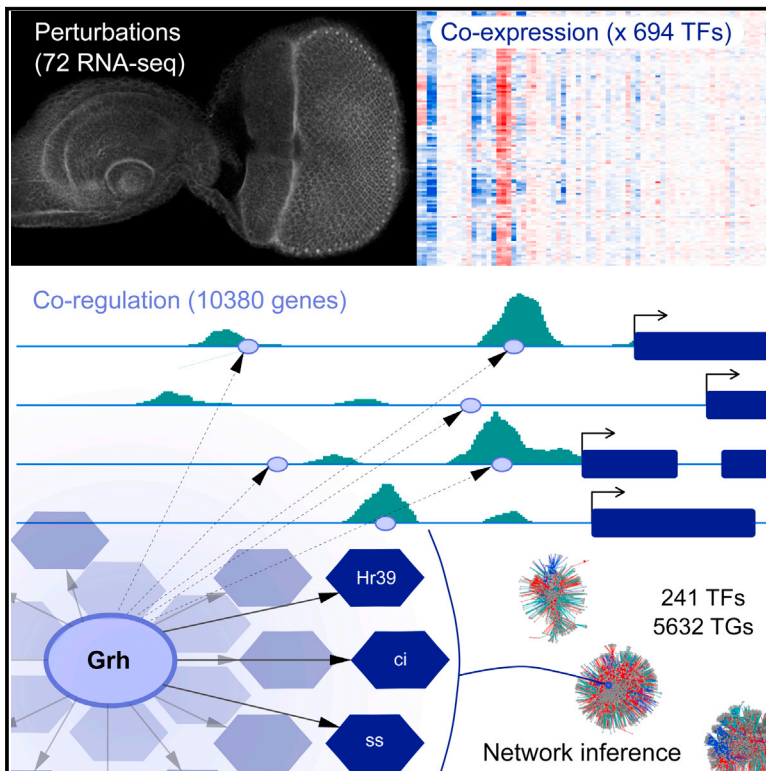# Mapping Gene Regulatory Networks in *Drosophila* Eye Development by Large-Scale Transcriptome Perturbations and Motif Inference

Graphical Abstract

Perturbations (72 RNA-seq)

Co-expression (x 694 TFs)

Co-regulation (10380 genes)

Hr39

Grh

ci

ss

241 TFs
5632 TGs

Network inference

Authors

Delphine Potier, Kristofer Davie, ...,
Valerie Christiaens, Stein Aerts

Correspondence

stein.aerts@med.kuleuven.be

In Brief

Potier et al. predict the gene regulatory network of *Drosophila* eye development by combining network inference on 72 perturbation RNA-seq experiments with motif prediction, adding many players to the eye network. ChIP-seq confirms the predicted widespread binding of Grainyhead and its correlation with open chromatin in the eye disc.

Highlights

- Developmental GRNs can be reverse engineered by RNA-seq and motif inference

- More than 200 TFs and their target genes are added to the *Drosophila* eye network

- Motifs of TFs are enriched in enhancers of their coexpressed genes

- Grainyhead targets are validated by ChIP-seq and correlate with chromatin opening

Accession Numbers

GSE62558

CellPress

Resource

Mapping Gene Regulatory Networks in *Drosophila* Eye Development by Large-Scale Transcriptome Perturbations and Motif Inference

Delphine Potier,[1] Kristofer Davie,[1] Gert Hulselmans,[1] Marina Naval Sanchez,[1] Lotte Haagen,[1] Vân Anh Huynh-Thu,[2,4] Duygu Koldere,[3] Arzu Celik,[3] Pierre Geurts,[2] Valerie Christiaens,[1] and Stein Aerts[1,*]

[1]Laboratory of Computational Biology, Center for Human Genetics, University of Leuven, Leuven 3000, Belgium
[2]Department of Electrical Engineering and Computer Science and GIGA-R, University of Liège, Liège 4000, Belgium
[3]Department of Molecular Biology and Genetics, Bogazici University, Istanbul 34342, Turkey
[4]Present address: School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK
*Correspondence: stein.aerts@med.kuleuven.be
http://dx.doi.org/10.1016/j.celrep.2014.11.038

SUMMARY

Genome control is operated by transcription factors (TFs) controlling their target genes by binding to promoters and enhancers. Conceptually, the interactions between TFs, their binding sites, and their functional targets are represented by gene regulatory networks (GRNs). Deciphering in vivo GRNs underlying organ development in an unbiased genome-wide setting involves identifying both functional TF-gene interactions and physical TF-DNA interactions. To reverse engineer the GRNs of eye development in *Drosophila*, we performed RNA-seq across 72 genetic perturbations and sorted cell types and inferred a coexpression network. Next, we derived direct TF-DNA interactions using computational motif inference, ultimately connecting 241 TFs to 5,632 direct target genes through 24,926 enhancers. Using this network, we found network motifs, *cis*-regulatory codes, and regulators of eye development. We validate the predicted target regions of Grainyhead by ChIP-seq and identify this factor as a general cofactor in the eye network, being bound to thousands of nucleosome-free regions.

INTRODUCTION

Gene regulatory networks (GRNs) control the functional expression of the genome. A GRN represents a particular combination of active transcription factors (TFs) that interacts with genomic enhancers, producing a specific gene expression profile for each cell type. Reconstructing GRNs and deciphering the underlying *cis*-regulatory logic at a genomic scale can provide mechanistic insight into developmental and disease processes (Davidson et al., 2002). Approaches to reverse engineer GRNs either concentrate on unraveling which TF physically binds to which genomic location, representing *TF-DNA* interactions, or identifying which TF functionally regulates the expression of

which target gene, representing *TF-gene* interactions. Ideally, network inference takes both *TF-DNA* and *TF-gene* relationships into account.

TF-DNA networks have been reconstructed using ChIP-seq in yeast (Harbison et al., 2004), in *Drosophila* and *C. elegans* cell cultures and embryos (Roy et al., 2010), and in human cell culture (Gerstein et al., 2012; Yan et al., 2013). Contrary to yeast, in higher eukaryotes, these networks did not reach whole-genome coverage due to technical limitations of ChIP-seq. Nevertheless, these studies have led to important insight into GRN topology, including a substantial role for feedforward and feedback loops. *TF-gene* networks, on the other hand, are often reconstructed using gene expression profiling, whereby computational methods are applied to reverse engineer a GRN based on coexpression of a TF with its candidate (direct or indirect) target genes (Marbach et al., 2012). Coexpression networks have been inferred from a variety of samples, including sorted cell types (Novershtern et al., 2011), cohorts of patient samples or cell cultures (Bonnet et al., 2010), or series of TF or chemical perturbations (Basso et al., 2005; De Cegli et al., 2013).

Ideally, network inference takes both physical *TF-DNA* and functional *TF-gene* relationships into account. By overlaying the indirect *TF-gene* network inferred from gene expression profiling with the direct *TF-DNA* network obtained by ChIP-chip, functional GRNs have been reconstructed in yeast (Kemmeren et al., 2014). Examples of integrative approaches in higher eukaryotes are the networks for Th17 differentiation, based on RNA-seq, ChIP-seq, and TF motif scanning (Ciofani et al., 2012; Yosef et al., 2013), a smaller GRN of muscle cell differentiation based on microarrays and motif detection (Warner et al., 2008), and GRNs underlying hematopoietic differentiation (Calero-Nieto et al., 2014; Novershtern et al., 2011). As these networks are obtained from in vitro cell cultures or through flow-cytometry-based cell sorting, they offer limited insight into how GRNs behave in vivo. Several small-scale efforts have been published that reconstruct in vivo networks "top down," usually focusing on a subset of prechosen master regulators. Examples include vulva development in *C. elegans* (Deplancke et al., 2006) (117 TFs and 72 gene
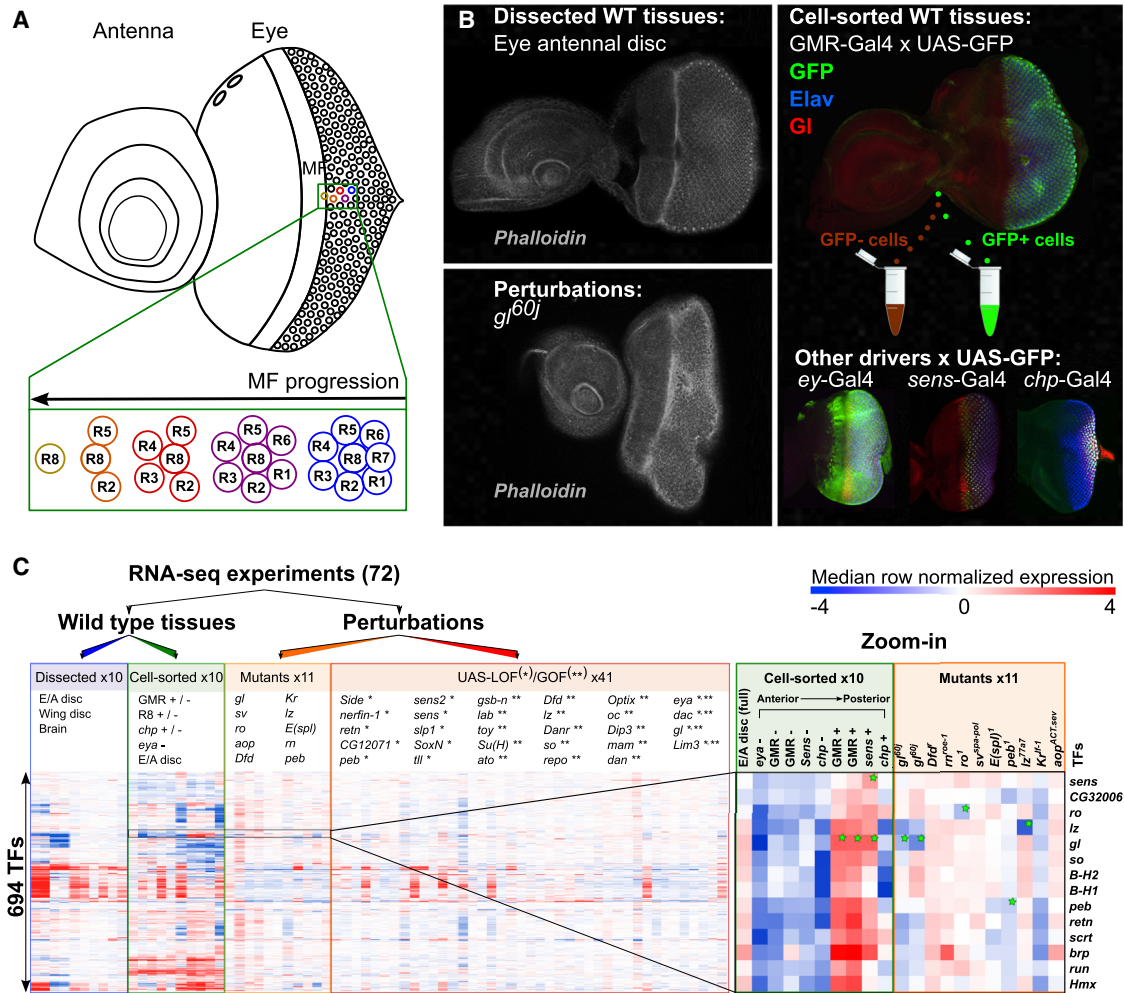
Figure 1. RNA-Seq on Transcriptome Perturbations

(A) Schema representing the third instar larval eye-antennal imaginal disc (posterior on the right and anterior on the left) showing cell types and temporal diversity.

(B) Examples of tissues and perturbations used: the left shows WT (top) and gl[60j] (bottom) discs stained with phalloidin. The right shows the different drivers used for cell sorting: GMR>GFP for PR ey>GFP for the eye part of the disc, sens-GPF for PR8 cells, and chp-GFP for the most posterior PRs.

(C) Heat map representing the expression profiles (median row normalized) of expressed TFs across the 72 RNA-seq experiments; a zoom is shown on cell sorting and mutant experiments for PR-specific TFs. The green stars points to expected variation. For example, glass is upregulated all PR-related samples and down in the glass mutant. Senseless is upregulated in GMR>GFP-positive cells, but more strongly in the sens-GFP-positive cells and in the UAS-Ato perturbation. Lozenge is enriched in PRs and downregulated in the glass mutant because it is a direct target gene of Glass (Naval-Sánchez et al., 2013; Yan et al., 2003). See also Table S1 for the genotypes and Table S2 for GMR>GFP-positive versus GMR>GFP-negative DESeq differential expression analysis.

promoters), heart development in *Drosophila* (Junion et al., 2012), anterior-posterior and dorsal-ventral patterning in the *Drosophila* (MacArthur et al., 2009), and mesoderm development in *Drosophila* (Sandmann et al., 2007). However, the bottom-up reconstruction of GRNs underlying the development of an entire tissue or organ is still a considerable challenge. In the present study, we investigate whether integrative GRN mapping approaches, combining both functional TF-gene and (predicted) physical TF-DNA interactions, can be used to unravel in vivo networks underlying complex developmental programs such as organ formation, in a genome-wide unbiased setting, using *Drosophila* eye development as a model system.

RESULTS

Transcriptome Changes during Eye Development Induced by TF Perturbations and Cell Sorting

The *Drosophila* larval eye-antennal imaginal disc is composed of several cell types that give rise to the antenna, the eye, the maxillary palpus, and the head cuticle. During the third instar larval stage, this tissue represents a very interesting biological system, as it contains various cell types at different temporal stages. In the eye part, these include undifferentiated pluripotent cells anterior to the differentiation wavefront (called the morphogenetic furrow [MF]) and retinal cells undergoing differentiation posterior to the MF (Figure 1A). We asked whether the

GRNs underlying this complex developmental program could be reconstructed. To this end, we first measured gene expression during *Drosophila* eye development using RNA-seq starting from 73 different samples of eye-antennal imaginal discs (Figures 1B and 1C). The 72 samples that passed quality control (QC) filters (see the Supplemental Experimental Procedures) include 10 tissue samples from WT *D. melanogaster* strains, 52 genetic perturbations of different TFs, and 10 samples representing specific cell types obtained by FACS-based cell sorting (Table S1). The genetic perturbations comprise 10 TF knockout conditions that result in rough eye (e.g., *gl, shaven (sv), rough (ro), anterior open (aop), lozenge (lz), enhancer of split m8 (E(spl)), pebbled (peb), rn (rotund), Kr (Kruppel), Dfd (Deformed)*) plus mild perturbations of 30 TFs that overall lack a clear adult eye phenotype. These mild perturbations were obtained using the Gal4-upstream activating sequence (UAS) system, whereby the UAS allowed for eye-specific expression of RNA interference (RNAi) constructs to knock down a TF or eye-specific overexpression of TF complementary DNA (cDNA) (Figure 1C; Table S1; Experimental Procedures). The samples used for gene expression profiling of specific cell types were obtained by fluorescence-activated cell sorting (FACS) using genetic markers that represent different subsets of cell types, namely *eyes absent*-Gal4, UAS-GFP (*eya*>GFP) for proliferating cells anterior and differentiating cells posterior to the MF, GMR-Gal4, UAS-GFP (GMR>GFP) (Glass Multimer Reporter) for differentiating photoreceptor (PR) cells posterior to the MF, *senseless*-GFP (*sens*-GFP) for the R8 PR, and *chaoptin*-GFP (*chp*-GFP) for the most posterior rows of PRs (Figure 1B; Table S1; Experimental Procedures). For each driver, we obtained RNA from both GFP-positive and GFP-negative cells. Each sample was subjected to RNA-seq and processed to obtain normalized gene expression levels (see the Experimental Procedures). Of the 15,216 annotated protein-coding genes, 10,380 are expressed in at least 5 of the 72 samples. Remarkably, of the 752 annotated TFs (Pfreundt et al., 2010), 694 are expressed (Figure 1C). To verify the quality of the perturbations and cell sorting experiments, we looked in detail at several marker genes that have specific roles during eye development (Figure 1C, "zoom-in"). Next, using genome-wide gene rankings based on the contrast of GMR-positive cells (representing differentiating PRs) versus the GMR-negative cells (representing undifferentiated cells), we found a very strong enrichment of the expected biological processes, such as "compound eye PR development" (adjusted p value 1.11×10^{-8} in the top 551 genes) (Table S2; Supplemental Experimental Procedures). Likewise, a contrast defined by more specific R8-positive or chp-positive cells versus all GMR-positive cells yields "R8 PR differentiation" (8.36×10^{-3}) and "pigmentation" (3.5×10^{-3}) as enriched terms, respectively. Although the comparison of one perturbation versus control or versus another perturbation can yield interesting results, we decided not to consider each sample separately but instead analyze the entire data set using network inference methods. The advantage of doing so is that the combination of multiple perturbations can provide a higher confidence for deriving coexpression by using all measurements simultaneously.
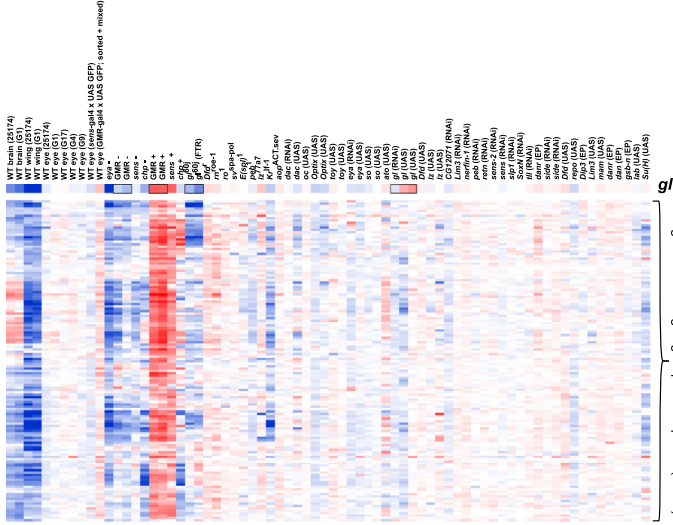
TF-Centric Inference of the Eye Network and Motif-Based Edge Pruning

To reconstruct the GRNs underlying cell proliferation, cell specification, and differentiation in the different developmental programs that occur simultaneously in the eye-antennal disc, we have taken a TF-centric approach. Particularly, for each expressed TF, we aim to identify its candidate target genes based on coexpression relationships (i.e., TF-gene relationships), followed by *cis*-regulatory sequence analysis (i.e., TF-DNA relationships). We first examine Glass as an example to illustrate a validated data analysis pipeline to be applied later to less characterized TFs. Glass is a zinc finger TF involved in PR differentiation (Moses et al., 1989a) for which several target genes and binding sites have already been discovered (Naval-Sánchez et al., 2013; Yan et al., 2003). We use two different methods to measure coexpression on our large expression matrix, one called Pavlidis Template Matching (PTM; see Figure 2A for an example), based on Pearson correlation, and the second called GENIE3, based on Random Forest machine learning (see the Experimental Procedures). Both the PTM and GENIE3 queries yield gene sets that represent candidate Glass targets, being a mixture of direct and indirect targets. In the second step, we test whether the Glass motif is enriched in these gene sets compared with the rest of the genome and compared with thousands of other "background" TF motifs. To this end, we use the method *i-cisTarget*, as it uses large collections of candidate motifs, applies motif clustering within *cis*-regulatory modules (CRMs), and integrates motif cluster scores across species (Herrmann et al., 2012). When i-cisTarget is applied to sets of genes coexpressed with *glass*, the Glass motif is found as the most significantly enriched motif. Particularly, the Glass motif is ranked first out of 6,782 PWMs with Normalized Enrichment Scores (NESs) ranging between 4.92 and 7.42, depending on which method (PTM or GENIE3) or threshold is used to define the coexpressed gene set (Figures 2B and 2C). As the Glass motif is only enriched in the set of positively correlated genes, and not in the set of negatively correlated genes, this indicates that Glass is likely an activator rather than a repressor. We also found that the fraction of predicted direct targets increases when the input set is smaller and more tightly coexpressed (Figure 2C). However, this stringency comes with a cost of lower sensitivity, and therefore, we maintain various stringency levels throughout further analyses.
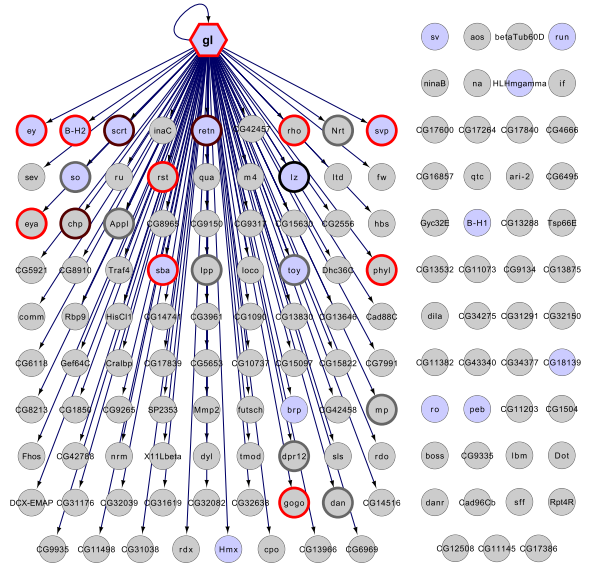
Once the subset of inferred direct targets is defined, each of them contains one or more high-scoring enhancers with one or more Glass binding sites per enhancer. To test the predictions of Glass target enhancers, we selected predicted regions that overlap with transgenic Gal4 enhancer-reporter lines from the FlyLight database (Jenett et al., 2012). We crossed these lines with UAS-GFP, followed by immunohistochemistry and confocal microscopy to monitor GFP expression in the eye. We tested 34 Janelia regions (accounting for 36 predicted enhancers, for 26 genes; see Table S3 for more details) of which 20 (22 predicted enhancers for 16 genes) were positive, being active posterior to the MF, either in PR (16), cone cells (3), or ocelli (1), thereby likely to be activated by Glass (Figure 2D).

Interestingly, the motif enrichment analysis identifies, besides the Glass motif, potential cofactors and their targets among the input set of *glass*-coexpressed genes. For example, Lozenge

A   Co-expression: PTM & GENIE3
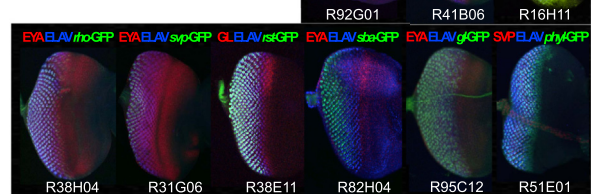
D   Gl predicted targets network

Known from previous studies:
H. Yan et al. 2003   M. Naval et al. 2012

Current study (Janelia Gal4 lines):
Expression in PRs ocelli or cone cells   No expression in PRs

B   Co-regulation: iRegulon

PWM rank (cluster rank)	Associated TF	NES	Number of targets by PWM (cluster)	Logo / PWM name	Confidence
1 (1)	Gl	7.42	75 (89)	flyfactorsurvey-gl_SOLEXA_5	(1) motif of the query TF *High confidence layer*
10 (2)	Pnt	3.93	66 (106)	jaspar-MA0062.1	(3) motif of another TF, not present in the co-expression cluster *Low confidence layer*
37 (9)	Lz	3.04	52 (88)	jaspar-MA0002.2	(2) motif of another TF, present in the co-expression cluster *Medium confidence layer*

C

Percentage of genes predicted as direct Gl targets across methods & thresholds

	GENIE3 0.005	GENIE3 0.005	GENIE3 0.005	GENIE3 0.005	PTM 2sided 10e-7	PTM 2sided 10e-8	PTM 2sided 10e-9	PTM 2sided 10e-10	PTM greater 10e-7	PTM greater 10e-8	PTM greater 10e-9	PTM greater 10e-10	PTM less 10e-7	PTM less 10e-8	PTM less 10e-9	PTM less 10e-10
Number of input genes	280	202	142	110	193	144	106	82	178	137	102	82	37	21	14	4
Number of predicted direct targets	112	85	66	57	105	84	64	58	111	90	73	60	0	0	0	0
Percentage of predicted direct targets	40%	42.1%	46.5%	51.8%	54.4%	58.3%	60.4%	70.7%	62.4%	65.7%	71.6%	73.2%	0%	0%	0%	0%

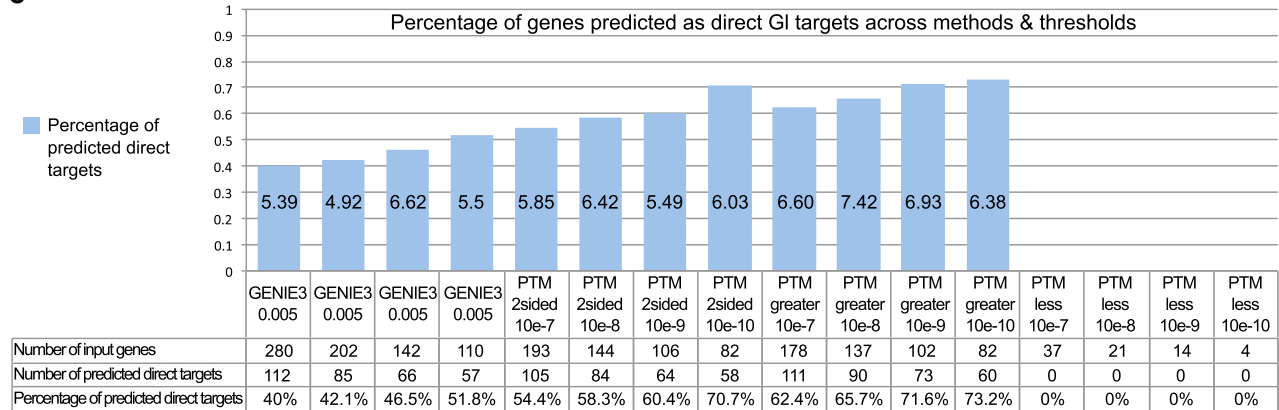Values above bars: 5.39, 4.92, 6.62, 5.5, 5.85, 6.42, 5.49, 6.03, 6.60, 7.42, 6.93, 6.38

Figure 2. Mapping a TF Targetome: The Glass Example
(A) Heatmap of the 137 genes coexpressed with *glass* (derived by correlation, PTM p value < 10 × 10⁻⁸).
(B) i-cisTarget results on the gene set shown in (A), indicating how different confidence layers are built.
(C) Different methods and thresholds for coexpression all identify the Glass motif enriched, but influence the fraction of predicted direct targets among the coexpressed gene set.
(D) Predicted Gl targetome. Gl predicted target regions fully overlapping with a FlyLight Gal4 enhancer-reporters have been tested: a red node border represents a gene with a tested region showing expression corresponding to a possible Glass activation (PRs, ocelli, or cone cells). Images for those enhancer-reporters showing activity in the PRs are shown (enhancer-GFP in green; Elav in blue; Eya, Gl, or Pros in red). A gray node border represents a gene with a tested region not showing an expression pattern corresponding to a possible Glass activation. Black and brown node borders represent Glass targets validated in other studies (see also Table S3 for Flylight enhance-reporter references).
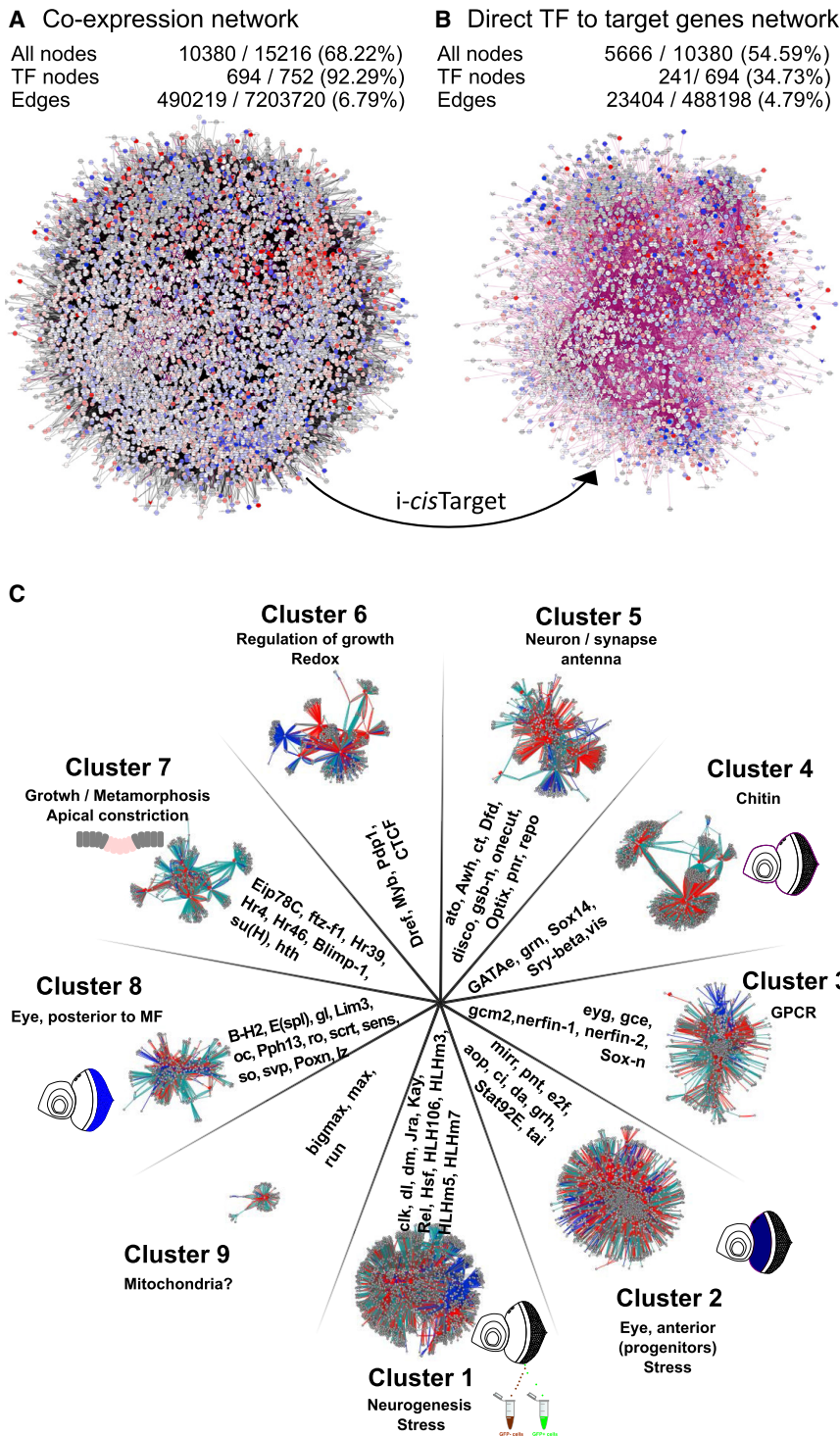
A Co-expression network

All nodes	10380 / 15216 (68.22%)
TF nodes	694 / 752 (92.29%)
Edges	490219 / 7203720 (6.79%)

B Direct TF to target genes network

All nodes	5666 / 10380 (54.59%)
TF nodes	241/ 694 (34.73%)
Edges	23404 / 488198 (4.79%)

i-*cis*Target

Figure 3. The Mapped Eye GRN

(A) Coexpression TF-gene network and network statistics.

(B) Direct GRN, combining TF-gene and TF-DNA information. TF-DNA relationships are derived by i-cisTarget motif inference.

(C) Clust&see network clustering. The FT algorithm clustered the direct GRN in nine distinct clusters all enriched for specific GO terms (relevant terms are listed). A subset of biologically relevant TFs is listed for each cluster (see also Figure S1 for a comparison of expression level and SD of the genes present in the network compared with the rest genome and the network connectivity, Table S4 for network clusters GO term enrichment, and Figure S6 for insights related to temporal aspects).

C

Cluster 6
Regulation of growth
Redox

Cluster 5
Neuron / synapse
antenna

Cluster 7
Grotwh / Metamorphosis
Apical constriction

Cluster 4
Chitin

Cluster 8
Eye, posterior to MF

Cluster 3
GPCR

Cluster 9
Mitochondria?

Cluster 2
Eye, anterior
(progenitors)
Stress

Cluster 1
Neurogenesis
Stress

Drei, Myb, PDP1, CTCF

ato, Awh, ct, Dfd, disco, gsb-n, onecut, Optix, pnt, repo

Eip78C, ftz-f1, Hr39, Hr4, Hr46, Blimp-1, su(H), hth

GATAe, grn, Sox14, Sry-beta, vis

gcm2, nerfin-1, nerfin-2, Sox-n

eyg, gce, eyg, gce, nerfin-2, Sox-n

B-H2, E(spl), gl, Lim3, oc, Pph13, ro, scrt, sens, so, svp, Poxn, lz

mirr, pnt, e2f, aop, ci, da, grh, Stat92E, tai

bigmax, max, run

clk, dl, dm, Jra, Kay, Rel, Hsf, HLH106, HLHm3, HLHm5, HLHm7

that it is feasible to identify high-confidence enhancers and direct target genes of TFs by combining TF-coexpression with TF-motif detection.

Discovery of Coexpressed Genes and Prediction of Direct Target Genes for 241 TFs in the Eye Network

Encouraged by the results for Glass, we now ask whether direct target genes can be identified for all TFs involved in the eye developmental program. To this end, we applied the same prediction pipeline to each of the 694 expressed TFs. The merged coexpression network for all TFs is composed of 10,380 nodes and 488,198 edges (Figure 3A). Applying i-cis-Target on each TF-centered coexpression module identified the corresponding TF motif as significantly enriched, along with an optimal subset of inferred direct targets for 241 TFs. These 241 TFs and their predicted direct targets, representing the high-confidence network, account for 23,404 regulatory interactions, being 5.16% of the total number of interactions in the coexpression network (Figure 3B; http://eyenetwork.aertslab.org/). Of the 10,380 expressed genes, 5,666 are part of this GRN, meaning that they have at least one upstream regulator or one downstream target gene. This set of connected network members shows overall a higher expression level and a higher vari-

and Pointed are two TFs with known roles in PR differentiation (Figure 2B). Predicted targets of these "secondary" factors (i.e., their motifs are enriched in coexpressed genes of another factor) are added to the eye network as "medium confidence" interactions when they are part of the input gene set like in the Lozenge case. With this Glass example, we have illustrated

ance across the 72 samples than the nonmembers (Figure S1A). Functionally, the connected genes are enriched for developmental processes, including sensory organ development (FDR 4.09×10^{-35}; Table S4), and they are very strongly enriched for TF activity (FDR 9.31×10^{-62}). The number of target genes per TF in this network ranges from 1 to 677 targets (median = 38;

mean = 97.11) and has a scale-free distribution (Figure S1B). Overall, compared with the limited knowledge in the literature of only a handful of known eye and antennal regulators, it is remarkable that we find more than 200 TFs (even up to 335 TFs in the medium-confidence network) for which a significant number of predicted direct targets are significantly coexpressed with the TF. This large network now allows more in-depth analyses of genomic regulatory programs of organ development.

Multiple Interconnected Subnetworks Represent Concurrent Developmental Programs and Biological Processes

Although the motif-inferred direct GRN represents only about 5% of the "parental" coexpression network, its size and complexity still limit its direct use as an organizer or as a code that may clearly describe the developmental program of eye development. We reasoned that this limitation is partly due to the complexity of the tissue under study, as in the eye-antennal disc multiple developmental subprograms are being deployed simultaneously to ultimately generate the eye, the antennae, the ocelli, and part of the head (Figure 1A) (Quan et al., 2012). If each subprogram is operated by distinct TFs, then the large GRN may be dissected into multiple subnetworks, each representing a distinct program. Alternatively, subprograms could exist as network modules fulfilling a particular patterning role, and these may be reused by the different sensory organs. To examine whether any of these models match with our network, we clustered the GRN with the Clust&See method (Spinelli et al., 2013) (see the Supplemental Experimental Procedures). This resulted in nine subnetworks, with sizes ranging between 69 and 1,499 genes. We used Gene Ontology enrichment analyses to analyze the putative processes controlled by each subnetwork and found several clusters where specific processes of eye development are most represented (Figure 3C; Table S4). Particularly, cluster eight contains several PR-related TFs (e.g., Glass, Rough, Senseless, Scratch), and this cluster is enriched for the GO term "compound eye PR development" (GO:0042051, p value 6×10^{-8}). Cluster 2 contains interactions likely to occur mainly anterior to the MF, being enriched for the GO term "regulation of growth" (GO:0040008, p value 3×10^{-17}). However, this cluster is also enriched for "PR cell development" (GO:0042461, p value 4×10^{-9}), indicating that these different processes are strongly interconnected at the network level. Five of the remaining six clusters do not represent particular anatomical structures such as eye, antennae, ocelli, or head, but rather depict functional network modules that can be invoked in several places in the tissue. These modules control neurogenesis, metamorphosis/growth, redox, synapse, and G protein-coupled receptor (GPCR) signaling. Clusters 1 and 2 are enriched for genes involved in stress response and are mainly activated in all the FACS sorted samples. Thus, the network clustering resulted in meaningful subclusters representing either particular parts of the tissue or reused functional modules.

The Regulatory Network Reveals Activators, Repressors, Feedforward Loops, and *cis*-Regulatory Grammars of Motif Combinations

Next, we asked whether we could decipher which TFs are activators and which are repressors within the predicted GRN. In-

formation about the mode of action can be derived from the sign of the correlation of the TF with the predicted target genes (see Experimental Procedures). We could obtain this information for 15,611 of 23,404 edges. For example, of 112 predicted Glass target genes, all are activating, while 49 of the 53 predicted Escargot (Esg) targets are repressed, as expected because Esg is a known repressor (Fuse et al., 1994), although in the eye-antennal disc its repressive role has not been described. In the entire predicted network, the majority of the direct edges represent activation (13,634), whereas fewer represent repression (1,977), and 7,793 edges are of an undefined direct TF to TG link (see the Experimental Procedures). At the TF level, among the 241 regulators in the network, 89 are predicted to only activate their targets, 65 to only repress their targets, and 62 are predicted to act as activator and repressor (Figure 4A). To assess the quality of these predictions, we performed a GO term-enrichment analysis with GOrilla (Eden et al., 2009) with all TFs as background and either activators or repressors as input. The set of predicted activators is enriched for "positive regulation of transcription, DNA templated" (q value FDR = 4.01×10^{-4}), and the predicted repressors are involved in "negative regulation of transcription, DNA templated" (q value FDR = 1.01×10^{-2}).

The network is also enriched for network motifs such as feedforward loops (FFLs) and double-feedback loops (Milo et al., 2002). For example, we recovered the known Sens-Ro repressive feedback loop (Frankfort et al., 2001; Pepple et al., 2008) in the subnetwork of PR differentiation (Figure 4B), among a total of 22,741 FFLs (Z score = 60). We also found that a very large proportion of TFs is autoregulatory (108 of 241, 44.8%). When two TFs in a FFL share their target genes through the same regulatory region, their binding sites are expected to show significant co-occurrence. This is indeed the case for 1,003 unique TF pairs (p value for motif co-occurrence < 0.001 and more than 10 regions in common). We found a striking overlap between Glass and Lozenge binding sites, both at the gene (Figure 5A) and enhancer level (enrichment 4.85; p value = 2.18×10^{-16}; Figure 5B). These two TFs also form a FFL because Glass targets *lz*, Lz targets *gl*, and they together coregulate 36 targets (Figure 5A; Table S5: enrichment 29×; p value 1.14×10^{-52}). Interestingly, at the enhancer level, the predicted Glass and Lozenge binding sites are largely overlapping, which was an unexpected finding given that their position weight matrices are rather different (STAMP e-value > 0.1; Figures 5C and 5C'; see the Supplemental Experimental Procedures).

Another interesting TF combination in the PR differentiation subnetwork is the predicted Glass and Escargot coregulation (Figure 4B). A large part of genes activated by Glass in PRs (including *gl* itself) is predicted to be repressed by Esg outside of the eye region. As *esg* is more expressed anterior to the MF (e.g., *eya>GFP* negative cells) than posterior to the MF (e.g., *eya>GFP* positive cells), we can hypothesize that Escargot represses Glass targets and likely other genes involved in PR differentiation, anterior to the MF and/or in the antenna. These findings at the *cis*-regulatory level provide a connection with the network motifs and show that multiple pair of coregulators not only share common targets, but often operate via the same enhancer.
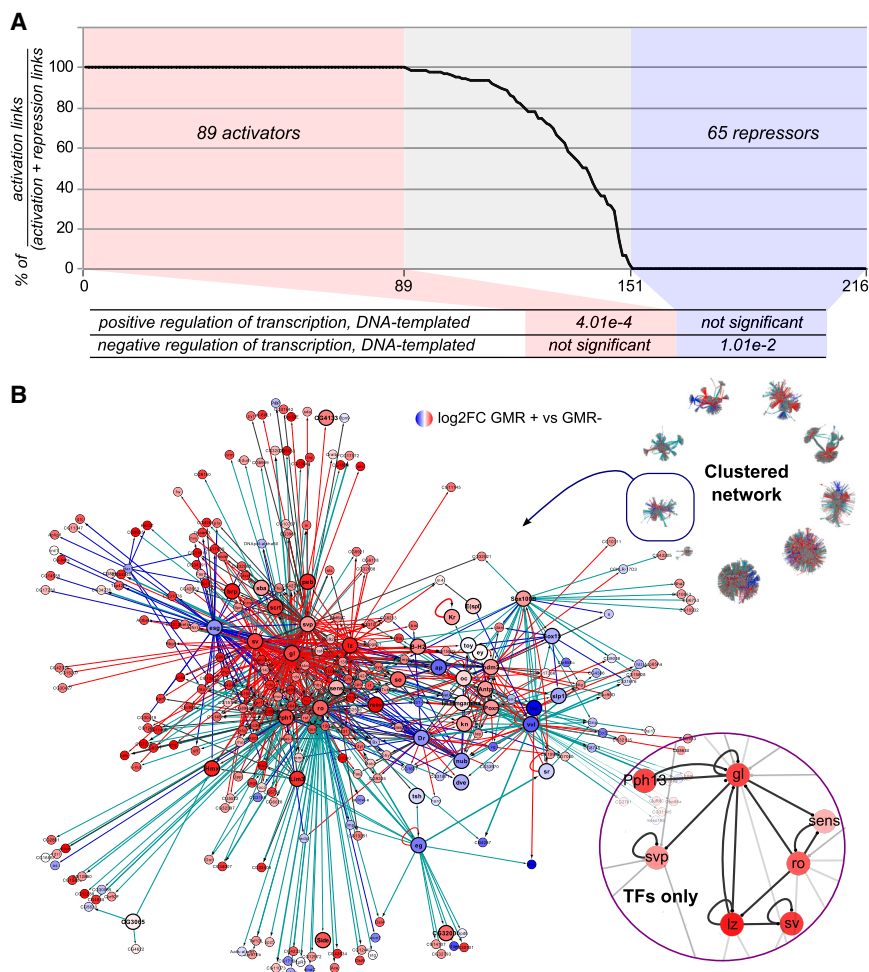
Figure 4. TF Mode of Action and Network Motifs

(A) Directionality (activation/repression) of TF-target relationships in the high-confidence network. TFs on the x axis are ranked from the 100% activators to 100% repressors. The fraction of activation edges is reported on the y axis for each TF.

(B) Activation and repression in the PR differentiation subnetwork (subcluster 8 from Figure 3). The node color represents the log2FC in PRs compared with the rest of the EA disc (GMR>GFP positive versus GMR>GFP negative). The bigger node size represents TFs. The edge color represents a predicted activation (red), repression (blue), or an undetermined action (turquoise). The edge thickness represents the number of methods that predict the TF-target edge. The inset shows that we identified a known feedback loop between *ro* and *sens* and a known activation of *lz* by Glass.

Casares, 2010; Kumar, 2009) (Figure S2). Another TF of the RDGN for which we found many target genes in the medium-confidence network is So. Particularly, we predicted 417 targets for the Eya/So pair (Figure S2A). We further validated the Ey and Eya/So target predictions and found their targets significantly enriched in the *Eya>GFP*-positive sorted cells (Figures S2B and S2C; Table S6). Moreover, some of the enhancers predicted to be bound both by So/Eya and Ey around *ey* and *so* reproduce the expected expression pattern of those genes, as indicated by enhancer-reporter assays (Figures S2D–S2F).

Among the 164 TFs in the high-confidence GRN, having more than 20 predicted target genes, the majority has not been investigated specifically for their role in eye development. One of the best scoring TFs in this analysis is Grainyhead, with only one eye-related publication, but a strong enrichment of "eye development" among its predicted targets (adjusted p value 8 × 10^{-24}; Figure 6A). Interestingly, Grainyhead has the second highest number of target genes (657) and is ubiquitously expressed in the eye disc (Yao et al., 2006). As indicated by the keyword "synapse," one of the TFs with strong GO enrichment is Onecut (Figure 6B), the human HNF6 homolog. The targets of Onecut could be discovered thanks to the motif of the human homolog and are significantly enriched for synapse related processes (e.g., "synaptic transmission" adjusted p value, 3 ×10^{-09}) (Figures S3A and S3B). This confirms previous studies (Janky et al., 2014; Weirauch et al., 2014) that PWMs derived for one species can be transferred to other species, provided that the DNA-binding domain is conserved, which is the case for Onecut and HNF6 (Figures S3B–S3D). In this case, the human HNF6 PWM outperforms the *Drosophila* Onecut PWM, which may be due to a difference in quality, particularly because the human PWM is
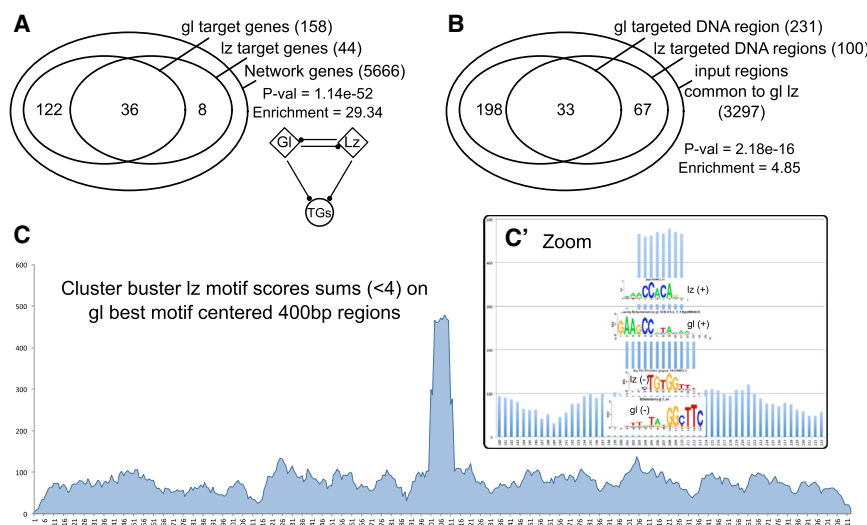
Identification of Known and New Regulators in the Eye-Antennal Developmental Program

To validate known TFs and to identify TFs with a possibly important, but until today uncharacterized role in eye development, we evaluated for each TF how many publications linked to that TF (as curated by FlyBase; St Pierre et al., 2014) contain the word "eye" in their title. We then compared this "eye knowledge value" with the enrichment of "eye"-related Gene Ontology terms in the set of predicted targets for the same TF (Figure 6A). Reassuringly, Eyeless and Sine Oculis are found in the top right of this plot, alongside other well-known TFs like Pointed (Pnt), Aop, Gl, Lz, and Ro. This indicates a high eye-related knowledge for these TFs, together with a strong enrichment for eye-related processes in their targetomes. The most famous master regulator of the classical retinal determination gene network is the PAX6 homolog Eyeless (Ey). As the expression pattern of *ey* is different from the pattern of its known targets (e.g., *so*) (Bessa et al., 2002a), we were not able to predict any of the Eyeless target genes within the high-confidence layer because this layer is restricted to targets that are tightly (anti-)correlated with their regulator, but when inspecting the medium confidence layer, we found 446 predicted Ey targets, including the known targets *eyes absent (eya)*, *so*, *teashirt (tsh)*, and *atonal (ato)* (Amore and

above input. The most strongly enriched motif in these peaks is the Grh motif with a high-significance score (i-cisTarget NES = 10.77), confirming the high quality and specificity of the ChIP-seq data.

The ChIP-seq data show a highly significant confirmation of our predicted Grh binding sites (FDR = 0.000, GSEA) (Figures 7A and 7B). By comparing Grh ChIP-seq peaks with open chromatin (i.e., FAIRE-seq peaks), we confirmed that Grh binds to a very large fraction of active regulatory regions (57.1% of all FAIRE-seq peaks contain a Grh peak; Figure 7C). Note that the remaining Grh peaks, those that do not overlap with open chromatin, are likely also true positive binding sites because this fraction of peaks still presents a very strong enrichment of the Grh motif (Figure S4). Two genes are shown in Figures 7D–7F as an example, namely *Hr39* and *ci*, both having multiple regulatory regions with a Grh ChIP-seq peak and a FAIRE-seq peak. The regions with the highest scoring Grh binding site prediction overlap with Flylight enhancers that drive reporter expression in the eye-antennal disc (Figures 7D and 7E) (Jory et al., 2012). In conclusion, ChIP-seq against Grainyhead validates the predicted Grainyhead binding sites and target genes and confirms its correlation with open chromatin.

constructed from curated binding sites, while the *Drosophila* PWM is derived by Bacterial-1-Hybrid (Noyes et al., 2008). We furthermore validated these targets experimentally by knocking down *onecut* in the eye disc (see Table S1; Experimental Procedures and Supplemental Experimental Procedures) and found the set of predicted targets significantly affected, as quantified by RNA-seq and GSEA (Figure S3E; Table S7).

Open Chromatin Underlies the Targetome of Ubiquitous TFs, such as Grainyhead

The TF-DNA network contains 24,926 candidate CRMs that are associated to 5,632 genes (i.e., on average 4.4 CRMs per gene) and covering 28.45 kbp (almost 20%) of the noncoding *Drosophila melanogaster* genome. We wondered if this predicted "eye regulatory landscape" is indeed active or nucleosome free at the chromatin level compared with locations not involved in the eye network. To test this, we used recently generated formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq data in the same system (i.e., the eye-antennal disc) at exactly the same stage as we examined here (McKay and Lieb, 2013). To test which TF-DNA interactions have the highest impact on chromatin activity, we tested all correlations between eye-specific FAIRE peaks and specific TF-target CRMs (Table S8) and found Grainyhead as the TF with the strongest enrichment of open chromatin (p value = 2.89 × 10^{-113}). Furthermore, chromatin opening correlates strongly with the precise location of Grh predicted binding sites (Figure 6C).

We decided to experimentally validate the Grh target predictions and the putatively widespread association with open chromatin using a transgenic *Drosophila* line carrying a GFP-tagged *grh* gene, expressing a chimeric Grh-GFP fusion under endogenous *grh* control (Spokony and White, 2012). This fusion protein is ubiquitously expressed in the eye-antennal disc (Figure 6D), confirming previous findings (Yao et al., 2006). We performed ChIP-seq against using an anti-GFP antibody (see the Experimental Procedures). Peak calling identified 10,407 Grh peaks (MACS2 FDR < 0.05), of which 5,874 are at least 2-fold enriched

DISCUSSION

The development of the *Drosophila* eye is a classical model system to study neuronal differentiation and patterning. The TFs that represent the core of the retinal determination network are Eyeless (Ey), Twin of Eyeless (Toy), Dachsund (Dac), Sine Oculis (So), and Eyes Absent (Eya). Although many regulatory interactions are known between these TFs, as they intensively cross-regulate each other (Bessa et al., 2002a), knowledge about interactions with downstream target genes and of other TFs involved in the eye-antennal GRN is sparse (Amore and Casares, 2010). Here, we aimed at combining classical reverse genetics—starting from a mutant allele and analyze its (molecular) phenotype—with genomics. Doing so, we aimed at unveiling genetic regulatory interactions in an unbiased way, and we identified many regulators of the eye and antennal developmental programs, for most of which we did not require or use any mutation or direct perturbation.

We began our mapping approach by systematically perturbing the developmental system. We attempted to include multiple
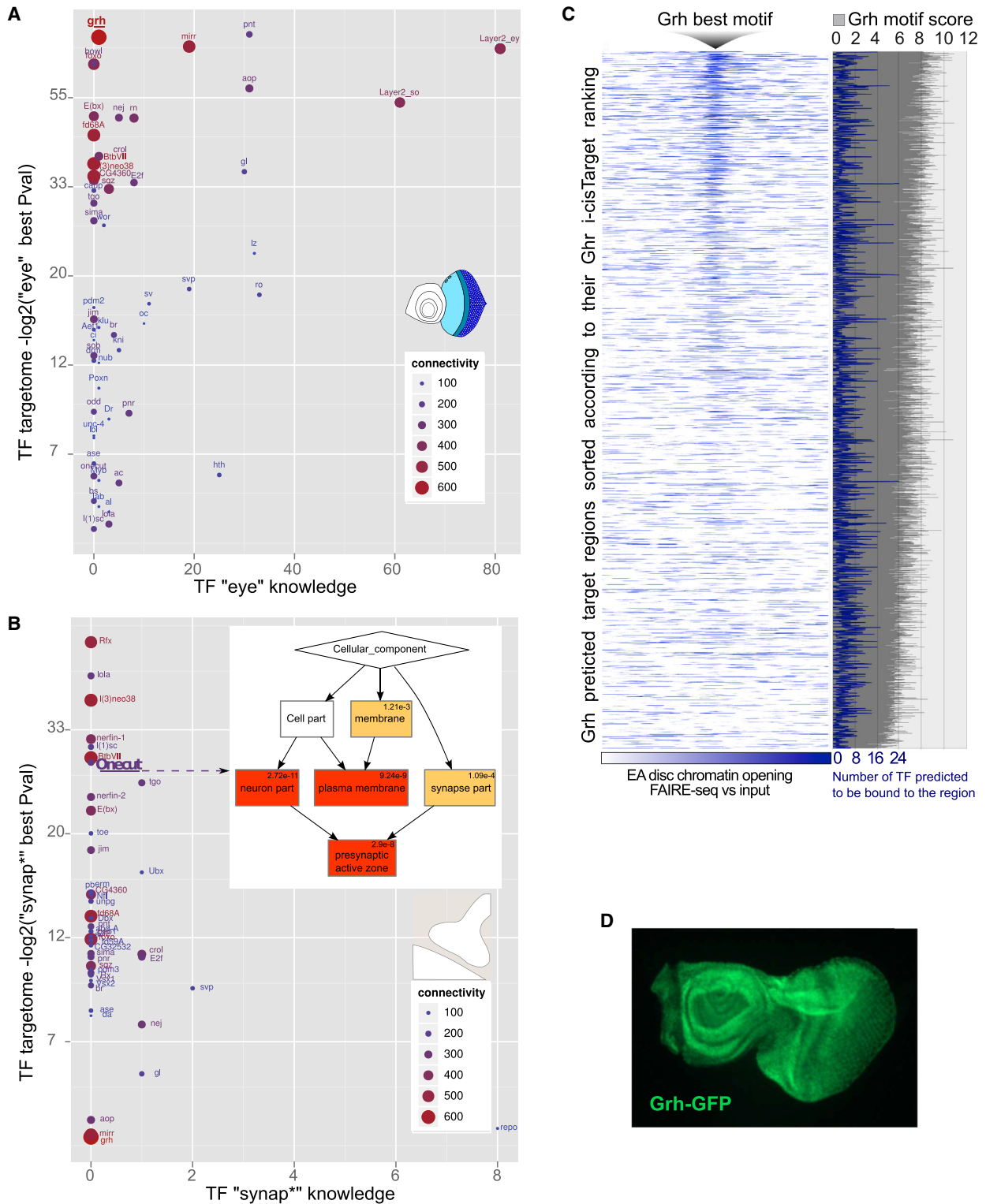
A

grh

pnt

Layer2_ey

bowl
foxo

mirr

55

aop

Layer2_so

E(bx)   nej   rn

fd68A

crol
BtbVII
l(3)neo38
CG4360   E2f
sgz
caup
tgo
sima
wor

gl

33

lz

20

svp

ro

pdm2
jim
Aef   klu   br
c   kni
scb
apm
nub

sv
oc

connectivity
• 100
• 200
• 300
• 400
● 500
● 600

Poxn

odd
Dr   pnr
unc-4
lbl

7

ase

one
cut
ac
myb
bs
hb   al
lola
l(1)sc

hth

TF targetome -log2("eye" best Pval)

0   20   40   60   80

TF "eye" knowledge

B

Rfx

lola

l(3)neo38

Cellular_component

33

nerfin-1
l(1)sc

Cell part

1.21e-3
membrane

BtbVII
Onecut

tgo

nerfin-2

2.72e-11
neuron part

9.24e-9
plasma membrane

1.09e-4
synapse part

E(bx)

toe

20

jim

Ubx

2.9e-8
presynaptic
active zone

pberm
CG4360
Nf1
unpg
fd68A
Dbx
Atf-2
sima   CG32532
pnr   crol
sqr   E2f
sdm3
vvl
vsx1
br

12

ase
da

nej

connectivity
• 100
• 200
• 300
• 400
● 500
● 600

svp

7

gl

aop

mirr
grh

repo

TF targetome -log2("synap*" best Pval)

0   2   4   6   8

TF "synap*" knowledge

C

Grh best motif   □ Grh motif score
                          0 2 4 6 8 10 12

Grh preticted target regions sorted according to their Ghr i-cisTarget ranking

EA disc chromatin opening
FAIRE-seq vs input

0 8 16 24

Number of TF predicted
to be bound to the region

D

Grh-GFP

Figure 6. Known and New Eye-Related TFs and the Grainyhead Targetome

(A) Representation of previously published "eye knowledge" (x axis), as the number of eye-related papers for that TF versus the enrichment for eye-related GO terms in the set predicted TF target (y axis). Node size and color represent the targetome size. For example, Ey, So, and Glass are well known eye-related TFs with many papers and an eye-enriched targetome. Grh is located in the top left corner, with only one eye-related publication, but its targetome has the highest enrichment for "eye"-related GO terms.
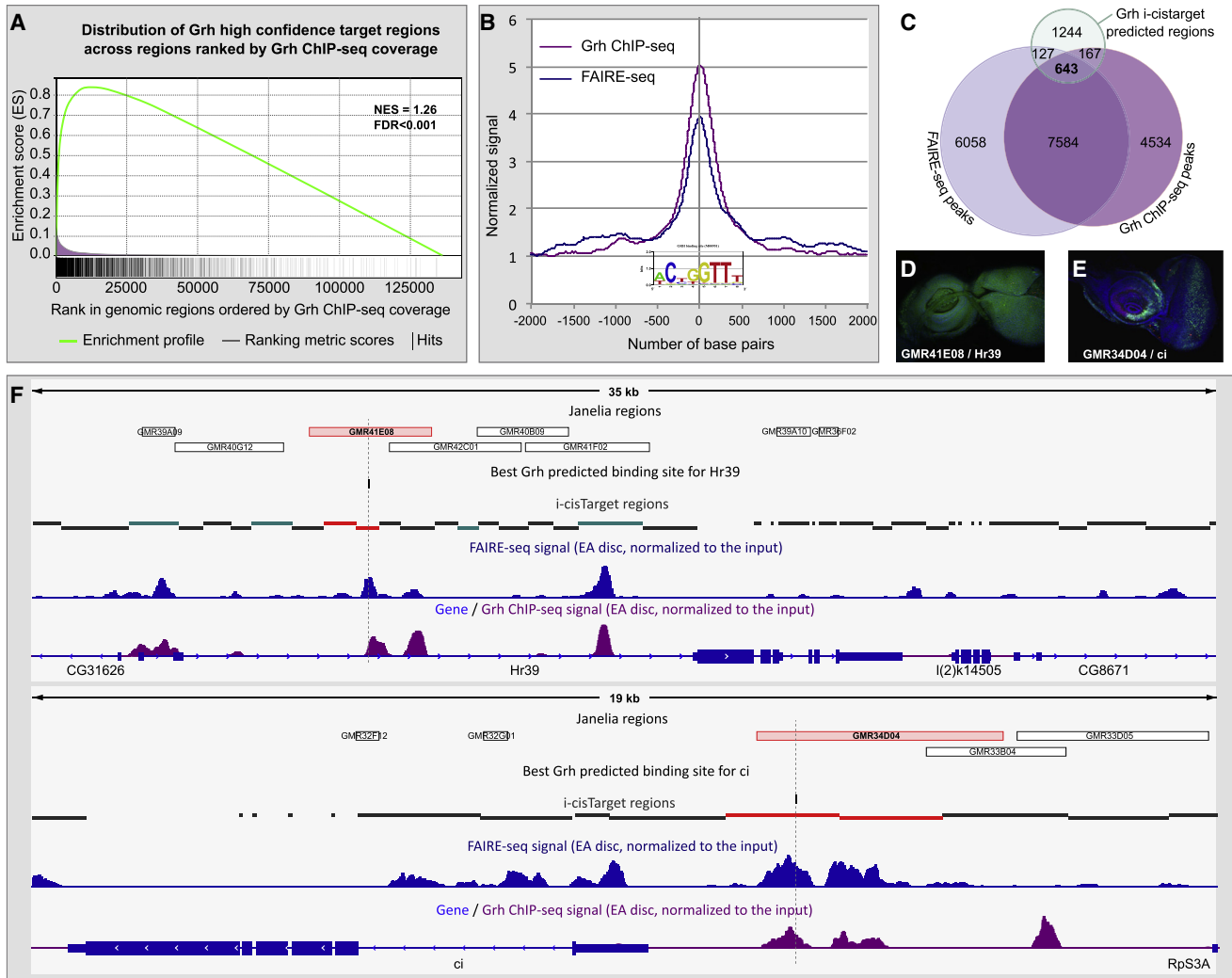
Figure 7. Validation of Predicted Grh Targets by ChIP-Seq

(A) Comparison of predicted Grh binding sites (black lines below x axis) with Grh ChIP-seq signal (x axis) showing significant (FDR = 0.000) enrichment of predicted sites among high Grh peaks.

(B) Aggregation plot showing total Grh ChIP-seq (purple) and chromatin opening (blue) signal in the 4 kb centered around the best Grh binding sites at each Grh predicted region. The best Grh motif is displayed in the center of the panel.

(C) Venn diagram representing the overlap of predicted Grh binding sites (turquoise), open chromatin peaks (blue), and Grh ChIP-seq peaks (purple).

(D) Janelia Gal4 region (GMR41E08, linked to Hr39) that is ubiquitously active in the EA disc.

(E) Janelia Gal4 region (GMR34D04, linked to ci) that is broadly active across the EA disc. Both (D) and (E) are reproduced with permission of Janelia Farm.

(F) Two examples of predicted Grh target regions (red enhancers, other Grh predicted target regions are colored in Turquoise) for Hr39 and ci, with the best Grh binding site represented by a black tick. These predicted enhancers overlap FAIRE-peaks and Grh ChIP-seq peaks (wiggle plot respectively in blue and purple) and represent the enhancer-reporters in (D) and (E) (see also Figure S4 for motif enrichment predictions of each Venn diagram segment).

perturbations into one data matrix to obtain a broad spectrum of expression profile changes. These perturbations included TF mutants, TF overexpression, TF knockdown, and cell sorting.

We dissected eye-antennal discs at the stage where in the WT discs about half of the eye disc contains pluripotent cells that are dividing asynchronously, while the other half contains

(B) Knowledge plot of "synapse"-related publications of a TF (x axis) versus "synapse"-related GO term enrichment of the TF target predictions (y axis). The Onecut cellular component term enrichment is highlighted.

(C) Chromatin opening in the 4 kb centered around the best Grh binding sites at each Grh predicted region. Regions are ranked according to the i-cisTarget ranking. The number of predicted TF inputs for the region is shown in blue.

(D) Anti-GFP immunostaining of Grh-GFP fusion third instar larval eye-antennal disc (see also Figure S2 and Table S6 for the Ey and So predicted targetome, Figure S3 for the Onecut targetome, Figure S5 for the "antennal" knowledge plot, Table S7 for onecut[x652] versus WT RNA-seq, and Table S8 for FAIRE-seq peak enrichment with the set of predicted target enhancers of each TF).

differentiating PR neurons, in consecutive stages of differentiation. Simultaneously, the antennal disc contains neuronal precursors that are undergoing specification. The expression changes induced by the perturbations often result from a shift in proportion of cell types. This is trivial for the cell-sorting experiments; for example, the GMR>GFP-positive cells show, as expected, a very strong enrichment of genes related to PR differentiation (Table S2). TF mutants and TF perturbations can also result in cell type shifts; for example, overexpression of Atonal yields more R8 PRs (Aerts et al., 2010), and the *glass* mutant results in fewer differentiated PRs (Moses et al., 1989a). Other TF perturbations cause changes in gene expression downstream of the TF without changing the cell type composition, such as Retained, which disturbs axonal projection (Ditch et al., 2005). The key technique that we applied, however, is not to compare each TF perturbation with WT discs to identify differentially expressed genes. Rather, we used (linear and nonlinear) correlations of gene expression profiles across the entire vector of 72 gene expression measurements. This *TF-gene* coexpression network contains both direct and indirect edges, and although this network is informative, we added a second layer of predicted *TF-DNA* interactions, thus making this a direct GRN. To increase the sensitivity, we used a very large collection of TF motifs, also including position weight matrices derived for yeast and vertebrate TFs and including computationally derived motifs (e.g., highly conserved words). Using motif-motif similarity measures and TF-TF orthology relationships, we link each motif to a candidate binding factor, as described elsewhere (Herrmann et al., 2012; Janky et al., 2014). This yielded a large network with 335 TFs and their predicted direct targets. The only functional network of comparable size and comparable directedness to our in vivo network is the TH17 GRN that was derived in vitro in a recent study (Yosef et al., 2013). Yosef et al. (2013) used a microarray time course of naive CD4+T cells differentiating into TH17. From these gene expression data, they derived TF-gene interactions by clustering and filtered those with TF-DNA interactions obtained by ChIP-seq data, TF perturbations, and *cis*-regulatory sequence analysis.

The predicted direct and functional eye-antennal GRN includes many previously reported interactions, such as known target genes for Eyeless and Sine Oculis (Figures S2A–S2D). We also found target genes in our network for late factors (e.g., Glass, Onecut) and very late factors (e.g., Pph13). The fact that we capture information at different time points during development is because we sorted several cell populations that are loosely correlated with the temporal axis of development, consisting of undifferentiated pluripotent cells anterior to the furrow, all PR cells undergoing differentiation posterior to the MF, R8 PR cells (Figures S6A and S6B), and late populations of *chp*-positive cells (Figure S6C). However, the temporal information encoded in the network is limited to these broad domains, and a more detailed reconstruction of the time axis would require higher resolution cell sorting or microdissection experiments. Although the perturbed TFs were mainly chosen for their development of the retina, we could also identify master regulators of antennal development, such as *aristaless* (Figure S5).

Interestingly, we also found general factors like Grainyhead that are ubiquitously expressed. Grh was found as one of the TFs with the largest number of target enhancers and its binding correlates with open chromatin. Previous studies have shown that Grainyhead may interact with Polycomb and Trithorax proteins to regulate (both activate and repress) target gene expression (Blastyák et al., 2006; Schuettengruber et al., 2007). We speculate that this observed correlation can be explained by the fact that Grh is present ubiquitously in the eye disc, thus yielding many sequence fragments from bound and nucleosome-free enhancers by FAIRE-seq.

It is well known that network motifs such as FFLs play an important role in biological networks (Le and Kwon, 2013; Mangan and Alon, 2003). We examined one such network motif in more detail, namely the TF pair Glass-Lozenge, and their common targets, in more detail. These TFs constitute a double-feedback loop (Glass regulates Lozenge, Lozenge regulates Glass, and they together regulate 36 targets). For this network motif, we found that Glass and Lozenge motifs co-occur at the same enhancer, where they furthermore overlap; this may indicate competition for binding between Glass and Lozenge. Given that Lozenge, an important regulator of cone cell differentiation, could be a repressor (Canon and Banerjee, 2003) and Glass, an important regulator of PR differentiation, could be an activator (Naval-Sánchez et al., 2013), such a competition at the CRM level could indeed be a plausible mechanism for their regulatory action.

Another interesting feature that can be derived from a GRN is the proportion of autoregulatory TFs (108 autoregulatory TFs in the eye network) and the proportion of activating versus repressive TFs. A recent large-scale study in yeast (Kemmeren et al., 2014) found a small majority of yeast TFs to have a repressive role. In that study, each individual TF was perturbed, thereby providing information on positive versus negative edges from the TF to its direct predicted targets, whereby TF-DNA information was used from ChIP-chip data. Since we started building the eye GRN from a coexpression TF-gene network, we revisited the correlations between TFs and their candidate targets and found 151 TFs that have their motif enriched in the positively correlated target genes, but not in the negatively correlated targets, and 127 TFs showing the opposite; 62 TFs show enrichment in both. This finding agrees, to some extent, with the results in yeast from Kemmeren et al. (2014) concerning the high amounts of gene-specific repressors. On the other hand, our network suggests relative more TFs with a dual activator/repressor function, while Kemmeren et al. (2014) found in yeast only a few such cases. In conclusion, starting from an expression matrix derived from large-scale perturbations and combining TF-gene coexpression with TF-DNA interactions based on motif inference enabled drawing an extensive eye-antennal GRN. All predicted regulatory interactions, target genes, and candidate regulatory regions are stored in a Neo4J database and can be queried from our website (http://eyenetwork.aertslab.org). The database can also be accessed directly from Cytoscape using the CyNeo4j plugin or can be queried programmatically using the Neo4j query language Cypher. Although we recovered many known regulators and *cis*-regulatory elements and revealed several other ones, a large part of the predicted network, including how the dynamics of the developmental program are encoded in the *cis*-regulatory regions and in the topology of the network, remains to be explored.

EXPERIMENTAL PROCEDURES

Fly Stocks

Fly stocks used for RNA-seq experiments are listed in Table S1. The Janelia FlyLight enhancer-Gal4 lines used for GI binding site predictions are listed in Table S3. The Grh-GFP fusion used is Bloomington stock number 42272. All flies were raised at 25°C on standards fly food. All crosses were set up at 25°C. Crosses with the heat-shock Gal4 driver were treated with a 60 min heat pulse at 37°C every 24 hr starting from first instar larval stage.

RNA-Seq

We used a classical Illumina TruSeq protocol to build the libraries after RNA extraction and sequenced them on the Illumina HISeq 2000 (see the Supplemental Experimental Procedures for more details).

RNA-Seq Data Analysis

Reads were first cleaned using fastx clipper (from fastx_toolkit-0.0.13, options -Q33 -M15 -n -v -l 20) to remove adaptor sequences, and their quality after cleaning was checked using the FastQC software (in case of overrepresentation from different primers, samples were cleaned again to remove the contamination). The reads passing the cleaning step were mapped to *D. melanogaster* genome release 5 using TopHat2.0.3 (default parameters). Gene expression measures were quantified by HT-seq count (option -str = no) using *D. melanogaster* gene annotation release 5.45 from FlyBase. The entire matrix of raw counts was first filtered (minimum of 1 read per million in 5 samples across the 72) and then normalized using edgeR (v.2.4.6) and DESeq (v.1.6.1) variance-stabilizing transformation in R (v.2.14.0).

TF-Gene Coexpression Network

Two methods were used to build the coexpression network, namely, GENIE3 (Huynh-Thu et al., 2010) and PTM (Pavlidis and Noble, 2001). GENIE3 was run with mtry option set to sqrt, and we selected all TF to TG pair with an interaction score higher at least than 0.005. Four different thresholds were selected from the most permissive to the most stringent: 0.005, 0.006, 0.007, and 0.008. PTM was performed starting from the template (expression profile) of each expressed TF in R (v.2.14.0) using three different options—greater (for coexpressed genes), less (for antiexpressed genes), and two sided (both at the same time)—and four different p value thresholds (10×10^{-7}, 10×10^{-8}, 10×10^{-9}, 10×10^{-10}).

Direct TF-DNA Network by Motif Inference

Each of the 694 TFs times four thresholds (GENIE3 data sets) and the 694 TFs times four thresholds times three correlations (PTM data sets) (16 data sets per TF for a total of 11,072 data sets) was converted to FlyBase annotation r5.37 (corresponding to the i-cisTarget database), and an i-cisTarget analysis was performed. We used the "motif2TF" links from iRegulon to link PWMs to *Drosophila* TFs (Janky et al., 2014). From this analysis, three different layers of confidence were derived: (1) a high-confidence layer, where the TF to TG link is drawn from direct evidence. The motif of the TF is found enriched in the regulatory sequences of one of its 16 corresponding data sets. If it is found in a PTM "greater" coexpression data set, we attribute a direct activation; in PTM "less," we attribute a direct repression. In PTM "two-sided" or in GENIE3 data sets, we attribute an undefined direct link. (2) In a medium-confidence layer, the TF to TG link is drawn if the TF corresponding to the motif found to be enriched is part of the input data set. (3) In a low-confidence layer, the link between TF and TG is drawn when the TF motif is discovered in any coexpressed gene set of another TF.

Immunohistochemistry

For immunohistochemistry, imaginal discs of wandering third instar larvae were dissected and processed as described in (Wang et al., 2002). The antibodies against Glass, Eya, and Elav raised by G.M. Rubin were obtained from the Developmental Studies Hybridoma Bank, developed under the auspices of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and maintained by the University of Iowa, Department of Biology. The antibody against GFP, the Alexa Fluor 647-488-555, and secondary antibodies were obtained from Life Technologies.

FAIRE-Seq Peaks and TF Correlations

EA disc FAIRE-seq and input data from (McKay and Lieb, 2013) were download from the NCBI Gene Expression Omnibus (GSE38727: samples GSM1261348, GSM948724, and GSM948725) and peaks called using MACS2 (Zhang et al., 2008) callpeak (options: -g dm -bdg -nomodel) comparing merged EA with merged input to produce a peak file (_peak.encodePeak), an EA-treated file (_treat_pileup.bdg), and a control file (_control_lamba.bdg). The two last files are used with macs bdgcmp (-m ppois -t _treat_pileup.bdg -c _control_lambda.bdg) to create the signal file. To visualize EA FAIRE-seq signal, we plotted read coverage per base pair for each predicted region in the 4 kb centered on the best Grh motif occurrence (cbust/options -m3 -c0) inside the i-cisTarget predicted Grh CRM. Enrichments for TF targetomes in EA FAIRE-seq peaks were calculated by counting the number of i-cistarget region overlapping a peak (minimum of 40% in both ways, 4,105/24,926 regions present in the network) for each targetome; p values for all TF region targetomes were then computed using a hypergeometric test comparing the number of region with a peak in a targetome with the number of regions having a peak in the network as reference.

Grh ChIP-Seq

Eye-antennal imaginal discs are dissected from Grh-GFP (Bloomington stock 42272) third instar larvae and fixed with formaldehyde. Chromatin is prepared and sonicated until fragments reach an average size of 500 bp. Chromatin is immunoprecipitated with an anti-GFP Ab (ab290, Abcam), and the immunocomplexes are recovered with protein A/G magnetic beads (Millipore). ChIP libraries are prepared with the Truseq DNA library prep kit (Illumina), and the samples are sequenced on a HiSeq 2000 (Illumina) (see Supplemental Experimental Procedures for the extensive protocol). For the data analysis, 101 bp sequence reads were first cleaned using fastx clipper, checked using the FastQC software, and mapped to the dm3 genome using bowtie2-1.0.0-beta3 (options -q-seed 1 -local). Reads were filtered for a minimum mapping quality of four. We used macs2 "callpeak" and "bdgcmp" with the same options as in the FAIRE-seq analysis above (using another input sample from WT CantonS flies). For the aggregation plots, we summed the read coverage per base pair in the 4 kb regions centered on the best Grh motif occurrence (cbust/options -m3 -c0) inside the i-cisTarget predicted Grh CRM and normalized it by the total signal.

Data and Network Availability

The high- and medium-confidence network is available via an online web application hosting a Neo4j graph database: http://eyenetwork.aertslab.org. A dump of the database can be downloaded, and the database can be queried Cypher or Cytoscape using the CyNeo4j plugin (see the help page). Finally, A UCSC track hub is available with all CRM predictions for Grainyhead, alongside with FAIRE-seq and ChIP-seq data, from http://genome.ucsc.edu/cgi-bin/hgTracks?db=dm3&hubUrl=http://ucsctracks.aertslab.org/Network-paper/hub.txt.

ACCESSION NUMBERS

All RNA-seq and Grh-GFP ChIP-seq experiment fastq files and processed files were deposited to the NCBI Gene Expression Omnibus under accession number GSE62558.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.11.038.

AUTHOR CONTRIBUTIONS

REFERENCES

Aerts, S., Quan, X.-J., Claeys, A., Naval Sanchez, M., Tate, P., Yan, J., and Hassan, B.A. (2010). Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in Drosophila uncovers a regulatory basis for sensory specification. PLoS Biol. 8, e1000435.

Amore, G., and Casares, F. (2010). Size matters: the contribution of cell proliferation to the progression of the specification Drosophila eye gene regulatory network. Dev. Biol. 344, 569–577.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382–390.

Bessa, J., Gebelein, B., Pichaud, F., Casares, F., and Mann, R.S. (2002a). Combinatorial control of Drosophila eye development by eyeless, homothorax, and teashirt. Genes Dev. 16, 2415–2427.

Blastyák, A., Mishra, R.K., Karch, F., and Gyurkovics, H. (2006). Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol. Cell. Biol. 26, 1434–1444.

Bonnet, E., Michoel, T., and Van de Peer, Y. (2010). Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data. Bioinformatics 26, i638–i644.

Calero-Nieto, F.J., Ng, F.S., Wilson, N.K., Hannah, R., Moignard, V., Leal-Cervantes, A.I., Jimenez-Madrid, I., Diamanti, E., Wernisch, L., and Göttgens, B. (2014). Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. EMBO J. 33, 1212–1226.

Canon, J., and Banerjee, U. (2003). In vivo analysis of a developmental circuit for direct transcriptional activation and repression in the same cell by a Runx protein. Genes Dev. 17, 838–843.

Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkurst, C.N., Muratet, M., et al. (2012). A validated regulatory network for Th17 cell specification. Cell 151, 289–303.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science 295, 1669–1678.

De Cegli, R., Iacobacci, S., Flore, G., Gambardella, G., Mao, L., Cutillo, L., Lauria, M., Klose, J., Illingworth, E., Banfi, S., and di Bernardo, D. (2013). Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. Nucleic Acids Res. 41, 711–726.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J.S., Hope, I.A., et al. (2006). A gene-centered C. elegans protein-DNA interaction network. Cell 125, 1193–1205.

Ditch, L.M., Shirangi, T., Pitman, J.L., Latham, K.L., Finley, K.D., Edeen, P.T., Taylor, B.J., and McKeown, M. (2005). Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. Development 132, 155–164.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Frankfort, B.J., Nolo, R., Zhang, Z., Bellen, H., and Mardon, G. (2001). senseless repression of rough is required for R8 photoreceptor differentiation in the developing Drosophila eye. Neuron 32, 403–414.

Fuse, N., Hirose, S., and Hayashi, S. (1994). Diploidy of Drosophila imaginal cells is maintained by a transcriptional repressor encoded by escargot. Genes Dev. 8, 2270–2281.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature 489, 91–100.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99–104.

Herrmann, C., Van de Sande, B., Potier, D., and Aerts, S. (2012). i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Res. 40, e114.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5, e12776.

Janky, R., Verfaillie, A., Imrichová, H., Van de Sande, B., Standaert, L., Christiaens, V., Hulselmans, G., Herten, K., Naval Sanchez, M., Potier, D., et al. (2014). iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput. Biol. 10, e1003731.

Jenett, A., Rubin, G.M., Ngo, T.-T.B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for Drosophila neurobiology. Cell Rep 2, 991–1001.

Jory, A., Estella, C., Giorgianni, M.W., Slattery, M., Laverty, T.R., Rubin, G.M., and Mann, R.S. (2012). A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of Drosophila melanogaster. Cell Rep 2, 1014–1024.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E.M. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell 148, 473–486.

Kemmeren, P., Sameith, K., van de Pasch, L.A.L., Benschop, J.J., Lenstra, T.L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C.W., et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell 157, 740–752.

Kumar, J.P. (2009). The molecular circuitry governing retinal determination. Biochim. Biophys. Acta 1789, 306–314.

Le, D.-H., and Kwon, Y.-K. (2013). A coherent feedforward loop design principle to sustain robustness of biological networks. Bioinformatics 29, 630–637.

MacArthur, S., Li, X.-Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keränen, S.V.E., et al. (2009). Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol. 10, R80.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. Proc. Natl. Acad. Sci. USA 100, 11980–11985.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., and Stolovitzky, G.; DREAM5 Consortium (2012). Wisdom of crowds for robust gene network inference. Nat. Methods 9, 796–804.

McKay, D.J., and Lieb, J.D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. Dev. Cell 27, 306–318.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Moses, K., Ellis, M.C., and Rubin, G.M. (1989a). The glass gene encodes a zinc-finger protein required by Drosophila photoreceptor cells. Nature 340, 531–536.

Naval-Sánchez, M., Potier, D., Haagen, L., Sánchez, M., Munck, S., Van de Sande, B., Casares, F., Christiaens, V., and Aerts, S. (2013). Comparative motif discovery combined with comparative transcriptomics yields accurate targetome and enhancer predictions. Genome Res. *23*, 74–88.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277–1289.

Pavlidis, P., and Noble, W.S. (2001). Analysis of strain and regional variation in gene expression in mouse brain. Genome Biol. *2*, H0042.

Pepple, K.L., Atkins, M., Venken, K., Wellnitz, K., Harding, M., Frankfort, B., and Mardon, G. (2008). Two-step selection of a single R8 photoreceptor: a bistable loop between senseless and rough locks in R8 fate. Development *135*, 4071–4079.

Pfreundt, U., James, D.P., Tweedie, S., Wilson, D., Teichmann, S.A., and Adryan, B. (2010). FlyTF: improved annotation and enhanced functionality of the Drosophila transcription factor database. Nucleic Acids Res. *38* (Database issue), D443–D447.

Quan, X.J., Ramaekers, A., and Hassan, B.A. (2012). Transcriptional control of cell fate specification: lessons from the fly retina. Curr. Top. Dev. Biol. *98*, 259–276.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al.; modENCODE Consortium (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787–1797.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E.M. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev. *21*, 436–449.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Spinelli, L., Gambette, P., Chapple, C.E., Robisson, B., Baudot, A., Garreta, H., Tichit, L., Guénoche, A., and Brun, C. (2013). Clust&See: a Cytoscape plugin for the identification, visualization and manipulation of network clusters. Biosystems *113*, 91–95.

Spokony, R., and White, K. (2012). Spokony insertions. http://flybase.org/reports/FBrf0220060.html.

St Pierre, S.E., Ponting, L., Stefancsik, R., and McQuilton, P.; FlyBase Consortium (2014). FlyBase 102—advanced approaches to interrogating FlyBase. Nucleic Acids Res. *42* (Database issue), D780–D788.

Wang, V.Y., Hassan, B.A., Bellen, H.J., and Zoghbi, H.Y. (2002). Drosophila atonal fully rescues the phenotype of Math1 null mice: new functions evolve in new cellular contexts. Curr. Biol. *12*, 1611–1616.

Warner, J.B., Philippakis, A.A., Jaeger, S.A., He, F.S., Lin, J., and Bulyk, M.L. (2008). Systematic identification of mammalian regulatory motifs' target genes and functions. Nat. Methods *5*, 347–353.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

Yan, H., Canon, J., and Banerjee, U. (2003). A transcriptional chain linking eye specification to terminal determination of cone cells in the Drosophila eye. Dev. Biol. *263*, 323–329.

Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M., and Taipale, J. (2013). Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. Cell *154*, 801–813.

Yao, L.-C., Blitz, I.L., Peiffer, D.A., Phin, S., Wang, Y., Ogata, S., Cho, K.W.Y., Arora, K., and Warrior, R. (2006). Schnurri transcription factors from Drosophila and vertebrates can mediate Bmp signaling through a phylogenetically conserved mechanism. Development *133*, 4025–4034.

Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M., et al. (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461–468.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.