Supplementary Material

Comparative motif discovery combined with comparative transcriptomics yield accurate targetome and enhancer predictions

Marina Naval Sanchez[1], Delphine Potier[1], Lotte Haagen[1], Máximo Sánchez[2], Sebastian Munck[3], Bram Van de Sande[1], Fernando Casares[2], Valerie Christiaens[1], and Stein Aerts[1 #]

[1] Laboratory of Computational Biology, Department of Human Genetics, University of Leuven, Leuven, Belgium

[2] Centro Andaluz de Biología del Desarrollo (CABD) CSIC-UPO-Junta de Anadalucía, Sevilla, Spain

[3] LiMoNe, VIB Center for the Biology of Disease, Leuven, Belgium

[#] Correspondence to Stein Aerts, Herestraat 49, PO Box 602, B-3000 Leuven, Belgium. Email: stein.aerts@med.kuleuven.be

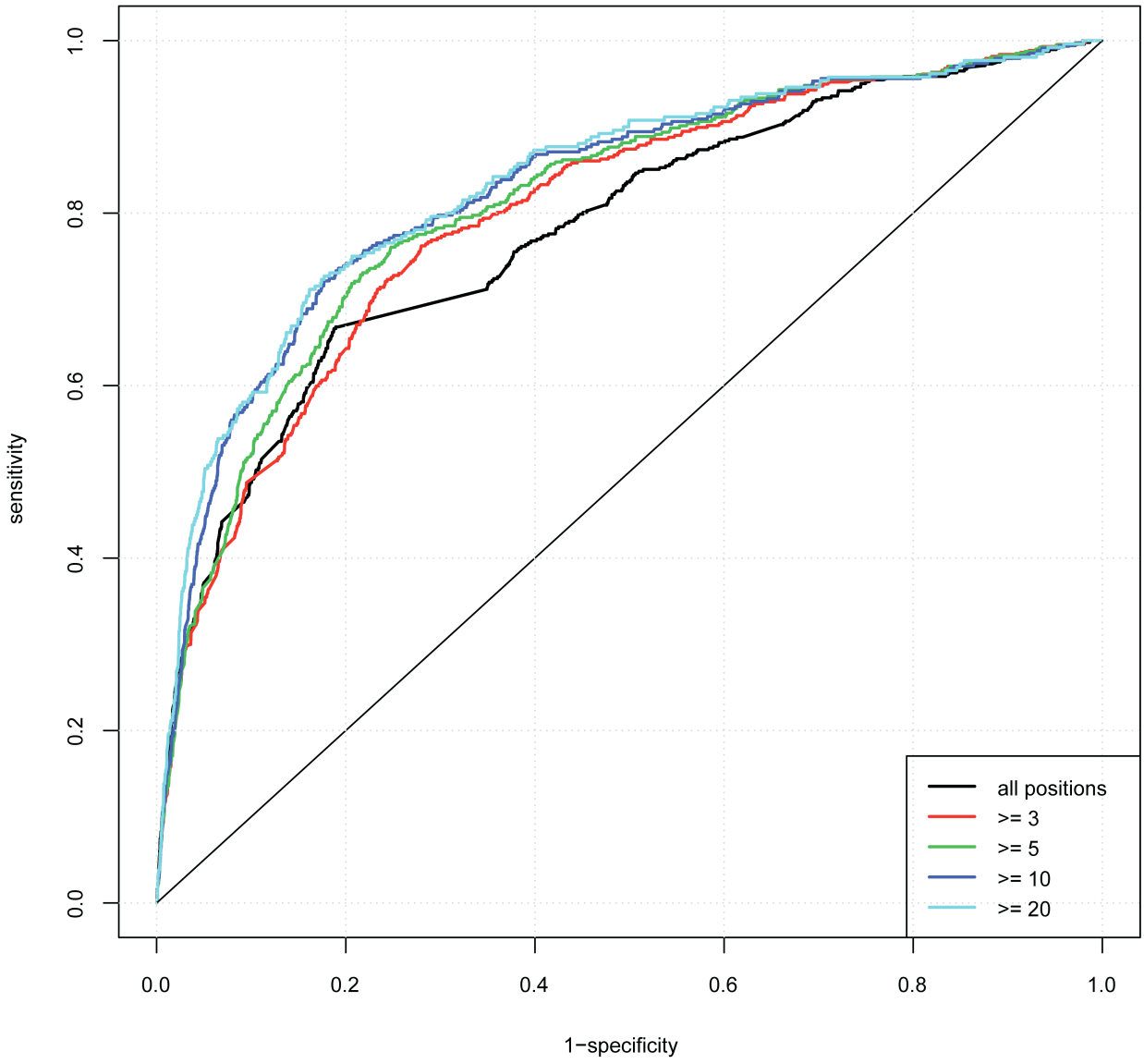


Figure S1: Tag-seq filtering

Comparison of different cutoff values for the minimal number of reads in at least one sample (eye-antennal or wing). X-axis represents the ranking of all genes according to the log2 (eye/wing), using different filtering thresholds. Y-axis represents the sensitivity to recover the "true positives", for which we used a set of 507 eye-enriched genes obtained by microarray data (Ostrin et al. 2006).

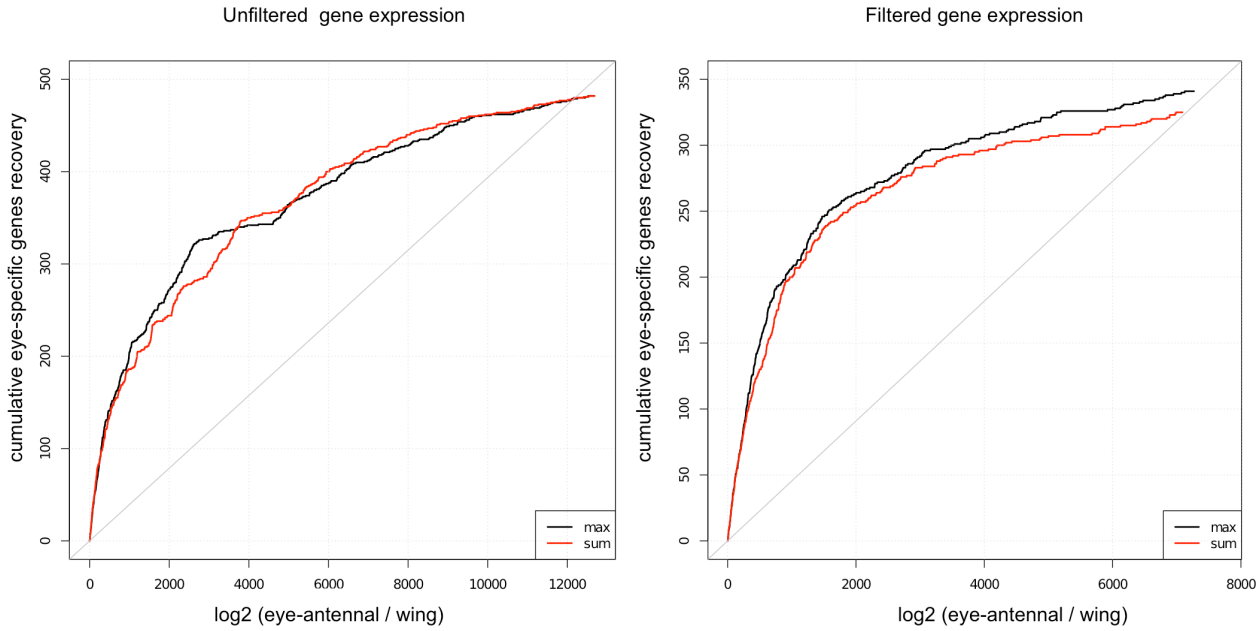


Figure S2: Max versus Sum

Comparison of using the max (peaks) for a gene to derive the expression value, or the sum of all peaks.

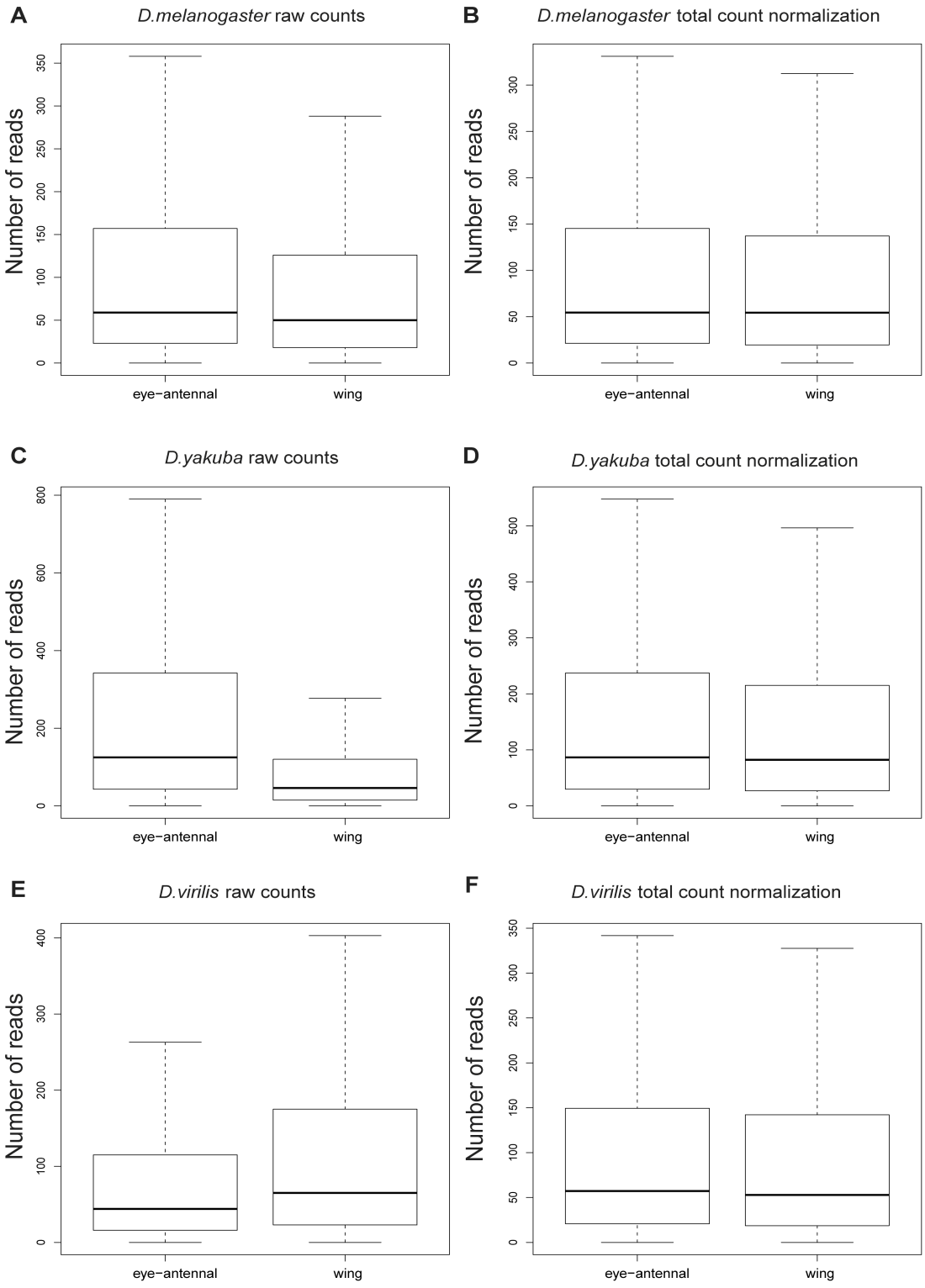


Figure S3: Distribution of gene expression values before and after total-count normalization. Boxplots from raw counts (A,C,E) and total-count normalized distributions (B,D,F) per species (A-B for *D. melanogaster;* C-D for *D. yakuba*; E-F for *D. virilis*) and tissue samples.

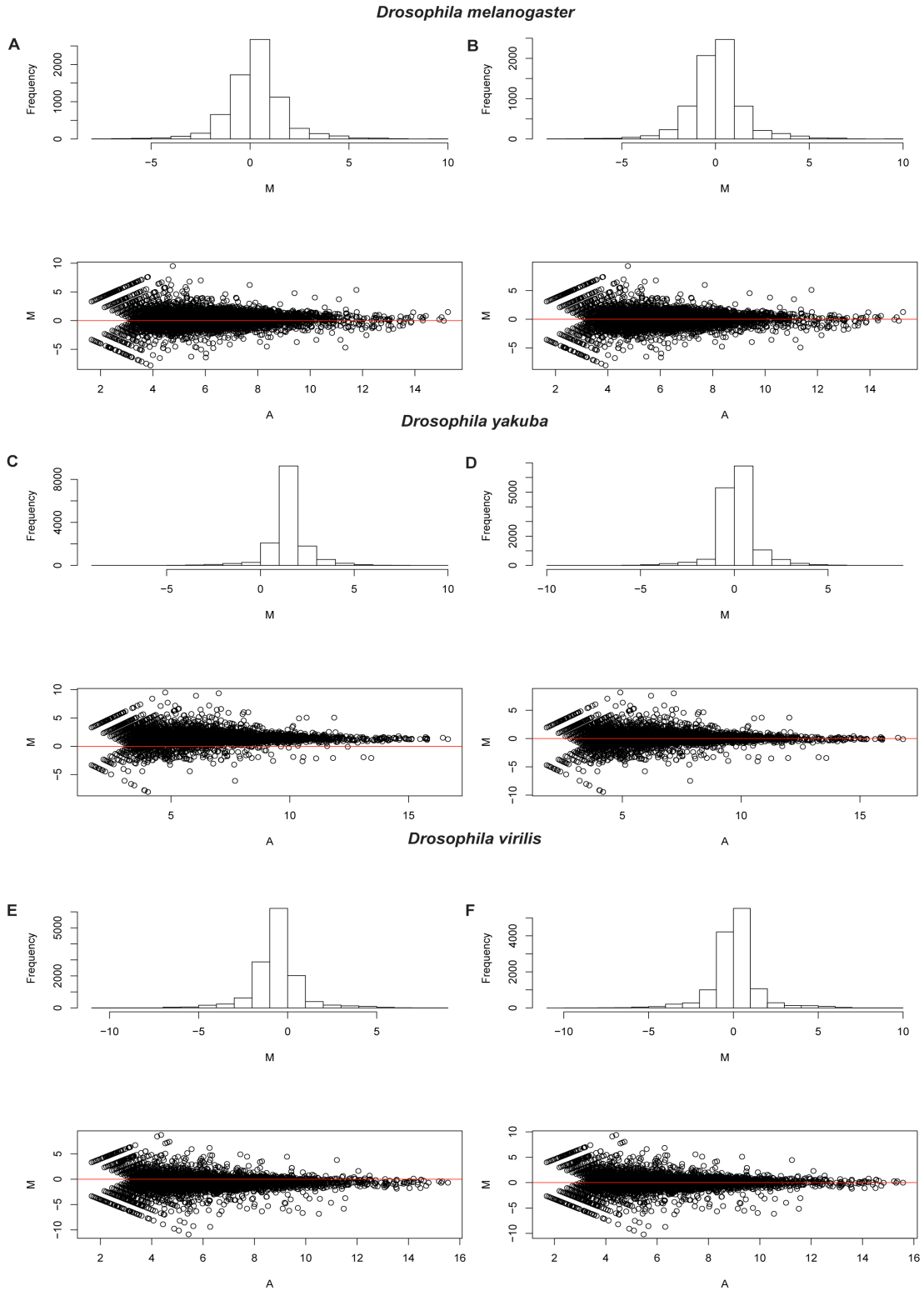


Figure S4: Distribution of log2 (eye-antennal / wing) ratios and MA plots before and after normalization.

Histograms and MA plots before (A,C,E) and after (B,D,F) total-count normalization.

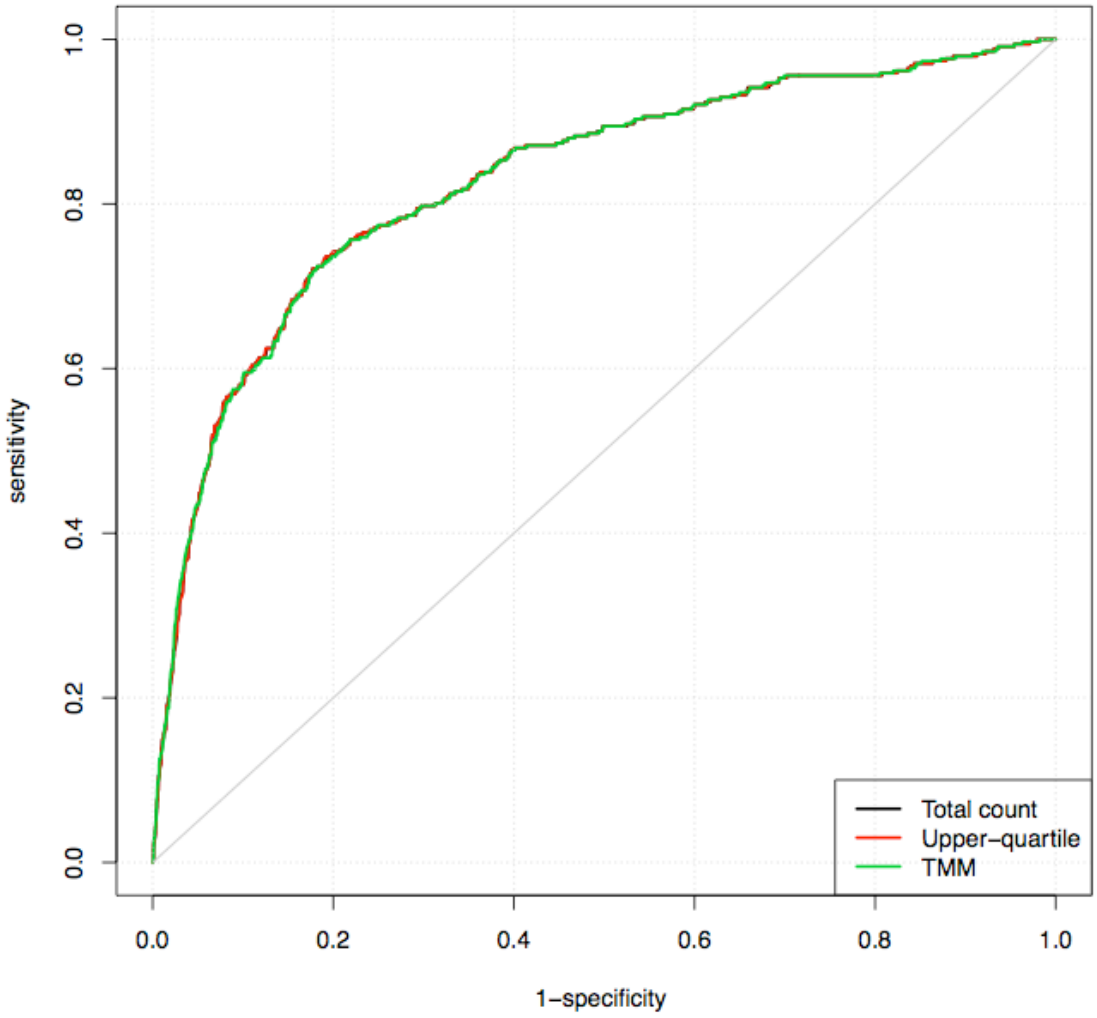


Figure S5: Alternative methods for normalization of Tag-seq data

Different methods were compared to normalize the Tag-seq read counts, namely total-count, upper-quartile (Bullard et al.), and trimmed mean of M-values (TMM) (Robinson and Oshlack 2010) The different methods yield similar results, which we believe is due to the fact that we also normalize within each species, by using the log2(eye/wing). The recovery curves are constructed in a similar way as in Figure S1, using the same set of true positives.

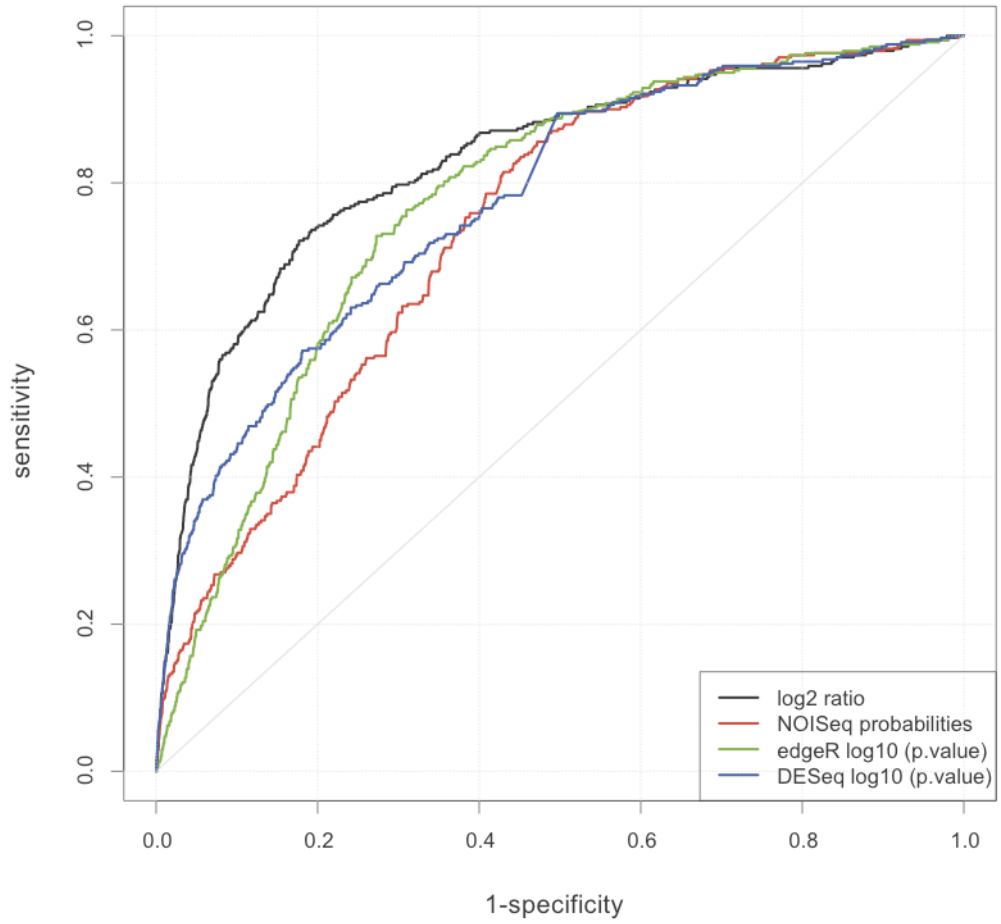


Figure S6: Differential expression by Tag-seq in *D.melanogaster* only

The x-axis represents all *D.melanogaster* genes ranked according to either of four measures: logratio in black, NOISeq probabilities in red, signed -log10(p value) of edgeR in green, and signed -log10(p value) of DESeq in blue. As true positives we used a set of 507 eye-enriched genes obtained from microarray data (Ostrin *et al.* 2006).

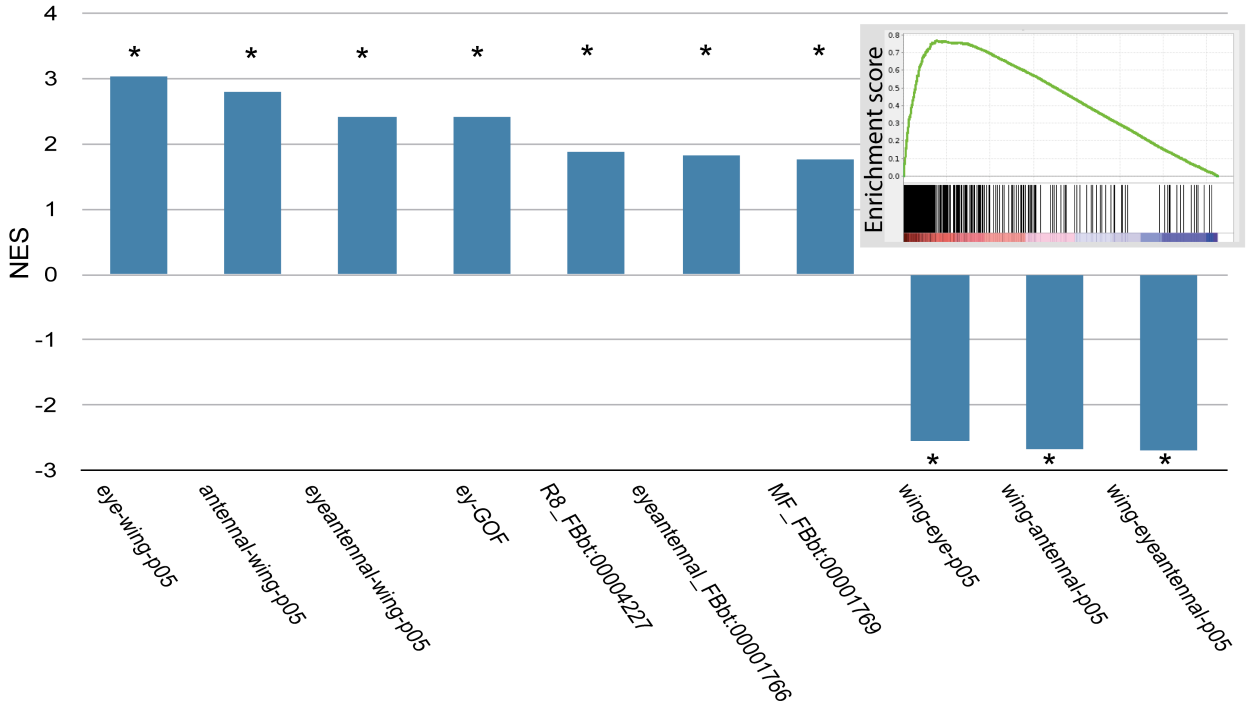


Figure S7: GSEA results on eye-vs-wing comparison

Normalized enrichment scores (NES) obtained by GSEA on the *D.melanogaster* log2(eye-antennal/wing) based ranking, using a selection of eye and wing related gene sets.

Eye-enriched gene sets: “eye-wing-p05”: set of 507 genes used in Figure S1; “antennal-wing-p05” and “eyeantennal-wing-p05” are derived from microarray data from (Aerts et al. 2010), containing 173 and 1383 genes respectively; “ey-GOF”: overexpressed genes when performing ey-GOF in imaginal discs (Ostrin et al. 2006), 178 genes ; FlyBase TermLink sets: FBbt:00004227 (34 genes), FBbt:00001766 (209 genes), FBbt:00001769 (72 genes). Wing-enriched gene sets: “wing-eye-p05”: 178 wing-specific genes from microarray data, compared to eye expression (Ostrin et al. 2006); “wing-antenna-p05” and “wing-eyeantennal-p05” : 125 and 182 wing specific genes (Aerts et al. 2010).

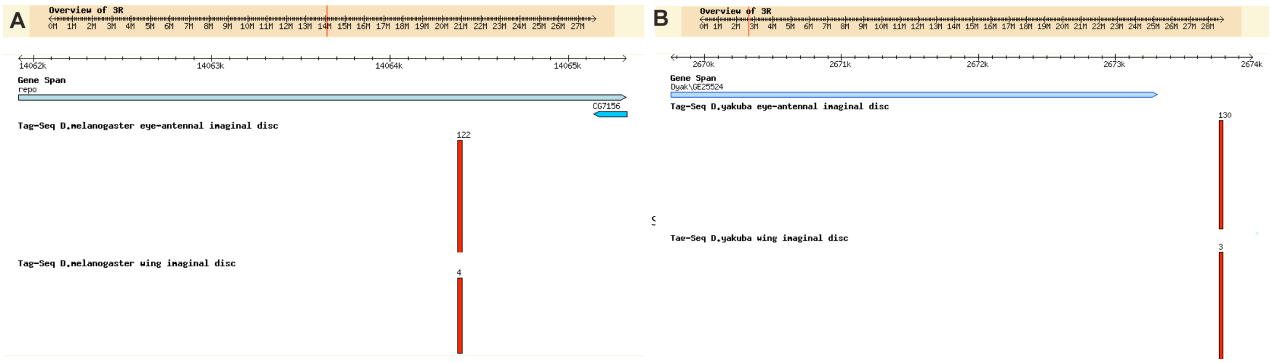


Figure S8: An example where non-model organisms require annotation amendments.

FlyBase Genome Browser Visualization of tags assigned to the *repo* gene in *D.melanogaster* (A) and *D. yakuba* (B), in eye-antennal and wing imaginal discs.

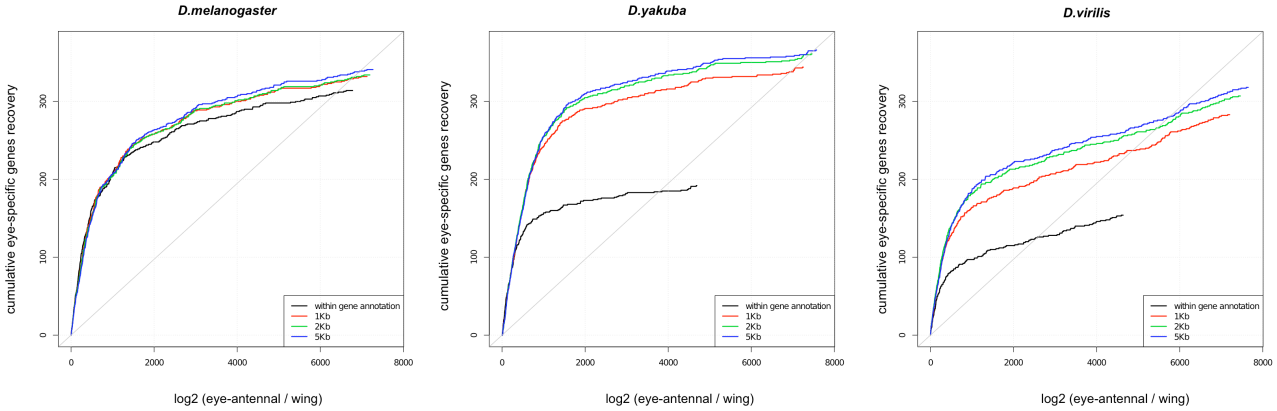


Figure S9: Assignment of free Tag-seq peaks to genes.

Comparison of 1kb, 2kb, and 5kb extensions of species-specific annotation to assign yet unassigned peaks to genes.

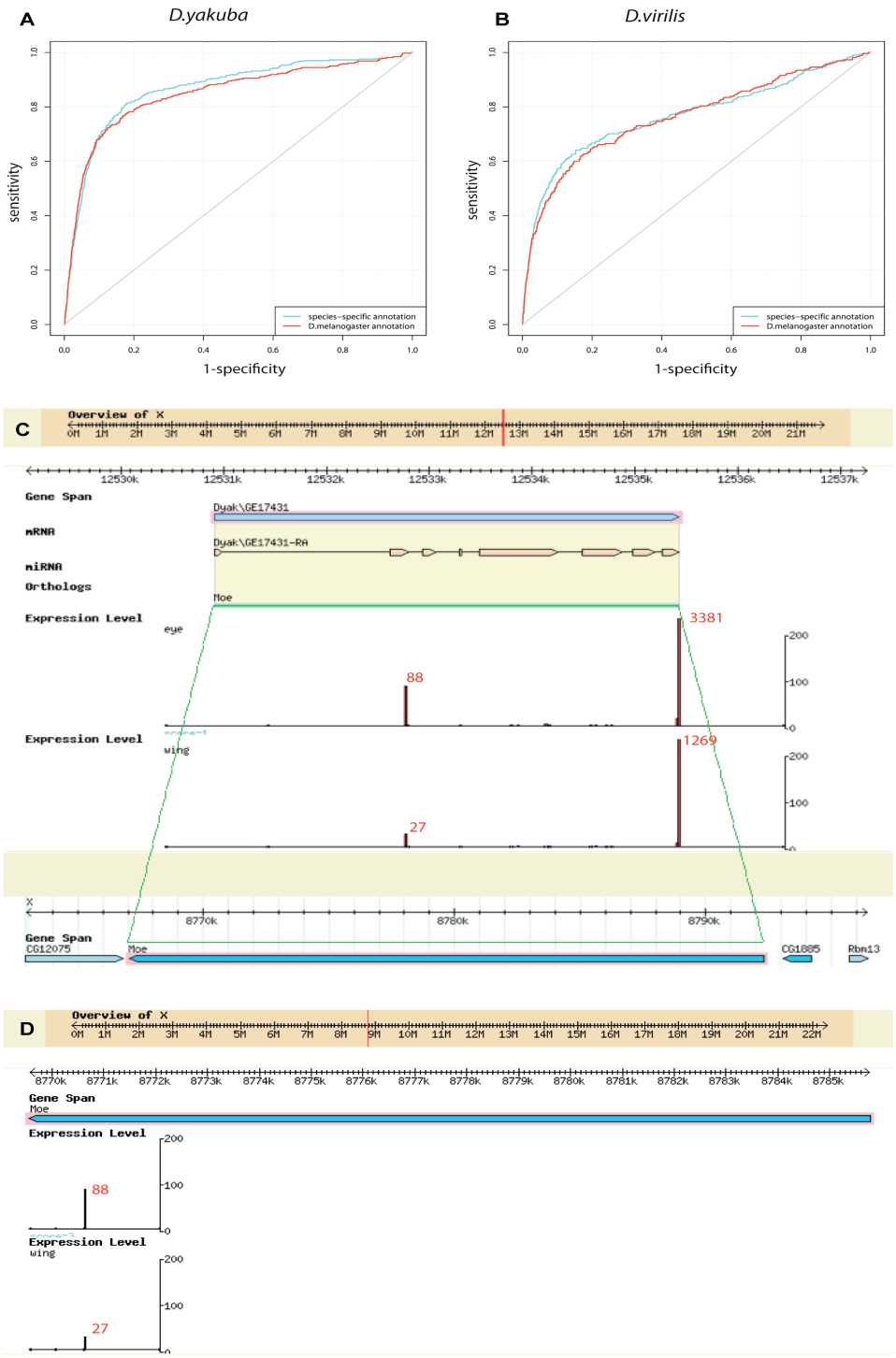


Figure S10: Differences between 3' UTR extension and conversion to *D.melanogaster* coordinates. (A-B) Comparison of annotation methods for non-model organisms: species-specific annotation with 3'UTR extension (green) and conversion to *D.melanogaster* coordinates (red) for *D.yakuba* (A), and *D.virilis* (B), by a similar recovery curve as in Figure S1. (C) Genome browser view of a gene involved in eye development, *Moe*, using the 3'UTR extension (C) and the conversion to *D.melanogaster* coordinates (D), yielding different peaks and different expression values in *D.yakuba*. In such cases we choose the logratio(eye/wing) that is most similar to the *D.melanogaster* ratio.

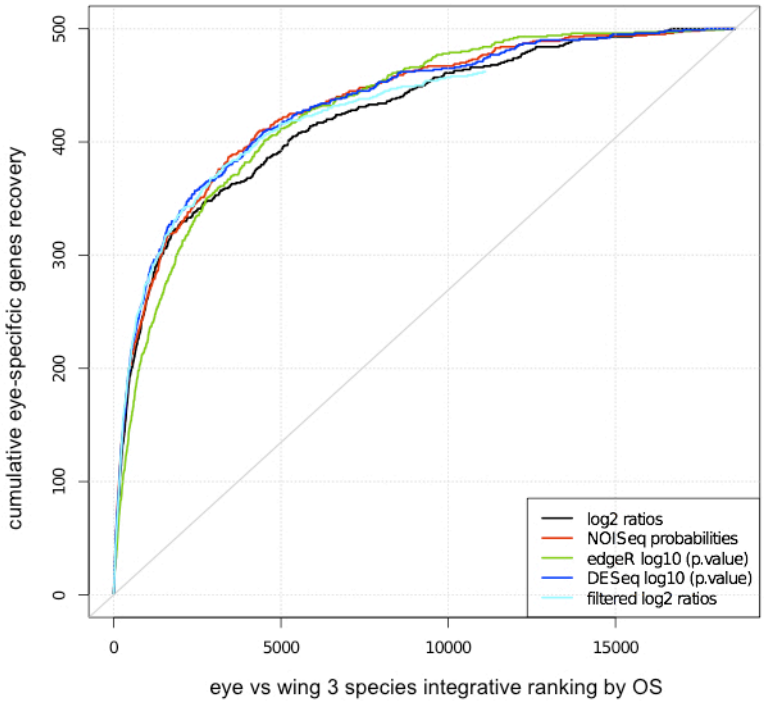


Figure S11: Alternative methods for rank aggregation across species.

Eye-enrichment gene rankings can be performed by multiple methods. The cumulative recovery curve is shown for 507 eye-specific genes (Ostrin et al. 2006) when performing the individual species differential expression analysis of eye-antennal versus wing by log2 ratio (eye-antennal / wing) (black), NOISeq (Tarazona et al. 2011) (red) , edgeR (Robinson, McCarthy, and Smyth 2010) (green), DESeq (Anders and Huber 2010) (dark blue), and log2 (eye-antennal / wing) ratio after filtering low expressed genes (see Figure S1) (light blue). The situation described in the text corresponds to the light blue curve.

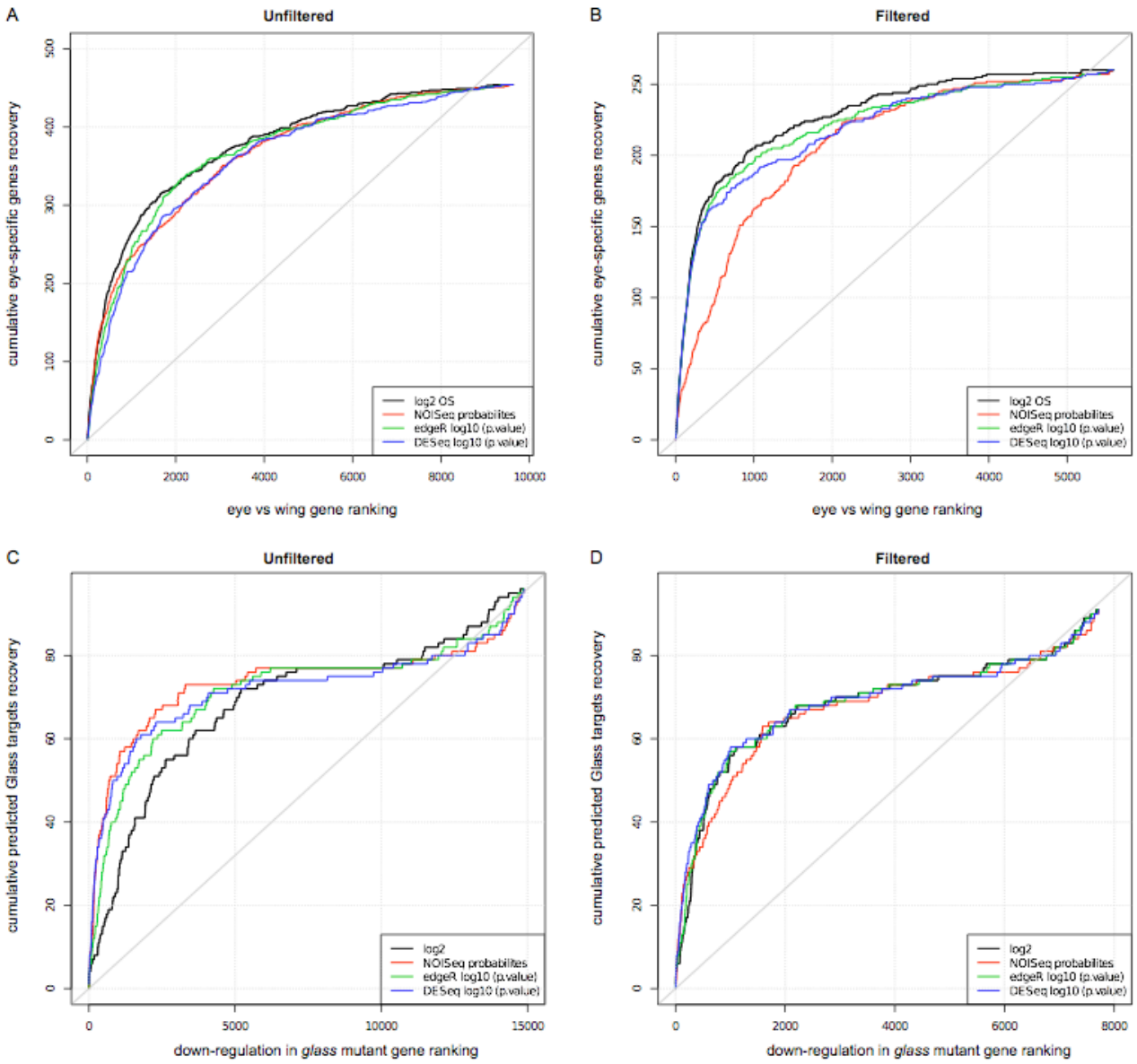


Figure S12: Differential expression by Tag-seq across species and by 'classical' RNA-seq in *D. melanogaster*

Differential expression methods applied are rank integration by Order Statistics (OS) (Aerts et al. 2006) (black), NOISeq (Tarazona et al. 2011)(red), edgeR (Robinson, McCarthy, and Smyth 2010) (green) and DESeq (Anders and Huber 2010) (blue). These methods are applied to cross-species Tag-seq data (A-B) and to classical differential expression in a single species, between two conditions, using replicates (C-D). True positives used for A-B are 507 eye-specific genes (Ostrin *et al.* 2006); while true positives used for C-D are 96 predicted Glasss targets. A) Unfiltered data, 9633 genes presenting expression in all three species considered in this study. B) Filtered data on number of counts (see Figure S1) and with expression in all three species, 5691 genes. C) Unfiltered RNA-seq data. D) Filtered data RNA-seq data using a threshold of 5 RPKM in at least

two of the four samples. This figure shows that the order statistics (on the logratio ranking) outperforms classical methods on cross-species data. On RNA-seq data sets, the logratio is valid when data are filtered, otherwise the statistical methods are more robust.

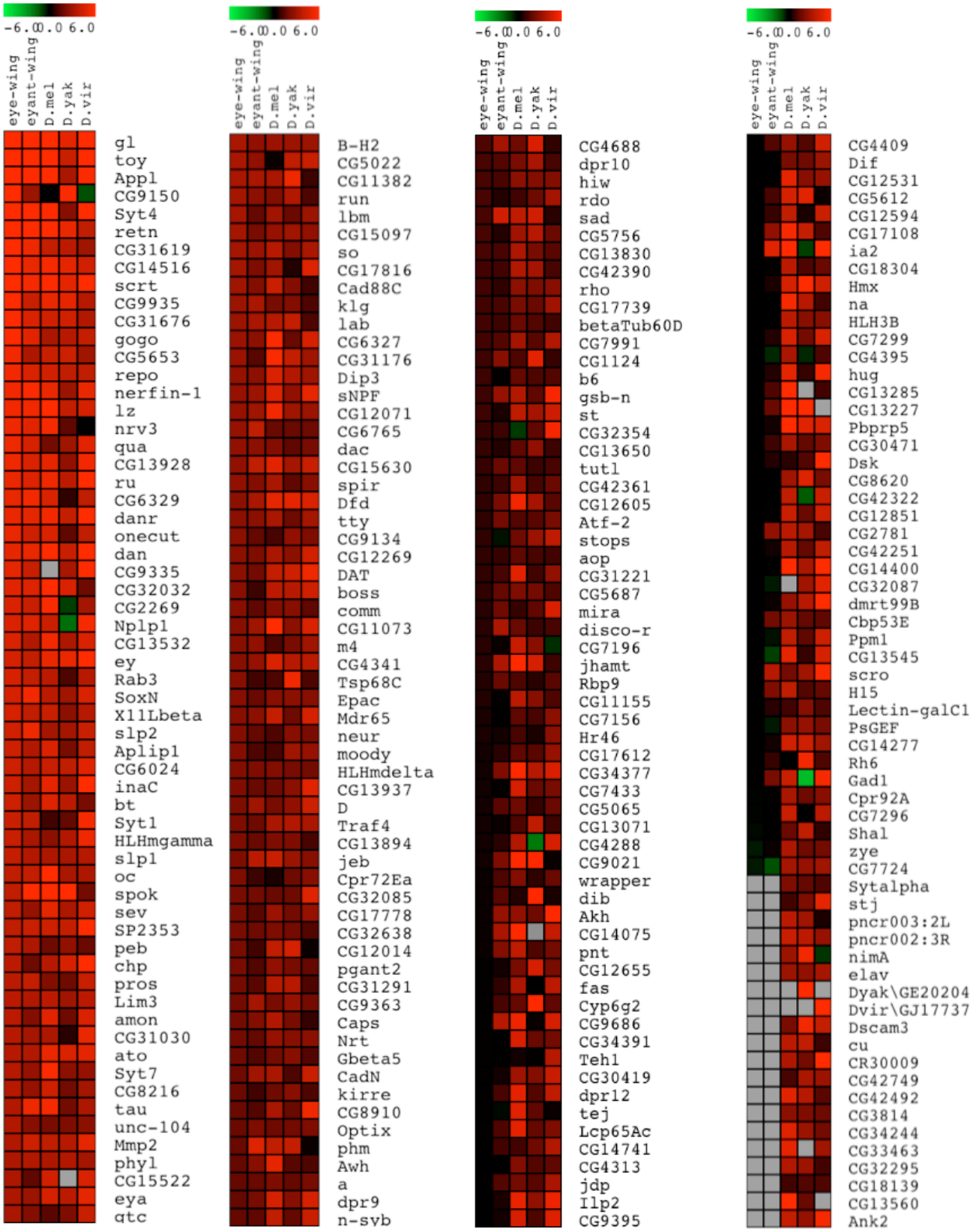


Figure S13: Log2 (eye-antennal / wing) ratio for the top 245 genes in the cross-species order statistics ranking. First two columns correspond to microarray data from (Ostrin et al. 2006), third to sixth column represent *D.melanogater, D.yakuba, D.virilis* log2 (eye-antennal / wing) measures by Tag-seq from this analysis.

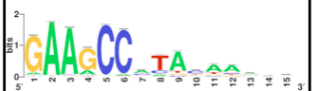
	Glass			SCRT		SU(H)		SOXN		EY	
											
	flyfactorsurvey_gl_SOLEXA_5_FBgn0004618			flyfactorsurvey-scrt_2.5_FBgn0004880		jaspar4-MA0085.1-Su_H		transfac10.1-M00410-V-SOX9_B1		transfac10.1-M00097-V-PAX6_01	
Top genes individual rankings	**Motif rank**	**NES**	**Number of Target Genes**	**Motif rank**	**NES**	**Motif rank**	**NES**	**Motif rank**	**NES**	**Motif rank**	**NES**
***D.melanogaster* (100)**	**29**	**2.507075**	**13**	2421	-0.375550856	2236	-0.24903562	707	0.862376692	193	1.645489979
***D.yakuba* (500)**	**1**	**3.678888**	**64**	3424	-1.345472222	1476	0.324799668	161	1.669020829	146	1.715076437
***D.virilis* (200)**	**1**	**3.678859**	**36**	933	0.6514162658	3253	-1.1136774	805	0.769967333	76	2.102569143

Figure S14: Motif discovery by cisTargetX when using the top eye-enriched genes from individual species. Gene set thresholds giving best enrichment score in cisTargetX (Aerts et al. 2010) for **flyfactorsurvey-gl_SOLEXA-5_FBgn0004618** motif (Zhu et al. 2011). Motif rank, enrichment scores and number of target genes for glass motif and motif rank and enrichment scores for motifs corresponding to SCRT, SU(H), SOXN and EY for which high enrichment is found when considering three species as replicates.

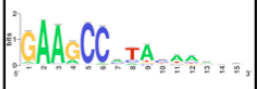
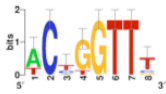
TF	LOGO		DESeq		edgeR		NOISeq		OS	
			250 genes		250 genes		250 genes		245 genes	
			Motif rank	NES	Motif rank	NES	Motif rank	NES	Motif rank	NES
GL		flyfactorsurvey-gl_SOLEXA_5_FBgn0004618	1	**5.3455**	2	4.4723	1	**4.8317**	1	**4.7523**
SCRT		flyfactorsurvey-scrt_2.5_FBgn0004880	16	3.0195	23	2.8229	-	-	3	***3.506898***
SU(H)		jaspar4-MA0085.1-Su_H	**6**	**3.7133**	1	4.681	8	3.47088	15	*3.104033*
GRH		transfac10.1-M00951-I-GRH_01	41	2.666	-	-	2	4.09867	-	-

Figure S15 cisTargetX results on various sets of conserved eye-specific genes. DESeq, edgeR, NOISeq are applied using species as replicates (3 eyes vs 3 wings). The top 250 most differentially expressed genes are used as input for cisTargetX. Note that this number cannot be derived by GORilla, but is used to compare sets with similar sizes. The sizes defined by GORilla on the edgeR and NOISeq results do not perform well on cisTargetX, which argues for using OS. The cisTargetX results regarding the glass motif are not very different between these sets, the glass motif is always found to be highly enriched, on conserved eye-specific gene sets, regardless of the method used to integrate the data across species. Even though at Gene Ontology level there are differences between the methods (see main text, Table 1), these differences are not reflected at the motif level, most likely because a minimum number of Glass targets is always present, leading to the robust identification of the motif by cisTargetX.

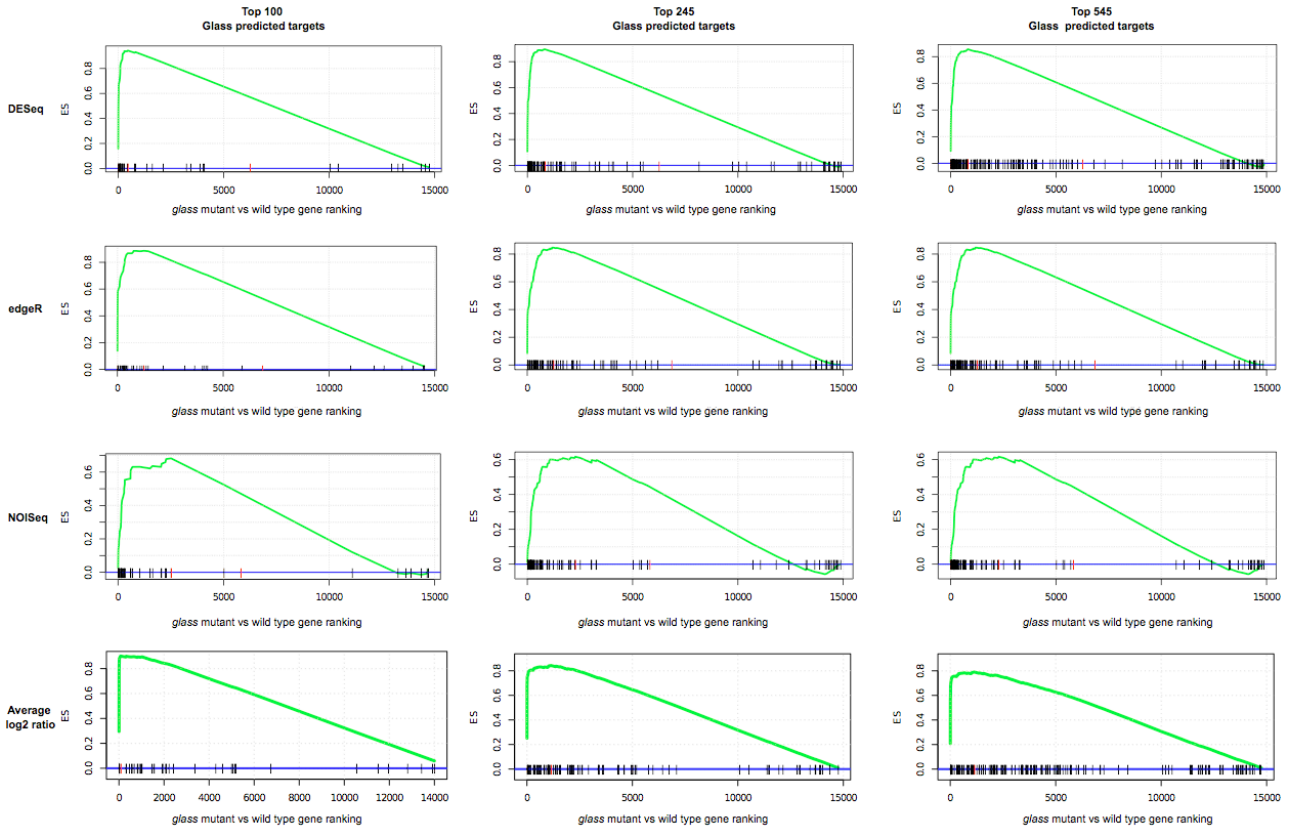


Figure S16: Gene Set Enrichment Analyses of predicted Glass targets in RNA-seq data

GSEA plots showing the significant enrichment of predicted Glass targets obtained from different sizes of input gene sets to cisTargetX (top 100, tops 245 and top 545 eye-enriched genes; in columns), using different differential expression methods for RNA-seq data on the x-axis (in the rows). All methods agree in finding an enrichment for down-regulation on the predicted Glass targets.

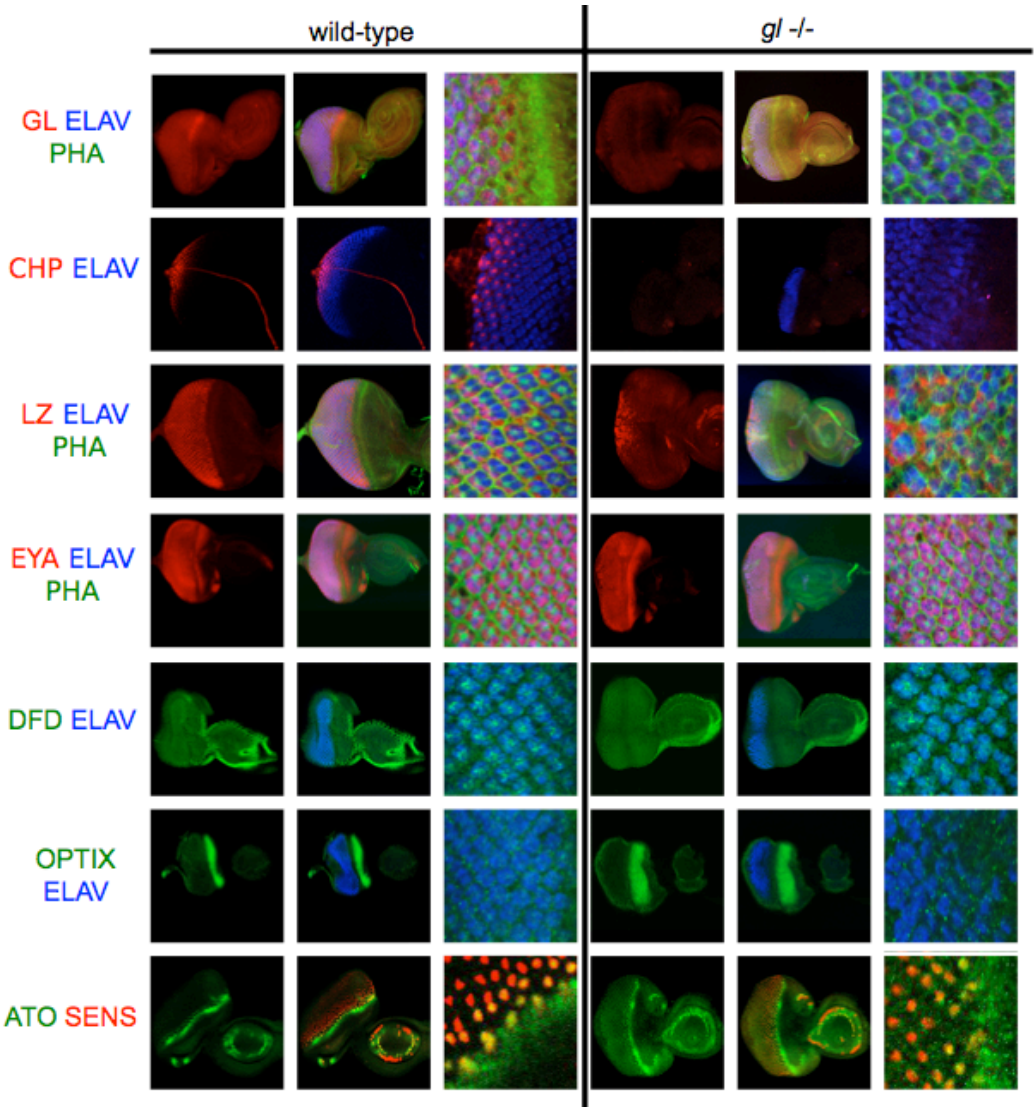


Figure S17: Antibody stainings for a set of predicted Glass targets in wild-type and *gl* -/- eye imaginal discs. Expression of validated targets (*gl*, *chp*, *lz)* is clearly affected in the *glass* mutant. Expression of invalidated targets (*eya*, *Dfd*, *Optix*, and *ato*) is unaffected in the *glass* mutant.

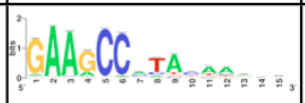
TF	LOGO	ID	Top 245 eye genes		96 predicted Glass targets		62 predicted Glass targets strongly down-regulated in *gl* -/-		34 predicted Glass targets no strongly down-regulated in *gl* -/-	
			Motif rank	NES	Motif rank	NES	Motif rank	NES	Motif rank	NES
GL		flyfactorsurvey-gl_SOLEXA_5_FBgn0004618	1	4.7523	**1**	**6.3807**	**1**	**6.3227**	-	-
SCRT		flyfactorsurvey-scrt_2.5_FBgn0004880	**3**	***3.5068***	**13**	**2.9795**	-	-	-	-
SU(H)		jaspar4-MA0085.1-Su_H	15	*3.1040*	12	2.98250	-	-	**12**	**4.0348**
SOB/ODD		flyfactorsurvey-sob_SOLEXA_5_FBgn0004892	41	2.5574	44	2.6033	-	-	**1**	**5.4369**

Figure S18: Motif discovery by cisTargetX on subsets of genes.

CisTargetX is first applied on the set of 245 conserved eye-specific genes, yielding 96 candidate Glass targets. This set is then split into 62 validated targets by RNA-seq, and 34 invalidated targets. All these sets are analyzed again with cisTargetX, showing that the invalidated targets are not enriched for the Glass motif anymore.

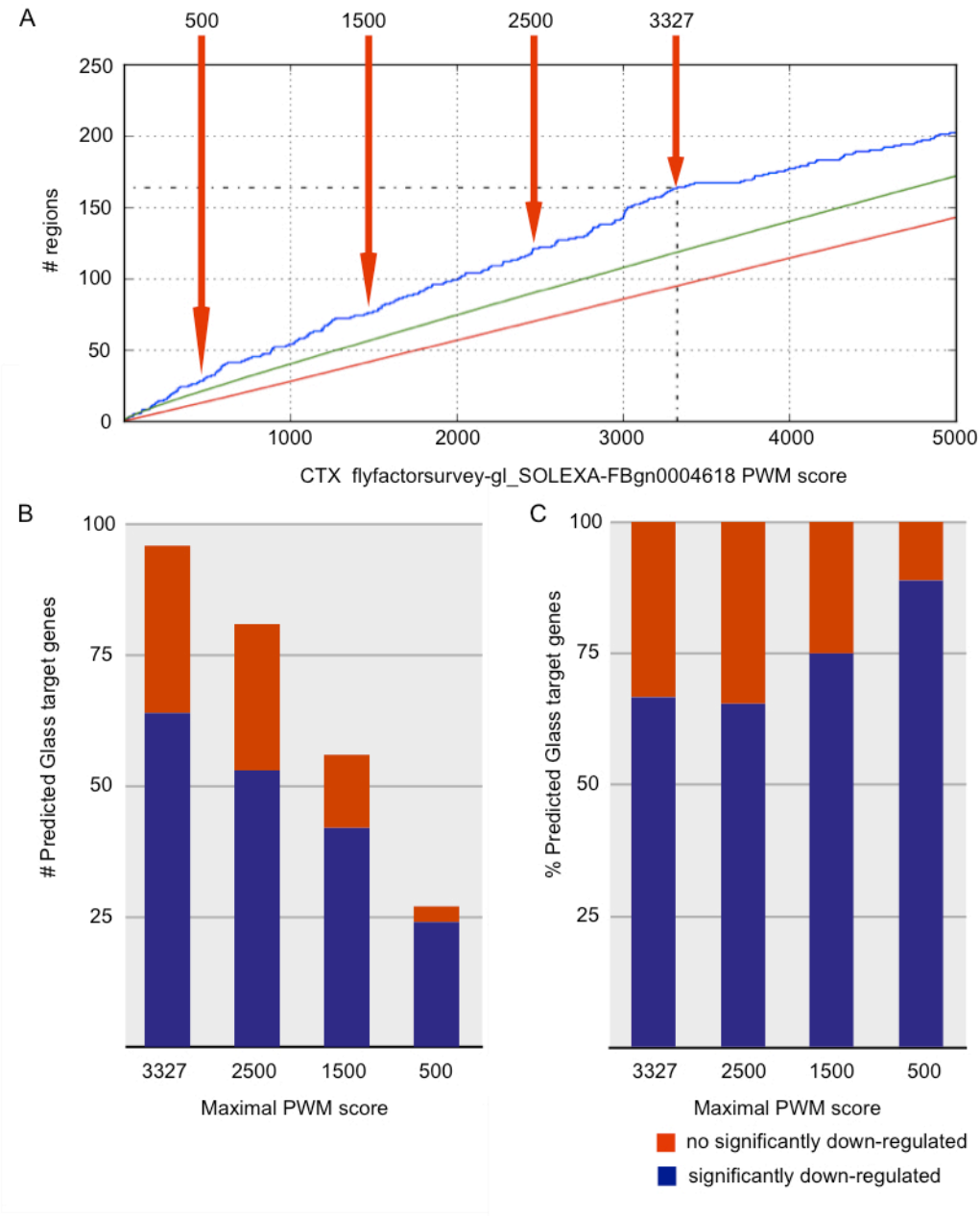


Figure S19: Relation between the cisTargetX genomic ranking and false positive rates.

A) cisTargetX result showing the recovery of the input set (245 conserved eye-specific genes) along the genomic ranking generated by a cross-species ClusterBuster scoring with the Glass PWM. This recovery curve is found with the highest, and most significant, Area Under the Curve (AUC) among all tested PWMs. The automatically defined threshold is at a genomic rank of 3327, determined as the point with the largest difference between the observed (blue) curve and the average curve plus 2 standard deviations (green curve). This is the situation used in the text, and yields 62 distinct predicted Glass targets. We then lowered this cut-off to select smaller sets of high-scoring genes. The more we move this threshold to the left, selecting only better ranked genes, the better the true positive rate (more significantly down-regulated genes in the RNA-seq data).

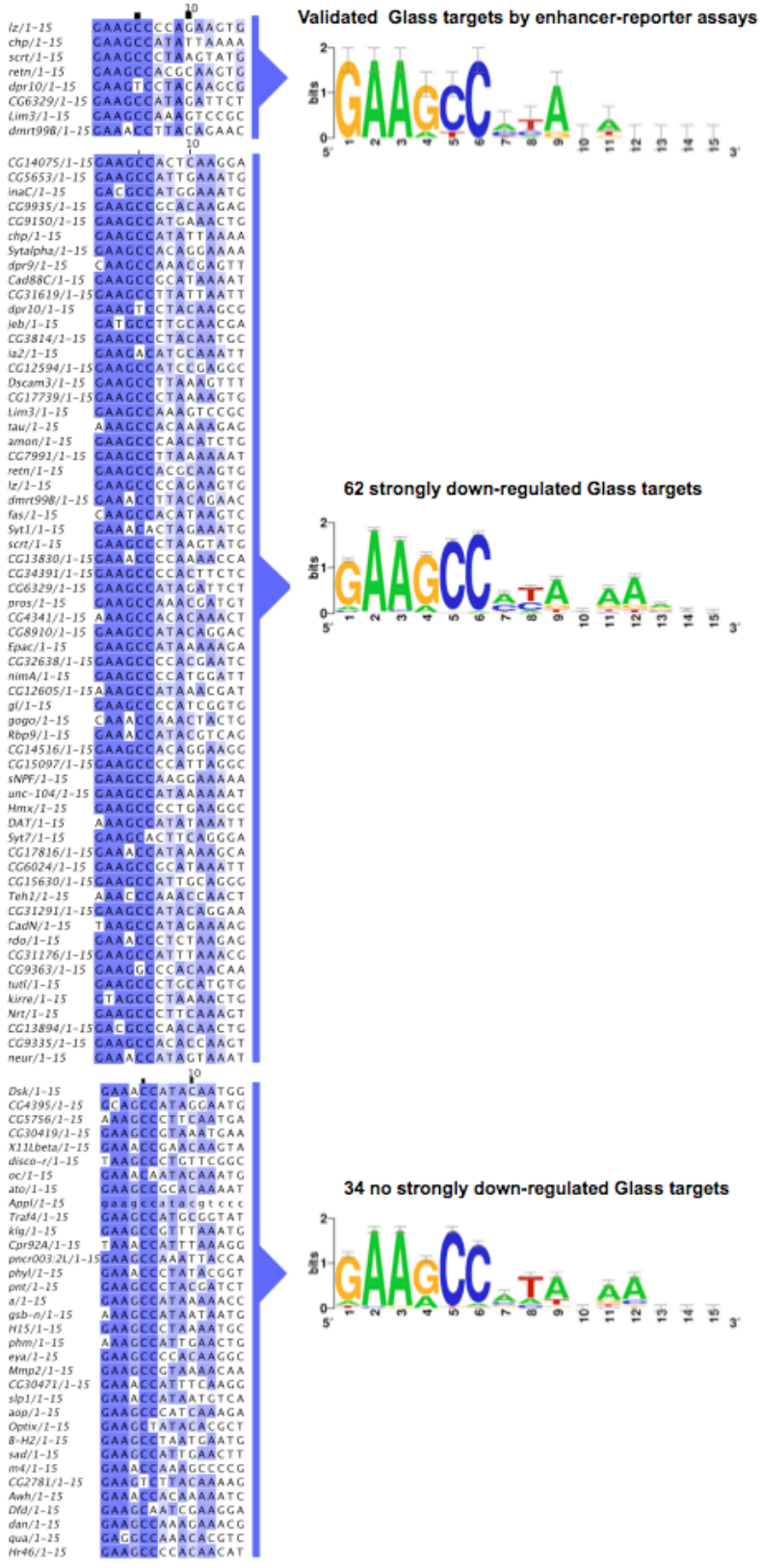


Figure S20: Individual Glass binding sites

Comparison of the maximum scoring Glass binding site, within the maximum scoring CRM, between validated Glass targets by enhancer-reporter assays, 62 predicted Glass targets significantly down-regulated in gl-/- and 34 invalidated Glass targets.

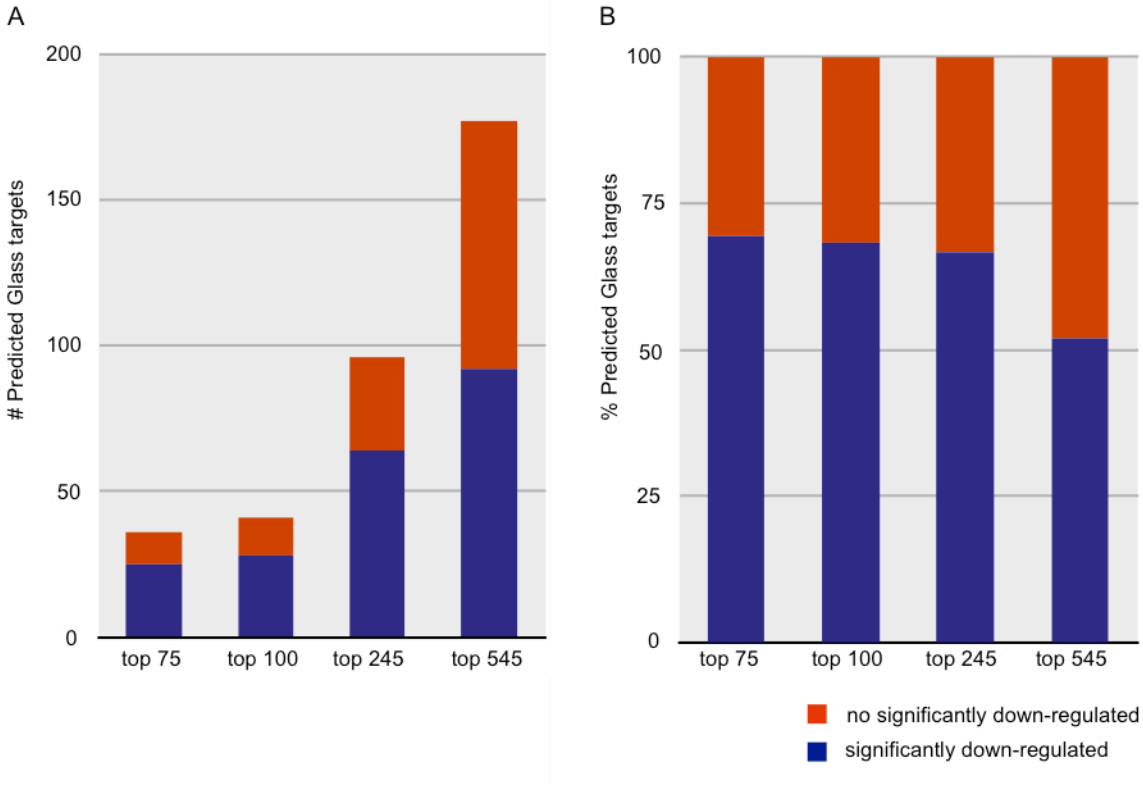


Figure S21: Relation between different sizes of input sets and the rate of false positive Glass target predictions.

Graphs presenting the absolute number (A) and the percentage (B) of significantly (blue) versus non-significantly (red) down-regulated Glass targets. Each bar represents a different gene set that was used as input for cisTargetX, using different thresholds (75, 100, 245, and 545) on the OS-based eye-vs-wing gene ranking across species. The percentage of predicted Glass targets that are significantly down-regulated in the *glass* mutant remains nearly constant while the input set increases in size from 75 to 245 genes. Only when using 545 genes, there is an increase in the number of false positive predictions (genes that are not significantly down-regulated).

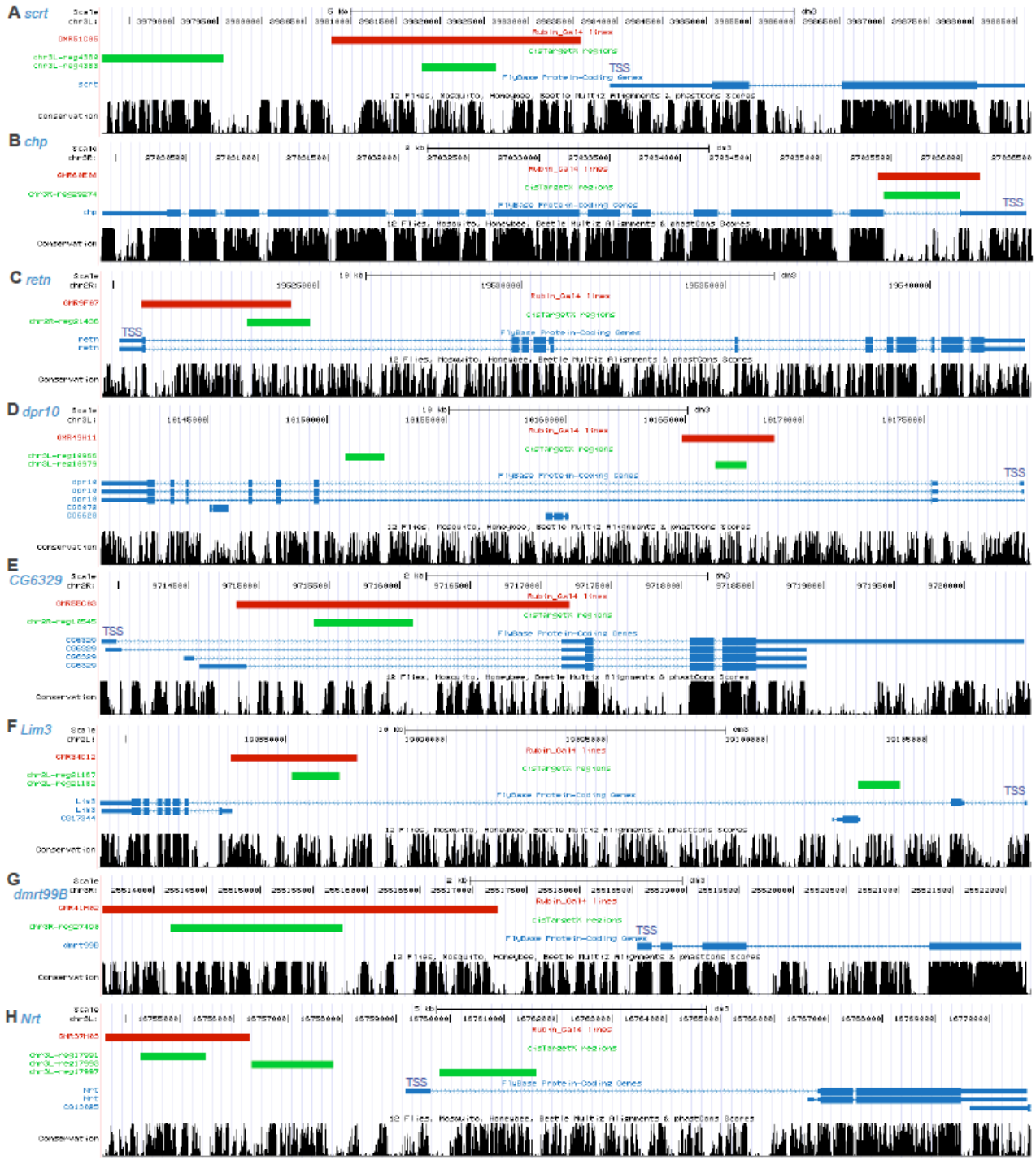


Figure S22: Visualization by UCSC genome browser of CRMs with predicted glass-binding sites (green) and enhancer-GAL4 regions (red) for nine predicted glass target genes. A) *scrt*, B) *chp*, C) *dpr10*, D) *CG6329*, E) *retn*, F) *Lim3*, G) *dmrt99B*, H) *Nrt*.

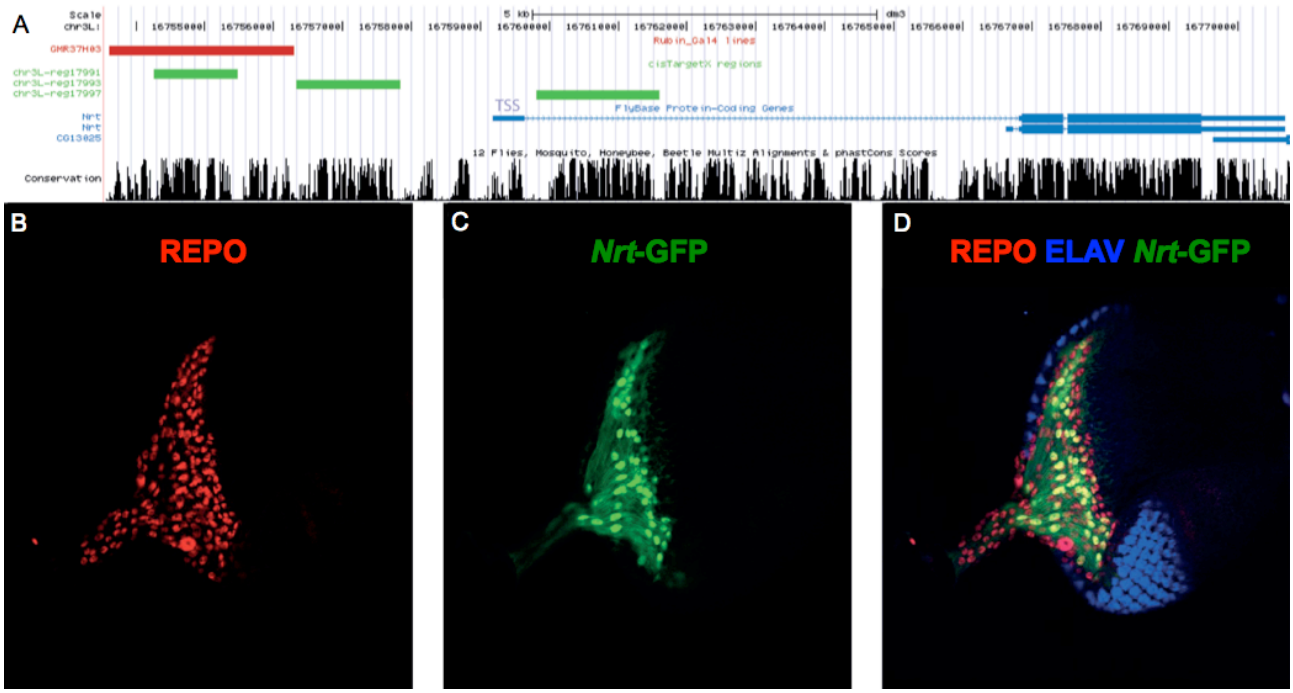


Figure S23: *Nrt* tested enhancer drives expression in glia. A) Visualization by UCSC genome browser of the *Nrt* gene (blue) with predicted cisTargetX glass-binding sites regions (green) and enhancer-GAL4 line (Pfeiffer et al. 2008) spanning a predicted CRM. A-C) Immunostaining of an eye-antennal imaginal disc in third instar larvae with enhancer-GAL4 x UAS-GFP. Antibodies are Anti-repo (red) , anti-elav (blue) and anti-GFP (green).

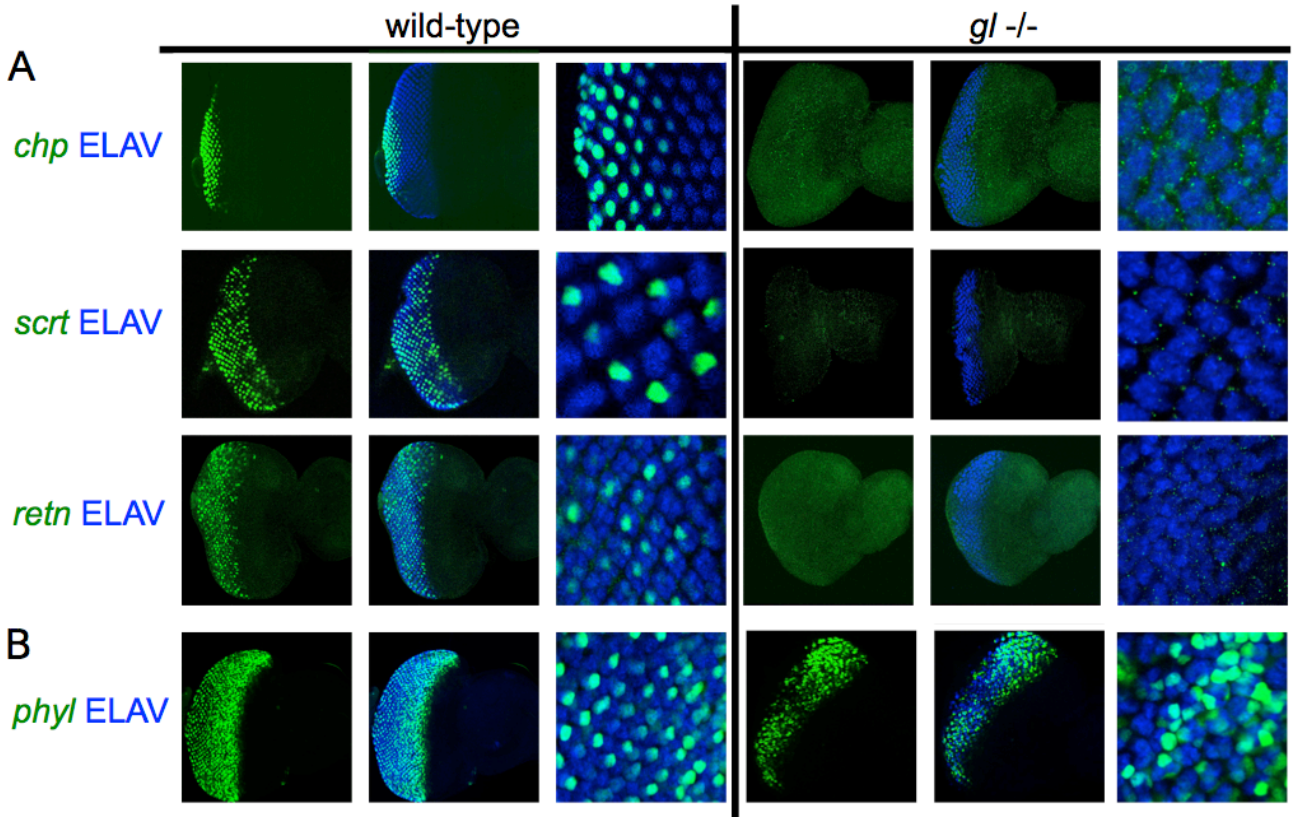


Figure S24: Enhancer-GFP in wt and *gl*-/-.

A) Enhancer-GFP for *chp*, *scrt* and *retn*, genes strongly down-regulated in *gl* -/-. The enhancer-reporter activity is gone or severely affected in the *glass* mutant. B) Same experiment for the enhancer-GFP reporter construct of the *phyl* CRM. This is the CRM targeted by Atonal, as previously reported in (Aerts *et al*. 2010).

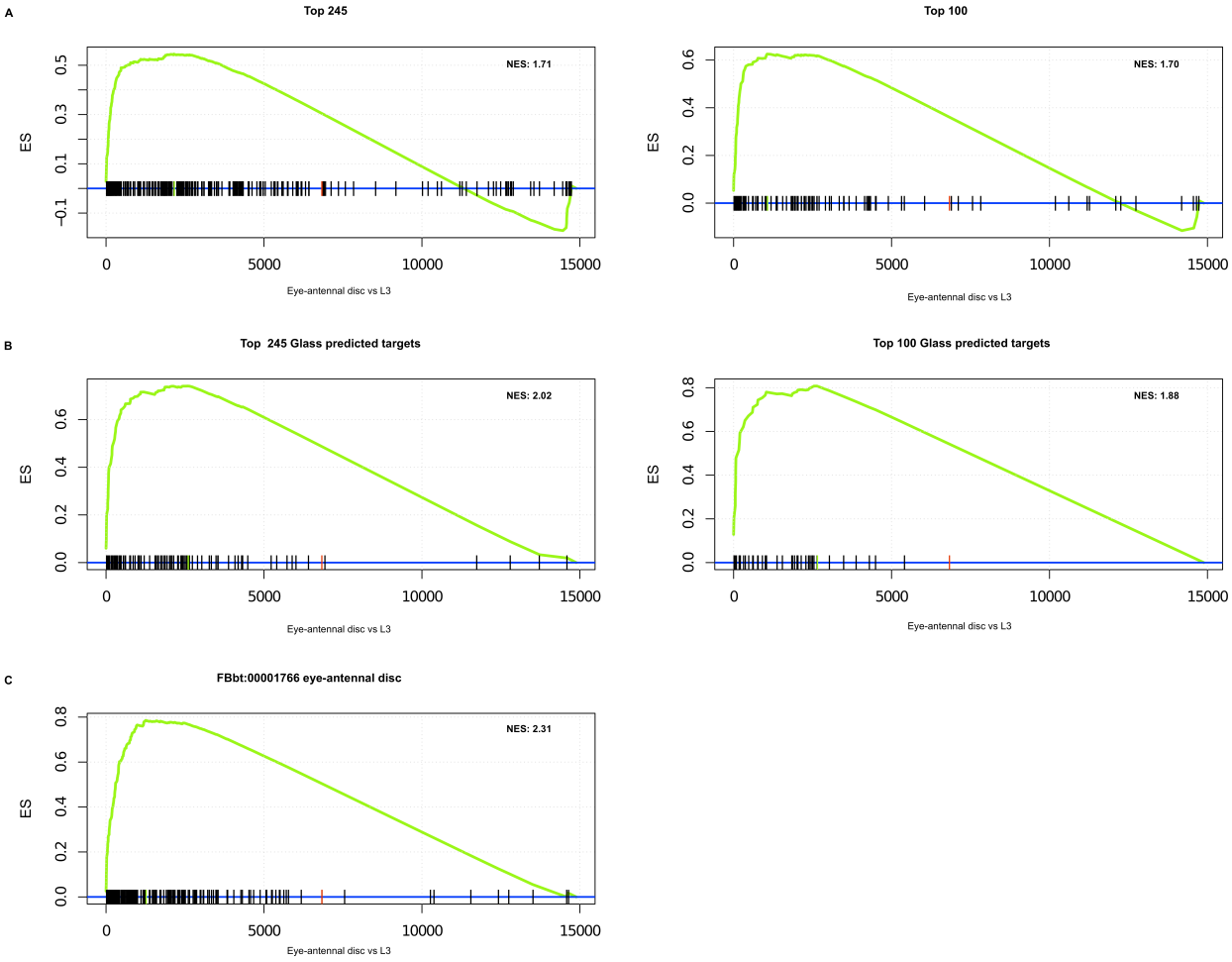


Figure S25: Gene Set Enrichment Analyses (GSEA) using L3 larvae RNA-seq data.

A) Top 245 and top 100 eye-enriched genes using Tag-seq cross-species. B) Predicted Glass targets from the top 245 and top 100 eye-enriched genes by cisTargetX, showing that Glass target gene predictions cause a selection for genes enriched in eye-antennal disc compared to total larval RNA. C) 220 genes expressed in “eye-antennal disc”, as annotated by the FlyBase Term FBbt:00001766, showing that such genes may still be enriched in total larvae RNA compared to the eye-antennal disc.

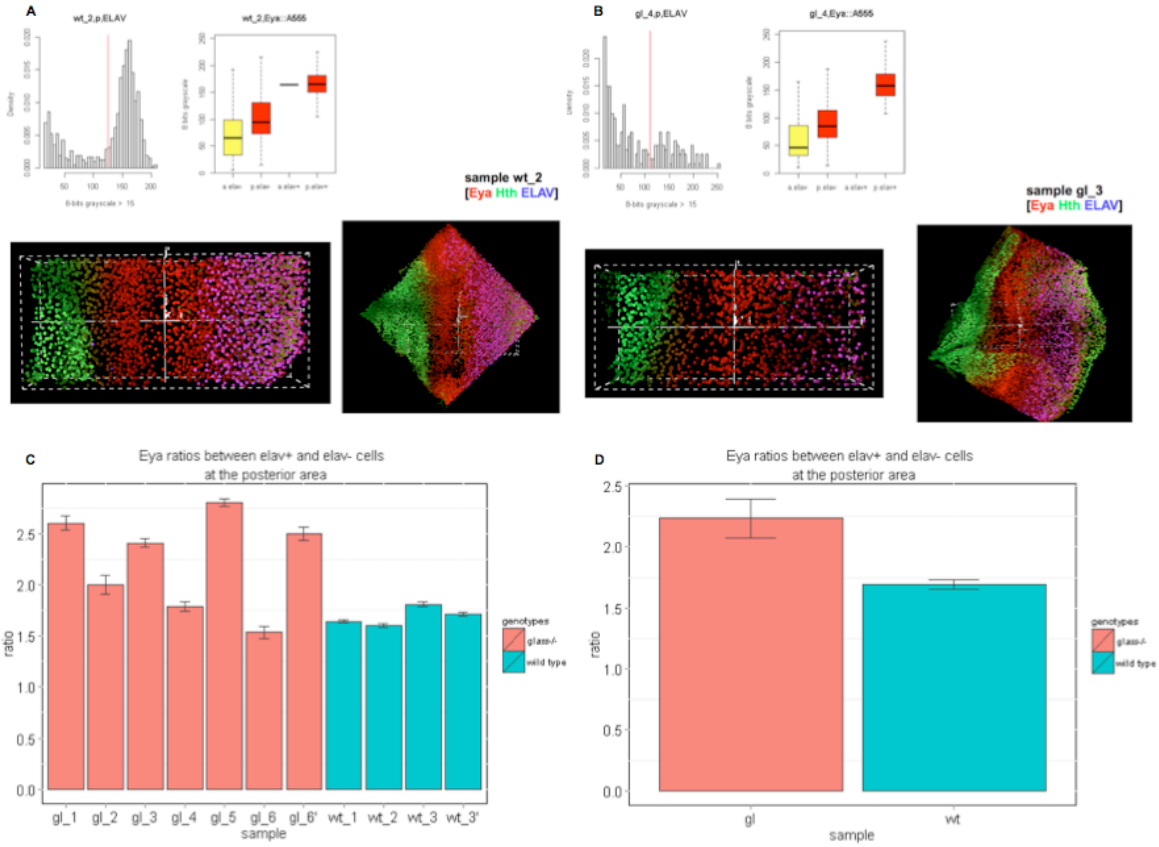


Figure S26: EYA quantification in wild-type and glass mutant eye discs.

A) Example of one wild type eye disc showing the raw Eya quantification data in ELAV- and ELAV+ cells. B) Example of one glass mutant disc showing raw EYA quantification in ELAV- and ELAV+ cells. C) Normalized ratios (ELAV+ over ELAV-) of EYA expression comparing multiple wild type (blue) and *glass* mutant discs (red). D) Average of the ratios shown in C). Details on the quantification steps are provided in the Materials and Methods Section.

Table S1: Mapping results

Species	Tissue	Number of outputted reads	Genome Assembly	% mapped reads	Annotation	% Reads falling within annotation out of total	% Tags falling within annotation out of mapped tags
Drosophila melanogaster	eye-antennal	4667043	FlyBase 5	47,30	FlyBase 5.30	41,30	87,32
	wing	2447992	FlyBase 5	76,00	FlyBase 5.30	69,39	91,31
Drosophila yakuba	eye-antennal	6844469	FlyBase 1	82,89	FlyBase 1.3	34,38	41,48
			UCSC droYak2	83,61	*Drosophila melanogaster* FlyBase 5.30	70,54	84,37
	wing	2719615	FlyBase 1	80,92	FlyBase 1.3	34,15	42,21
			UCSC droYak2	82,08	*Drosophila melanogaster* FlyBase 5.30	69,41	84,56
Drosophila virilils	eye-antennal	4676180	FlyBase 1	44,18	FlyBase 1.2	22,18	50,20
			UCSC droVir3	44,18	*Drosophila melanogaster* FlyBase 5.30	27,59	62,44
	wing	4173261	FlyBase 1	75,06	FlyBase 1.2	41,88	55,79
			UCSC droVir3	75,06	*Drosophila melanogaster* FlyBase 5.30	50,59	67,39

Table S2: Gene Ontology Enrichment Analysis for individual species *D.melanogaster*, *D.yakuba*, *D.virilis* log2 (eye-antennal) rankings, and considering species as replicates. For edgeR (Robinson, McCarthy, and Smyth 2010) and DESeq (Anders and Huber 2010) genes are ranked according to their differential expression between eye-antennal and wing imaginal discs using the signed -log10(p.value), for one-to-one orthologs. For NOISeq (Tarazona et al. 2011) is the ranking is based on the probability of differential expression. The cross-species ranking is based on the order statistics integration of the three independent logratio rankings of the three species (Aerts et al. 2006).

Independent file : TS2.xls

Table S3: Gene counts for eye-antennal and wing imaginal discs in *D.melanogaster*, *D.yakuba* and *D.virilis*. Raw counts, normalized counts and logratio (eye-antennal / wing). For *D.yakuba* and *D.virilis* the corresponding *D.melanogaster* FBgn identifiers are added, as well as the methodology used, being either mapping by orthology using GeneTrees (Vilella et al. 2009) or mapping by whole-genome alignment using the UCSC genome browser (Fujita et al. 2010). Expression values are those after low-count filtering. Raw count data with genomic positions can be obtained from GEO (accession number GSE39784).

Independent file : TS3.xls

Table S4: Predicted Transcription Factor – Target Gene regulatory interactions extracted from cisTargetX outputs from the top 100 and 245 genes in cross-species order statistics ranking.

ey	*Appl*	*gl*	*CG5653*	*gl*	*pros*
ey	*CG31176*	*gl*	*CG5756*	*gl*	*qua*
ey	*CG31221*	*gl*	*CG6024*	*gl*	*Rbp9*
ey	*CG6024*	*gl*	*CG6329*	*gl*	*rdo*
ey	*CG6327*	*gl*	*CG7991*	*gl*	*retn*
ey	*CG7991*	*gl*	*CG8910*	*gl*	*sad*
ey	*dan*	*gl*	*CG9150*	*gl*	*scrt*
ey	*DAT*	*gl*	*CG9335*	*gl*	*slp1*
ey	*dmrt99B*	*gl*	*CG9363*	*gl*	*sNPF*
ey	*dpr12*	*gl*	*CG9935*	*gl*	*Syt1*
ey	*dpr9*	*gl*	*chp*	*gl*	*Syt7*
ey	*gsb-n*	*gl*	*Cpr92A*	*gl*	*Sytalpha*
ey	*Lim3*	*gl*	*dan*	*gl*	*tau*
ey	*oc*	*gl*	*DAT*	*gl*	*Teh1*
ey	*ru*	*gl*	*Dfd*	*gl*	*Traf4*
ey	*slp1*	*gl*	*disco-r*	*gl*	*tutl*
ey	*sNPF*	*gl*	*dmrt99B*	*gl*	*unc-104*
ey	*Traf4*	*gl*	*dpr10*	*gl*	*X11Lbeta*
gl	*a*	*gl*	*dpr9*	*scrt*	*a*
gl	*amon*	*gl*	*Dscam3*	*scrt*	*amon*
gl	*aop*	*gl*	*Dsk*	*scrt*	*Ank2*
gl	*Appl*	*gl*	*Epac*	*scrt*	*aop*
gl	*ato*	*gl*	*eya*	*scrt*	*ato*
gl	*Awh*	*gl*	*fas*	*scrt*	*Awh*
gl	*B-H2*	*gl*	*gl*	*scrt*	*betaTub60D*
gl	*Cad88C*	*gl*	*gogo*	*scrt*	*B-H2*
gl	*CadN*	*gl*	*gsb-n*	*scrt*	*CadN*
gl	*CG12594*	*gl*	*H15*	*scrt*	*Cbp53E*
gl	*CG12605*	*gl*	*Hmx*	*scrt*	*CG12594*
gl	*CG13830*	*gl*	*Hr46*	*scrt*	*CG12605*
gl	*CG13894*	*gl*	*ia2*	*scrt*	*CG13560*
gl	*CG14075*	*gl*	*inaC*	*scrt*	*CG13650*
gl	*CG14516*	*gl*	*jeb*	*scrt*	*CG13928*
gl	*CG15097*	*gl*	*kirre*	*scrt*	*CG14075*
gl	*CG15630*	*gl*	*klg*	*scrt*	*CG15097*
gl	*CG17739*	*gl*	*Lim3*	*scrt*	*CG15630*
gl	*CG17816*	*gl*	*lz*	*scrt*	*CG18304*
gl	*CG2781*	*gl*	*m4*	*scrt*	*CG2781*
gl	*CG30419*	*gl*	*Mmp2*	*scrt*	*CG30419*
gl	*CG30471*	*gl*	*neur*	*scrt*	*CG31176*
gl	*CG31176*	*gl*	*nimA*	*scrt*	*CG31221*
gl	*CG31291*	*gl*	*Nrt*	*scrt*	*CG31291*
gl	*CG31619*	*gl*	*oc*	*scrt*	*CG31619*
gl	*CG32638*	*gl*	*Optix*	*scrt*	*CG32354*
gl	*CG34391*	*gl*	*phm*	*scrt*	*CG33463*
gl	*CG3814*	*gl*	*phyl*	*scrt*	*CG34391*
gl	*CG4341*	*gl*	*pncr003:2L*	*scrt*	*CG42251*
gl	*CG4395*	*gl*	*pnt*	*scrt*	*CG42322*

scrt	*CG42390*
scrt	*CG42749*
scrt	*CG4341*
scrt	*CG6024*
scrt	*CG6327*
scrt	*CG7991*
scrt	*CG8216*
scrt	*CG8910*
scrt	*CG9134*
scrt	*CG9363*
scrt	*comm*
scrt	*CR30009*
scrt	*cu*
scrt	*dac*
scrt	*dan*
scrt	*DAT*
scrt	*Dfd*
scrt	*Dif*
scrt	*disco-r*
scrt	*dpr10*
scrt	*Dscam3*
scrt	*eya*
scrt	*fas*
scrt	*gl*
scrt	*gogo*
scrt	*gsb-n*
scrt	*H15*
scrt	*hiw*
scrt	*Hmx*
scrt	*Hr46*
scrt	*inaC*
scrt	*ia2*
scrt	*jdp*
scrt	*jeb*
scrt	*kirre*
scrt	*klg*
scrt	*lab*
scrt	*Lim3*
scrt	*m4*
scrt	*Mmp2*
scrt	*nerfin-1*
scrt	*neur*
scrt	*nrv3*
scrt	*pgant2*
scrt	*phyl*
scrt	*pncr002:3R*
scrt	*pnt*
scrt	*qtc*
scrt	*qua*
scrt	*Rab3*
scrt	*rdo*
scrt	*retn*
scrt	*rho*
scrt	*ru*
scrt	*scrt*
scrt	*sNPF*
scrt	*so*
scrt	*SoxN*
scrt	*spir*
scrt	*stj*
scrt	*stops*
scrt	*Syt1*
scrt	*Sytalpha*
scrt	*Teh1*
scrt	*tej*
scrt	*toy*
scrt	*Traf4*
scrt	*tutl*
scrt	*X11Lbeta*
SoxN	*amon*
SoxN	*CG11073*
SoxN	*CG12071*
SoxN	*CG12605*
SoxN	*CG15630*
SoxN	*CG30419*
SoxN	*CG31176*
SoxN	*Dscam3*
SoxN	*ey*
SoxN	*gl*
SoxN	*Hmx*
SoxN	*Lim3*
SoxN	*oc*
SoxN	*ru*
SoxN	*scrt*
SoxN	*slp1*
SoxN	*SoxN*
Su(H)	*a*
Su(H)	*aop*
Su(H)	*ato*
Su(H)	*Cbp53E*
Su(H)	*CG15097*
Su(H)	*CG31176*
Su(H)	*CG31676*
Su(H)	*CG42390*
Su(H)	*CG42492*
Su(H)	*CG8910*
Su(H)	*CR30009*
Su(H)	*D*
Su(H)	*DAT*
Su(H)	*Dfd*
Su(H)	*Dif*
Su(H)	*disco-r*
Su(H)	*eya*
Su(H)	*fas*
Su(H)	*gl*
Su(H)	*HLHmgamma*
Su(H)	*Hmx*
Su(H)	*inaC*
Su(H)	*jeb*
Su(H)	*kirre*
Su(H)	*m4*
Su(H)	*nerfin-1*
Su(H)	*neur*
Su(H)	*nimA*
Su(H)	*phyl*
Su(H)	*pnt*
Su(H)	*pros*
Su(H)	*PsGEF*
Su(H)	*qua*
Su(H)	*retn*
Su(H)	*rho*
Su(H)	*run*
Su(H)	*Teh1*
Su(H)	*Traf4*
Su(H)	*tutl*

Table S5: Conserved Glass targetome. Eye-enriched genes across species predicted as Glass target genes *and* significantly down-regulated in gl[60j] versus *D.melanogaster* wild-type.

Gene Name	*D.melanogaster* Gene FBgn	*D.melanogaster* wild-type	*D.yakuba* wild-type	*D.virilis* wild-type	gl[60j]	Gl[60j] *P.value*	Gl[60j] *P.adj*	cisTargetX predicted glass binding region
CG14075	FBgn0036835	6.97	NA	4.31	-10.00	8.39E-114	3.67E-111	chr3L:18976192-18978113
CG5653	FBgn0035943	4.14	4.25	2.96	-8.80	3.80E-223	4.16E-220	chr3L:8959462-8959760
inaC	FBgn0004784	5.02	2.77	6.01	-8.68	2.06E-179	2.00E-176	chr2R:12783842-12786149
CG9935	FBgn0039916	5.15	5.01	3.51	-6.37	4.10E-237	7.17E-234	chr4:667146-667344
CG9150	FBgn0031775	0.11	5.75	-1.21	-5.02	3.88E-65	5.06E-63	chr2L:6068579-6069220
chp	FBgn0000313	3.54	4.51	6.85	-4.44	0	0	chr3R:27035441-27035982
Sytalpha	FBgn0261089	2.13	1.56	1.20	-3.65	1.01E-29	3.78E-28	chr2L:17595313-17596866
dpr9	FBgn0038282	4.53	2.33	4.50	-2.85	3.77E-55	3.74E-53	chr3R:10892805-10893925
Cad88C	FBgn0038247	4.04	2.24	0.92	-2.82	5.94E-115	2.73E-112	chr3R:10455668-10456947
CG31619	FBgn0051619	3.89	3.70	2.90	-2.75	1.80E-140	1.31E-137	chr2L:21699086-21700671
dpr10	FBgn0052057	2.49	3.60	1.43	-2.67	3.44E-25	1.02E-23	chr3L:10150740-10152379 chr3L:10166262-10167555
jeb	FBgn0086677	4.14	2.87	2.29	-2.58	1.05E-79	2.03E-77	chr2R:8005565-8007491
CG3814	FBgn0025692	2.42	2.15	1.75	-2.42	1.27E-042	8.13E-043	chr2R:8764295-8765159
ia2	FBgn0031294	5.89	-0.93	6.35	-2.38	0.00E+000	0.00E+000	chr2L:1046329-1047548 chr2L:1029518-1032768 chr2L:1038422-1039823
CG12594	FBgn0037941	3.80	0.44	4.05	-2.29	6.87E-22	1.62E-20	chr3R:7546101-7547562
Dscam3	FBgn0261046	2.01	5.48	3.93	-2.24	1.94E-70	2.99E-68	chr3R:13324027-13325413

								chr3R:13334232-13335138
CG17739	FBgn0033710	2.41	2.10	3.33	-2.15	3.32E-32	1.40E-30	chr2R:8194164-8194792
Lim3	FBgn0002023	3.73	2.97	3.67	-2.07	3.66E-34	1.72E-32	chr2L:19085176-19086688
								chr2L:19102852-19104157
tau	FBgn0051057	5.59	1.87	3.33	-1.92	9.48E-47	7.03E-45	chr3R:23476471-23477585
								chr3R:23479049-23481095
amon	FBgn0023179	4.89	1.72	4.59	-1.89	1.41E-42	8.43E-41	chr3R:22530095-22531144
								chr3R:22533998-22535391
retn	FBgn0004795	6.82	5.18	3.83	-1.81	3.57E-65	4.73E-63	chr2R:19514693-19515935
								chr2R:19523270-19524801
CG7991	FBgn0035260	3.52	1.75	3.85	-1.80	2.74E-41	1.56E-39	chr3L:1679635-1680795
lz	FBgn0002576	6.38	3.99	4.50	-1.75	1.68E-51	1.40E-49	chrX:9180815-9181775
dmrt99B	FBgn0039683	2.71	3.51	6.56	-1.72	1.54E-18	2.99E-17	chr3R:25514163-25515780
fas	FBgn0000633	3.80	-0.04	4.05	-1.71	2.67E-31	1.07E-29	chr2R:9527732-9529592
								chr2R:9537569-9539962
								chr2R:9546761-9549179
Syt1	FBgn0004242	0.94	1.52	5.65	-1.71	6.80E-51	5.51E-49	chr2L:2793888-2794819
scrt	FBgn0004880	5.97	5.04	7.05	-1.68	4.72E-59	5.23E-57	chr3L:3978149-3979559
								chr3L:3981801-3982640
CG13830	FBgn0039054	3.80	2.39	1.77	-1.68	1.16E-62	1.43E-60	chr3R:18932619-18933865
								chr3R:18936623-18937928
								chr3R:18938656-18940394

CG34391	FBgn0085420	4.89	2.14	2.29	-1.66	8.29E-11	9.06E-10	chr3L:5423645-5424733
CG6329	FBgn0033872	4.68	0.73	3.78	-1.66	9.73E-43	5.95E-41	chr2R:9715391-9716094
pros	FBgn0004595	2.22	1.77	4.05	-1.59	2.14E-52	1.91E-50	chr3R:7191333-7193469 chr3R:7202998-7203728
CG4341	FBgn0028481	4.59	3.51	3.66	-1.58	3.48E-30	1.33E-28	chr2L:930423-931744 chr2L:943500-945037
CG8910	FBgn0025833	3.15	1.62	5.41	-1.51	2.45E-12	3.04E-11	chr2R:1276554212767777 chr2R:12772005-12773682 chr2R:12775699-12777015
Epac	FBgn0085421	2.29	2.49	2.29	-1.45	7.54E-17	1.31E-15	chr2R:2670842-2672958
CG32638	FBgn0052638	2.08	1.89	1.22	-1.43	2.00E-40	1.10E-38	chrX:13020247-13021312
nimA	FBgn0261514	2.83	5.41	-0.78	-1.43	6.32E-14	8.90E-13	chr2L:13958152-13958881
CG12605	FBgn0035481	5.59	3.46	1.78	-1.35	2.07E-26	6.52E-25	chr3L:3968149-3969195
gl	FBgn0004618	8.34	5.03	8.10	-1.27	2.11E-35	1.01E-33	chr3R:14199286-14201326
gogo	FBgn0052227	3.22	3.61	3.96	-1.25	1.47E-33	6.80E-32	chr3L:20287914-20288944
Rbp9	FBgn0010263	1.56	0.76	1.71	-1.12	1.04E-07	8.13E-07	chr2L:2959916-2961230
CG14516	FBgn0039640	5.64	5.03	4.36	-1.12	5.15E-27	1.67E-25	chr3R:24967291-24967931
sNPF	FBgn0032840	4.80	2.57	5.06	-1.10	1.85E-09	1.80E-08	chr2L:20029536-20030549
CG15097	FBgn0034396	2.62	2.90	2.52	-1.06	5.70E-14	8.05E-13	chr2R:14714225-14715260 chr2R:14717446-14718750
unc-104	FBgn0034155	1.98	2.14	2.48	-1.04	1.96E-24	5.58E-23	chr2R:12652314-12652894 chr2R:12657913-12658973
Hmx	FBgn0085448	4.48	6.43	4.26	-1.03	8.51E-10	8.53E-09	chr3R:13387092-13389664

								chr3R:13389665-13390518
DAT	FBgn0034136	3.89	4.20	5.29	-1.01	3.99E-05	0.0002	chr2R:12449980-12451275
Syt7	FBgn0039900	5.79	2.49	2.63	-0.98	7.25E-11	7.95E-10	chr4:311864-314134
CG17816	FBgn0037525	2.96	0.51	4.81	-0.94	8.06E-18	1.49E-16	chr3R:3701041-3701845
CG6024	FBgn0036202	4.07	3.86	4.88	-0.91	1.05E-18	2.07E-17	chr3L:11751117-11751844
								chr3L:11754456-11755537
CG15630	FBgn0031627	4.91	3.04	4.05	-0.80	1.20E-09	1.19E-08	chr2L:4763438-4764796
								chr2L:4766233-4767285
								chr2L:4770866-4773315
Teh1	FBgn0037766	0.45	0.14	3.93	-0.79	0.0079	0.0247	chr3R:5682324-5683532
CG31291	FBgn0051291	3.60	3.51	1.20	-0.72	0.0001	0.0006	chr3R:11746270-11748484
CadN	FBgn0015609	3.17	1.16	4.08	-0.72	6.67E-12	7.99E-11	chr2L:17721104-17723799
rdo	FBgn0243486	1.20	2.24	3.45	-0.67	8.84E-11	9.64E-10	chr2L:18013311-18014362
								chr2L:18022951-18024321
								chr2L:18033883-18036251
CG31176	FBgn0051176	5.72	4.02	2.90	-0.53	0.0003	0.0013	chr3R:17503347-17504463
								chr3R:17505224-17505941
								chr3R:17507016-17507973
								chr3R:17507974-17508959
								chr3R:17533316-17534391
								chr3R:17534392-17535455
								chr3R:17539746-17541493
								chr3R:17546238-17547251

CG9363	FBgn0037697	1.98	2.72	1.88	-0.53	2.56E-06	1.66E-05	chr3R:5286715-5290474
tutl	FBgn0010473	1.82	1.14	1.34	-0.52	7.90E-07	5.53E-06	chr2L:4283978-4286386
								chr2L:4287272-4289013
kirre	FBgn0028369	1.81	2.30	1.18	-0.51	3.60E-05	0.0002	chrX:2874471-2875512
								chrX:2894636-2896901
								chrX:2901563-2902587
								chrX:2904224-2905113
								chrX:2918585-2919771
								chrX:2944731-2945464
								chrX:2976182-2977483
								chrX:3010832-3011921
								chrX:3018454-3019320
Nrt	FBgn0004108	2.72	2.38	2.45	-0.46	1.76E-05	0.0001	chr3L:16754263-16755474
								chr3L:16756323-16757832
								chr3L:16759811-16761602
CG13894	FBgn0035157	2.15	2.14	1.10	-0.43	0.0002	0.0008	chr3L:699372-700357
CG9335	FBgn0032895	NA	3.11	9.83	-0.29	0.004	0.0142	chr2L:20853546-20854710
neur	FBgn0002932	1.20	2.30	2.00	-0.28	0.006	0.0202	chr3R:4859710-4860476

Table S6: Motif enrichment in various gene sets

Gene sets	Candidate Tfs/ motifs
Top 25	GL, KNI, ABD-B, OC
Top75	GL, SCRT, SOXNN, CG14962, EY
Top 100	GL, SCRT, SOXNN, SU(H)
Top 245	GL, SCRT, SOXN, SU(H)
Top 545	TRL, GL, SCRT, SU(H), HR46
96 predicted Glass targets	GL, PNT, SCRT
62 Glass targets significantly down-regulated in *gl*-/-	TRL, H, GRH
34 Glass targets no strongly down-regulated in the *gl*-/-	GL, DA, PNT
Intersection top 245 genes and genes up-regulated in ato-GOF (Aerts *et al.* 2010)	SCRT, GL, SU(H), ATO
Intersection top 245 and predicted EY targets (Ostrin *et al.* 2006)	E(SPL) DEAF1, OC
Eye-enriched genes eye-antennal (inhouse) vs L3 larva (modencode consortium)	SIX/Optix, EY