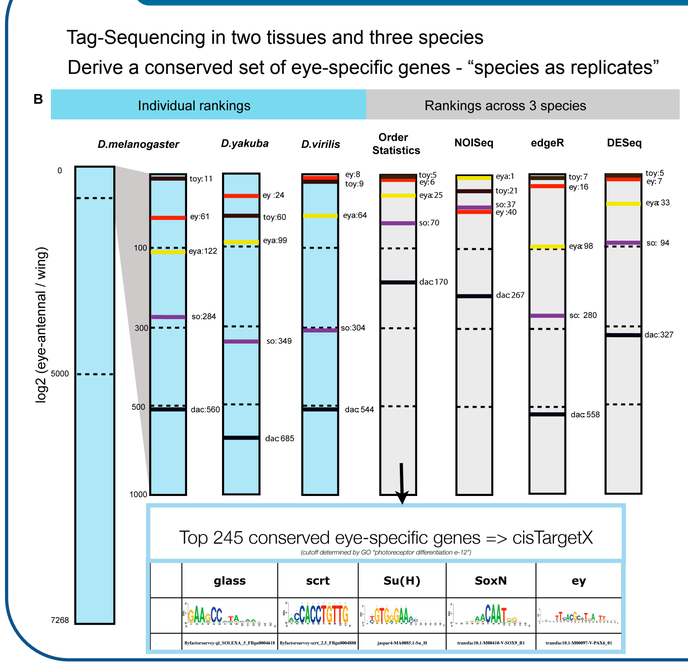


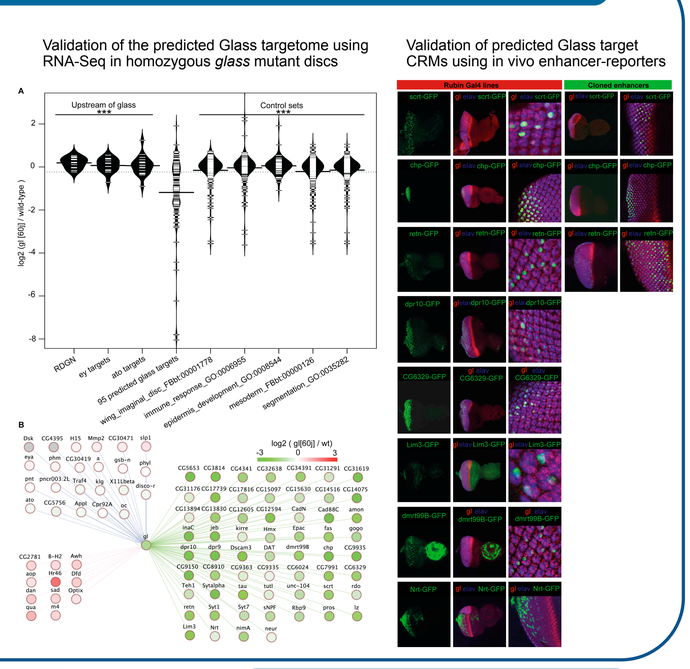
Combinations of motifs and In Vivo Tracks improve the accuracy of CRM predictions (TPR and PPV).
In this example we analyze a set of downregulated genes in the Zelda LOF (Liang et al.), and compare the predicted CRMs with Zelda ChIP peaks. In this example, the ChIP peaks are used as true positives, and not used as iVT.

Application to 628 co-expressed gene sets

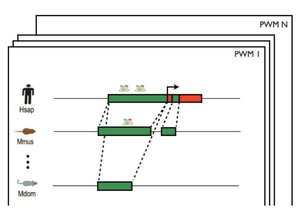


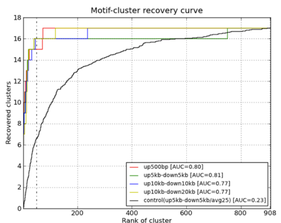
Comparative Transcriptomics + Comparative motif discovery



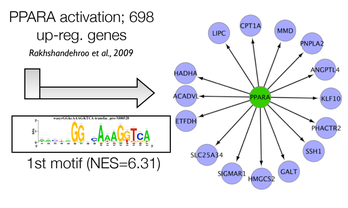


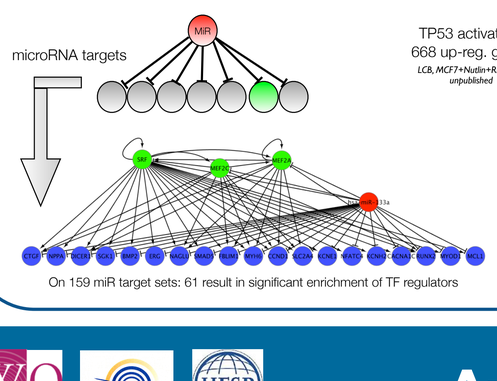
cisTargetNET for Human and Mouse

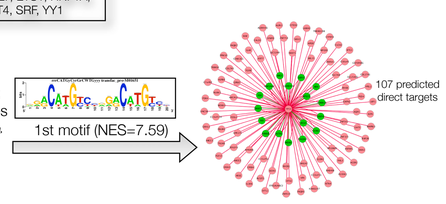


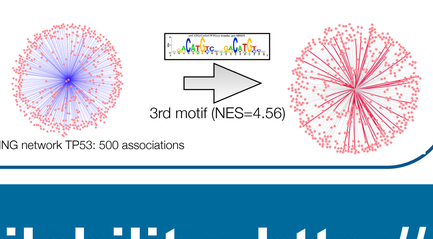


Validated (13/19) (incl. AMADEUS Benchmark)
PPAR, TP53, NFkB, LEF1, E2F, ETS1, HNF1A, HSF1, NFY, NRF1, OCT4, SRF, YY1







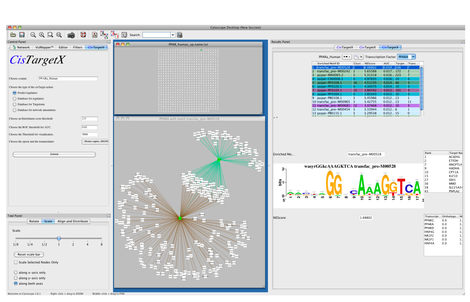


Meta-Targetomes obtained from 2211 cancer gene signatures



Meta-Targetomes: cisTargetNET was applied to 2211 cancer gene signatures. For each significant motif and linked TF, the number of target gene predictions across gene signatures are counted. The meta-targetome is the ranking of predicted targets based on the recurrence of finding the same target across gene signatures.

Cytoscape Plugin



Availability: http://med.kuleuven.be/lcb/i-cisTarget