A Cis-Regulatory Map of the Drosophila Genome

Nicolas Nègre1*, Christopher D. Brown1*, Lijia Ma1*, Christopher Aaron Bristow2*, Steven W. Miller3*, Ulrich Wagner5*, Pouya Kheradpour2, Matthew L. Eaton14, Paul Loriaux4, Rachel Sealfon2, Zirong Li5, Haruhiko Ishii3, Rebecca F. Spokony1, Jia Chen6, Lindsay Hwang5, Chao Chen,14,15, Richard P. Auburn7, Melissa B Davis1, Marc Domanus1, Parantu K. Shah8, Carolyn A. Morrison1, Jennifer Zieba1, Sarah Suchy1, Lionel Senderowicz1, Alec Victorsen1, Nicholas A. Bild1, A. Jason Grundstad1, David Hanley6, David M. MacAlpine14, Mattias Mannervik9, Koen Venken10, Hugo Bellen10, Robert White11, Steven Russell7, Robert L. Grossman1,6,12, Bing Ren5,13, Mark Gerstein,14,15,16, James W. Posakony3, Manolis Kellis2 and Kevin P. White1

1. Institute for Genomics & Systems Biology, Department of Human Genetics, The University of Chicago, 900 East 57th Street, Chicago, IL 60637, USA.
2. Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Broad Institute of MIT and Harvard, Cambridge MA 02139, USA.
3. Division of Biological Sciences/CDB, University of California San Diego, La Jolla, CA 92093, USA.
4. Signaling Systems Laboratory, Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA.
5. Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093-0653
6. National Center for Data Mining, University of Illinois at Chicago, 851 S. Morgan Street, Chicago IL 60607, USA.
7. Department of Genetics and Cambridge Systems Biology Centre, University of Cambridge, Cambridge, CB2 3EH, UK
8. Department of Biostatistics & Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston MA 02115, USA.
9. Department of Developmental Biology, Wenner-Gren Institute, Arrhenius Laboratories E3, Stockholm University, S-106 91 Stockholm, Sweden
10. Department of Molecular and Human Genetics, BCM, Houston, TX, 77030
11. Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3DY, UK
12. Department of Medicine, University of Chicago 5841 South Maryland Avenue, Chicago, IL 60637
13. Department of Cellular and Molecular Medicine, Institute of Genomic Medicine and Moores Cancer Center, 9500 Gilman Drive, La Jolla, CA 92093
14. Department of Pharmacology and Cancer Biology, Duke University Medical
Center, Durham, NC 27710 USA
15. Program in Computational Biology and Bioinformatics and Department of Molecular Biophysics and Biochemistry, Yale University, Bass 432, 266 Whitney Avenue, New Haven, CT 06520
16. Department of Computer Science, Yale University, Bass 432, 266 Whitney Avenue, New Haven, CT 06520

*These authors contributed equally to this work
*Corresponding authors: (KPW, project director, kpwhite@uchicago.edu; MK, data analysis, manoli@mit.edu; JWP, biological validation, jposakony@ucsd.edu; BR, chromatin, biren@ucsd.edu; RLG, informatics, grossman@labcomputing.org; SR, silencer/insulator analysis, sr120@cam.ac.uk).
Running title

Abstract

We produced a map of the *Drosophila melanogaster* regulatory genome based on the developmental dynamics of chromatin modifications and polymerase occupancy, the localization of chromatin modifying enzymes, and the binding of a wide range of regulatory proteins. We generated over 300 ChIP datasets for eight chromatin features associated with gene regulation, five histone deacetylases (HDACs) and thirty-eight site-specific transcription factors (TFs) at different stages of development. The entire dataset provided protein modification and binding annotations across 90% of the non-repetitive genome. Using these data we inferred more than 20,000 candidate regulatory elements including insulators, promoters, silencers and enhancers; and we validated a subset of predictions for promoters, enhancers and insulators *in vivo*. We also identified over 2,000 genomic regions of dense TF binding and showed that they are associated with chromatin activity and accessibility. We discovered hundreds of new TF co-binding relationships at target genomic sites and defined a TF network with over 800 potential regulatory relationships, including many that are associated with developmentally dynamic gene expression patterns. Together these data and results constitute the first attempt at a comprehensive cis-regulatory annotation of a metazoan genome throughout development.

Introduction

Systematic annotation of gene regulatory elements is a major challenge in genome science. Comparative sequencing, which identifies evolutionarily constrained regions of the genome, has given clues to the locations of a subset of non-coding regulatory elements [1]. However, the comparative approach identifies all constrained genome sequence regardless of function, and cis-regulatory elements have proven difficult to identify based solely on evolutionary conservation. An alternative approach is to directly map chromatin modification marks and transcriptional factor binding sites genome-wide [2]. Such mapping helps to identify specific subtypes of regulatory elements (e.g. promoters, enhancers, silencers, insulators), and this approach has been used with great success in specific cell types [3,4].

A major aim of the modENCODE Project is to systematically annotate cis-regulatory elements in the *Drosophila* genome [5]. In *Drosophila* several pioneering studies have provided genome-wide identifcation of Polycomb-Response Elements [6-8], chromatin states [9], transcription factors [10-15], PolII regulation [13], and insulator elements [16,17]. However, all of these studies were independent of one another and focused on a specific type of gene regulation or developmental process or tissue. Additionally, a limited number of factors have been examined in these studies. Thus, comprehensive annotation of the regulatory genome remains a significant challenge.

Here we describe results from the modENCODE cis-regulatory annotation project. We generated over 300 replicated datasets for a total of 55 different transcription factors (TFs), chromatin modifying enzymes and chromatin modifications in whole animals at different developmental stages. We used the structure in these data (**Supplemental Fig 1**) to provide annotation of putative promoters, enhancers, silencers and insulators, yielding the most comprehensive cis-regulatory map of the *Drosophila* genome to date.

Results

Strategy for systematic ChIP mapping to identify cis-regulatory domains

To maximize identification of cis-regulatory domains genome-wide we performed a developmental time course to reveal chromatin, promoter, and enhancer activity using whole animals. The use of whole animals identifies chromatin marks across tissues, generating the maximum number of marks per sample. While marks specific to rare cell types may not reach the threshold of detection, we nevertheless expect that a large proportion of the genome will be covered. This strategy mirrors successful developmental time courses for annotating gene expression in *Drosophila* [18]. We analyzed six histone modifications, the *Drosophila* CREB Binding Protein (dCBP) and RNA Polymerase II (PolII) across twelve stages of embryonic, larval, pupal and adult development (**Table 1**, **Supplementary Table 1; Supplemental Fig 2**; see **Supplementary Methods**). During development we identified 506,001 occurrences of these eight chromatin-associated features, creating annotations that correspond to 101MB (86.99%) of the non-repetitive genome. To relate these chromatin states to gene activity, we quantified transcript levels by high-throughput cDNA sequencing (RNAseq) using the same biological samples used for ChIP. Chromatin marks obtained from whole animals showed a strong correspondence with gene expression levels, validating that chromatin profiling of mixed tissues can identify regulatory elements (Figure 1, and described below).

Additionally, to further delineate potential cis-regulatory elements genome-wide, we mapped 38 transcription factors (TFs) in different developmental stages and cell types. We selected TFs involved in a broad range of biological functions and processes in order to capture the widest range of potential regulatory elements. A total of 155,048 TF binding sites were identified comprising 35,125 unique TF binding sites (TFBS). Of these, 93.76% (32,906 TFBSs) overlap at least one chromatin feature measured during development. We noted that while the majority of factors are bound in discrete peak regions, others such as Groucho (Gro), Distalless (Dll), Brakeless (Bks) or Chronologically inappropriate morphogenesis (Chinmo) are distributed in larger domains, occasionally interspersed with discrete peaks (**Table1**, **Supplementary Fig.3**). We also characterized the binding distribution of the full complement of *Drosophila* Histone Deacetylases (HDACs). We identified a total of 19,937 HDAC binding sites mapping to 7,692 unique genomic locations. Of these, 99.25% (7,634 unique HDAC genomic sites) overlap with at least one chromatin feature, and 94.58% (7,275 unique HDAC genomic sites) overlap with at least one TF binding location. Together, these data represent the largest genome-wide ChIP study of *Drosophila* TF and chromatin modification enzymes to date. All datasets produced for this study have been made immediately publicly available through the modENCODE consortium (www.modencode.org) and the BioNimbus (https://www.cistrack.org/#/public.php) portals.

Developmental dynamics of chromatin

To assess the distribution of regulation-associated chromatin marks during development in more detail, we determined the number of annotated genes that were associated with each mark at each developmental time-point (**Fig. 1a**). Very few genes displayed either repressive or activating marks across all of development; most genes were within dynamically marked regions (**Fig. 1a**). We observed distinct and dynamic combinatoric patterns of chromatin mark distribution, including those that corresponded to active chromatin (H3K4me3, H3K9Ac, HeK27Ac) that is expected to overlap with promoters, those associated with repressive states and silencers (H3K27me3, PHO/PREs), and those associated with enhancers (CBP, H3Kme3) (**Fig. 1b**).

We examined these different distributions in more detail. The repressive chromatin marks HK9me3 and H3K27me3 are distributed in large domains throughout development (**Supplementary Figs 2, 4, and 5)**. H3K9me3 marks localize to ~20 domains at centromeres as expected [19], and these are stable across development. H3K27me3 marks, in contrast, are remarkably dynamic (**Fig 1c**). Dynamic domains may be due to changes in specific cell populations during development or the active addition and removal of H3K27me3 marks [20]. Previous studies have implicated H3K27me3 mark dynamics in the regulation of homeotic genes [21], in the differentiation of stem cells [4,22], and in developmental processes in vertebrates [23]. We found an average of 258 discrete domains present at the various developmental stages assayed (range 123 in adult males … 438 in 0-4 hr embryos), each with an average length of approximately 70kb (**Supplementary Fig. 5a; Supplementary Table 2**). A total of 1,264 genes are associated with H3K27me3 in at least one stage of development, with 397 genes (31%) in H3K27me3 domains present in all stages of development and 867 genes (69%) in dynamic domains (**Fig. 1c**). Stable H3K27me3 domains correspond to those reported in embryos and tissue culture cells, and are enriched in genes involved in development, transcription and segmentation [6,7,24]. However, identification of stage-specific H3K27Me3 domains revealed previously unappreciated H3K27Me3 targets, including genes that control apoptosis, regulation of growth, and neurotransmitter transport (respectively larval, pupal and adult stages) (**Supplementary Fig. 6**). Taken all together, H3K27me3 domains are highly enriched for genes that exhibit stage and tissue-specific expression, and are depleted for ubiquitously expressed genes (**Supplementary Fig 7**).

The activating histone modifications H3K4me3, H3K9Ac, and H3K27Ac are, as expected, positively correlated with gene expression levels (**Fig. 1b)**. Regions marked by each activating modification are significantly enriched within the marked regions of the other activating modifications, but also with class I insulator binding sites, PolII binding sites, and a large fraction of transcription factor binding sites (**Supplemental Fig 1**). We classified transcripts at each of 12 developmental stages as detectably expressed ('detected') or not detectably expressed ('undetected') (see **Supplementary Methods**). Consistent with previous observations, hypophosphorylated PolII and H3K4me3 are tightly associated with detected TSSs, with peak occupancy approximately 150 bp downstream of the TSS (**Fig. 2a**) [7,25], [26], [27]. H3K27Ac, H3K9Ac, and H3K4me1 are also enriched at active TSSs (**Supplementary Fig 8**). In contrast, H3K27me3, a hallmark of repressive chromatin states, is not enriched at active TSSs (**Fig. 1b** and data not shown).

Although combinations of chromatin marks and TF binding sites are very powerful for identifying cis-regulatory elements, we noted that a substantial fraction of genes show evidence of expression but have no detectable H3K4me3 marks at their TSSs. We therefore classified genes at each stage as H3K4me3 "marked" or "unmarked" (see **Supplementary Methods**). Surprisingly, 32% of genes were detected but lacked significant H3K4me3 at their annotated TSS. To further investigate the large number of unmarked-detected genes we built, for each gene at each developmental stage, a logistic regression model to predict gene expression status based on the local ChIP signal for all 6 activating marks, CBP, and PolII (**Supplementary Figs. 9,10**). At 95% precision, between 13% and 31% of genes are unmarked-detected, depending on developmental stage (**Fig. 2d; Supplementary Fig. 10**). Compared to marked genes, unmarked-detected genes have similar spatial expression patterns (**Supplementary Fig. 11a**) and GO annotations. However, genes expressed in stage-specific expression patterns are more frequently unmarked (**Supplementary Fig. 11b**, **Supplementary Fig. 12**). The absence of epigenetic marks at active promoters could reflect a lack of sensitivity of the ChIP assay in a mixed population of embryonic cells. However, we find that in synchronized Kc167 cells, we also observe a significant proportion (9%) of genes are unmarked-detected (**Supplementary Fig13**). Furthermore, 5 of 6 unmarked-detected genes do not show significant H3K4me3 signal when assayed by ChIP-qPCR (**Supplementary Fig13**).

H3K4me1-marked and CBP/p300-bound regions form a third, intermediate class of genomic elements (**Figs. 1b and 2c**). Previous work has demonstrated that H3K4me1 and CBP/p300 are associated with active enhancers [3,26-29]. Accordingly, across the developmental time course, these elements are moderately associated with active promoters, activating histone marks, and transcription factor binding sites. H3K4me1 and CBP are bound more broadly across gene TSSs, typically positioned 1-2 kb upstream and downstream of the TSS, consistent with previous observations (**Fig. 2a** and **Supplementary Fig 8**) [27,30].

To characterize the regulators of chromatin mark dynamics, we characterized the genome-wide distribution pattern of all five known *Drosophila* HDACs (HDAC1/Rpd3, HDAC3, HDAC4a, HDAC6 and HDACX/11), reasoning that this approach would detect both acetylated regions and the marks that replace the acetyl residues, thus covering a substantial fraction of the regulatory genome. All five HDACs are enriched at active promoters, and HDAC enrichment is correlated with target gene expression level over several orders of magnitude, with highly expressed genes more likely to be bound (**Supplementary Fig. 14c**). Closer examination of the binding profiles revealed two additional associations: first, HDAC3 is primarily associated with transcribed exons (**Supplementary Fig. 14a, d**), which are also marked by H3K36me3[31]. Second, we noted that in addition to their association with active genes, HDAC4a and HDAC1/Rpd3 binding sites are frequently located within H3K27me3 repressive domains (**Supplementary Figs. 15 and 16**), raising the possibility that they identify the Polycomb-group responsive element (PRE) class of silencers. Indeed, HDAC4a and HDAC1 binding sites are significantly enriched at embryonic PHO (a PcG recruiter protein) bound regions (**Supplementary Fig. 14f**). To further characterize this relationship, we identified HDAC1 and HDAC4a binding sites contained within H3K27me3 domains but not overlapping with H3K4me3 (**Supplementary Fig 16**) to predict a set of 537 putative PREs. Out of 350 embryonic PHO sites previously described [8], 149 overlap with our predicted PREs (**Supplementary Fig 15 and 16**), indicating that HDAC1/Rpd3 and HDAC4a are indeed strongly associated with PREs.

Additionally, our observations indicate an evolutionary shift in the association of HDACs with corresponding chromatin modification marks. In CD4+ T cells human HDAC 1, 2, and 3 are associated with active promoters, while human HDAC6 is associated with transcribed exons [32]. Human and Drosophila HDAC 1, 3 and 6 appear to be orthologous [33], but *Drosophila* HDAC3 is associated with transcribed exons while HDAC6 is associated with active promoters. These findings were consistent for two independent antibodies raised against different domains for each factor. Therefore the specificity of human and *Drosophila* HDAC3 and HDAC6 distributions appears to have swapped, and the structure-function relationships of these proteins have not been conserved at the level of histone associations.

Annotation and prediction of cis-regulatory sequences

Using the dynamic chromatin signatures and RNAseq data described above, we sought to identify putative cis-regulatory elements. To identify previously unannotated promoters, we compiled genomic regions where coincident H3K4me3, PolII, and RNA signals are at least 1,000 base pairs away from any annotated transcription start site (see Supplemental Methods). In each developmental stage we found several hundred such regions (average, 485; range, 179-885), resulting in 2,307 total novel promoter predictions. We compared these predictions to modENCODE cap analysis of gene expression (CAGE) data from embryos [34]. Of the 2,307 novel promoter predictions from all developmental stages, 1,117 are supported by embryonic CAGE data (**Fig. 2a**). Independent of the CAGE comparison, we subjected 110 novel promoter predictions to biological validation using a luciferase reporter assay in Kc167 cells, including similarly predicted Kc167 cell novel promoters and novel promoters from embryonic stages [35]. 75 of these 110 predicted promoters (69%) yield significant luciferase activity in at least one orientation, with 26 displaying bi-directionality (**Fig. 2b; Supplementary Table 3**). Together, the CAGE data and the reporter gene validation assay indicate that a high proportion of these novel promoter predictions indeed correspond to previously unannotated transcriptional start sites.

In order to identify enhancers on a genome-wide scale, we examined two signatures of enhancers, H3K4Me1 and CBP/p300 [27,29,30]. To quantify the association between CBP, H3K4me1 and enhancers, we compared these data with the recently published CRM Activity Database (CAD) [14](**Fig. 2c, Supplementary Fig. 17**). Known enhancers are more likely to overlap CBP and H3K4me1 regions from early embryos, highlighting the bias of the CAD database for embryo-specific regulatory elements. For example, we found a 15-fold (z-score of 26) and 5.9-fold (z-score of 10) enrichment for CBP and H3K4me1 overlap, respectively, with blastoderm-specific enhancers, indicating that our dynamic chromatin map successfully recovers previously annotated enhancers.

To characterize the cis-regulatory function of CBP binding sites, we grouped distal CBP regions based on the binding pattern of TFs and chromatin remodeling factors from this and published work. CBP binding sites were converted into binary TF occupancy vectors, which were used to cluster regions with a finite mixture model of multivariate Bernoulli distributions (**Supplementary Figs. 18a,b**). As a control, we compared these data with a length-matched set of random regions. Several CBP clusters are bound by TFs known to physically interact with CBP, such as Bicoid, Dorsal, and Trl/GAF. A subset of clusters are enriched for TSS proximal marks (PolII and H3K4me3) and likely represent unannotated promoters, whereas other clusters are enriched for known enhancers (**Figs. 2d,e**) and are strongly enriched in K3K4me1 and the repressive mark H3K27me3 (S**upplementary Fig. 18c**). In total, 12,285 distinct putative enhancers were identified across the genome using this method.

To validate the ability of CBP binding data to accurately identify cis-regulatory modules, we tested 33 predicted enhancer elements using reporter gene assays in transgenic animals. We focused on putative enhancers that have developmentally restricted CBP association during embryogenesis. These 33 non-coding sequences are positive for CBP and H3K4me1 at an average of 4.3 and 4.4 stages, respectively, across development; four are within static H3K27me3 domains, 19 are within developmentally dynamic H3K27me3 domains, and 11 enhancer predictions fall outside of any H3K27me3 domain **(Supplementary Table 4).** In contrast, seven negative controls that show no detectable reporter gene expression in transgenic assays are positive for CBP and H3K4me1 at an average of 1.2 and 0.1 developmental stages, respectively. Thirty of the 33 predicted enhancer elements produce specific reporter expression patterns, identifying previously unknown cis-regulatory modules active during embryonic stages (**Fig 2f; Supplementary Fig 19**). These results indicate that information from whole-animal chromatin dynamics can guide the identification of regulatory sequences regulating tissue and stage-specific expression patterns. Such enhancer discovery/validation approaches will lead to a more precise evaluation of how broadly CBP, H3K4me1, and other marks coincide with functional enhancers.

Finally, we selected a set of putative insulator binding sites and tested their activity in an enhancer-blocking assay based on the *eve* stripe 2 and 3 enhancers. We assayed a set of 15 genomic fragments associated with the binding of CP190 + CTCF (class I), CP190 + Su(Hw) (class II) and GAF [17]. In agreement with our subdivision of insulator-binding proteins into two classes, we found that five of eight CTCF sites showed strong enhancer-blocking activity and the remaining three showed weak or variable activity. In contrast, neither of the GAF sites nor any of the five Su(Hw) sites we tested blocked enhancer-promoter interactions in our assay (**Supplementary Fig 20**). These results support a role for CTCF in insulator activity in vivo, but suggest that other proteins that have classically been associated with insulator activity are not strictly linked to this function.

Transcription factor binding sites and gene regulation

To further annotate predicted enhancers and to determine whether dynamics of chromatin and gene expression from whole animals can be associated with specific factors, we mapped 38 TFs involved in a range of known biological processes at various developmental stages. As expected, we found many examples of TF binding to well characterized enhancers. We compared our data with the CAD database (**Fig. 2g**) and observed that many factors are specifically enriched in particular enhancer classes. For example, blastoderm enhancers showed enrichment for Engrailed (En), Even-skipped (Eve) and Gro binding sites, among other TFs.

Indeed, enhancers are usually characterised by multiple TFs binding in concert to target genomic DNA. Thus, determining which TFs bind in proximity to one another can potentially reveal relationships between TFs. We therefore compared genome-wide overlap for each pair of factors to identify significantly co-occurring transcription factor pairs (**Supplemental Fig 21**). We found extensive overlap in TF binding sites (**Fig. 3a**). Of 38,536 unique binding sites mapped by the 38 modENCODE TFs, 38.3% are bound by more than two factors, 5.2% sites are bound by more than eight factors, and 2.6% are bound by ten or more factors (Supplementary Table 5). Regions with large numbers of TF binding sites may reflect complex regulation, or may be areas of open chromatin that tend to attract TFs but have no function in gene regulation [11,36]. We wished to systematically define such regions. Using Gaussian kernel density estimation across the binding profiles of 41 TFs mapped in early embryos in this and two previously published studies [11,14], we defined a 'TF complexity' score based on the number and proximity of contributing TFs (see **Supplemental Methods**). We thus identified 2,006 regions with a complexity of eight or more, and considered these as High Occupancy Targets (HOT) regions. HOT regions appear to be a conserved property of metazoan genomes, as they also have been observed in *C. elegans* (Gerstein, submitted) and human (ENCODE project, unpublished results). As expected, factors that are significantly enriched within the binding regions of other factors are associated with higher complexity categories (**Supplementary Fig. 21, 22**). We found that regions of higher complexity are weakly associated with more highly expressed genes ($r^2 = 0.19$), suggesting that low-complexity binding sites are associated with more restricted expression patterns. Interestingly, annotated enhancers, CBP, activating histone marks including H3K4me1, and HDACs 1, 4a, 6, and 11 are most significantly enriched within low to moderate complexity category (CC) regions (CC2-CC8) (**Fig 3b**). These enrichments consistently decrease at regions of high complexity (CC8-16). In contrast, we found that coding exons and HDAC3, which marks actively transcribed exons (**Fig 3b**, **Supplemental Fig. 22** [31]), are depleted from moderate to high complexity regions (>CC4). To distinguish whether binding of TFs in HOT regions is indeed due to increased chromatin accessibility, we compared nucleosome enrichment data from embryos to HOT regions determined from TFs assayed in similarly staged embryos [37]. We observed a significant correlation between nucleosome depleted regions and HOT regions (**Fig. 3b**). Interestingly, when compared to our enhancer validations and negative controls that were selected independent of HOT spot determination, there appears to be no obvious relationship between enhancer activity and HOT spots; 13 validated enhancers overlap with HOT spots but so did several sequences that give no enhancer activity (**Fig. 2f**; **Supplementary Table 4 and data not shown**). Taken together, these results indicate HOT regions are primarily associated with open chromatin but not necessarily cis-regulatory elements.

The existence of HOT spots presents a problem for simple interpretation of which co-occurring TFs are functionally related. For example, pair-wise clustering of TF binding sites resulted in very large groups of co-occurring TFs, revealing few specific relationships **(Supplemental Fig. 21**). For this reason, we removed all HOT regions defined above from our binding site dataset and re-computed pair-wise TF overlap enrichments for all 38 modENCODE TFs as well as TFs from 20 publicly available datasets. The resulting clustergram shown in **Fig. 4c** reveals structure that is otherwise obscured when HOT regions are included. For example, binding sites from different stages assayed for the same TF show tighter clustering in the hotspot subtracted data (e.g. Trl, Ubx. EcR), positive controls from previous genome-wide mapping studies show tight clustering (e.g. Tin-Twi, Bin-Bam), and the same factor technically repeated at the same stage shows tight clustering (e.g. Cad, Gro).

TFs known to physically interact with one another at specific enhancers also showed highly significant association throughout the genome. For example, the co-repressor complex of Groucho (Gro) and Engrailed (En)[38] and the *Drosophila* SWI/SNF chromatin remodeling complex components Brahma (Brm) and Snf5-related 1 (Snr1) show significant co-binding (z > 20; [39]). Co-binding enrichment genome-wide was also observed for TFs that are known to bind independently to particular enhancers, such as Ultrabithorax (Ubx) and En that each bind to the DMX enhancer of the *distalless* (*dll*) gene, and each independently contribute to *dll* repression in different embryonic segments [40]. Dll was itself enriched for co-binding with En, Gro and Ubx, indicating common regulation of target genes. Interestingly, such previously undescribed interactions were seen at levels equal to or of greater significance for known interactions. For example, while the previously reported mesodermal TF dataset[14] (Tin, Twi, Bin, Bam, dMEF2) all had high overlap with one another as expected, these factors also all showed highly significant overlap with Gro, Cad and En. Many other notable overlap pairs were identified, including the Ecdysone Receptor with the GAGA factor Trithorax-like (Trl), the peripheral nervous system (PNS) master regulator Senseless [41] with the axon guidance TF Disconnected (Disco), and the Jak/Stat signaling pathway TF Stat92E with Brm and Snr1 chromatin remodeling complex factors - all potential new connections between well studied regulatory pathways or mechanisms. In total there are 831 very highly significant positive pair-wise co-binding interactions in Fig 4c (Z score > 20; bright red in Fig. 4c), most of which are previously undescribed.

While most significantly associated TF pairs did show positive overlaps, we observed a few instances of highly significant negative associations (shown in blue, **Fig. 3c**). One of the most anti-correlated pairs of TFs is Brakeless (Bks) and Caudal (Cad) in the early embryo. Bks is a co-repressor that has been implicated in gap gene regulation, for example acting to restrict the expression of *knirps* (*kni*) and *giant* (*gt*) in the posterior blastoderm [42]. In contrast, Caudal, activates *kni* and *gt* in the same embryonic domain [43]. Even when Bks and Cad have multiple binding sites nearby one another, they appear to be non-overlapping and in different putative cis-regulatory elements (**Supplemental Fig 23**). The biologically opposing roles of these two TFs appear to have led to the evolution of a very strong repulsion for occupying the same regulatory elements. To our knowledge this genome-wide aversion in terms of TFs co-occupancy has not previously been observed in a metazoan genome. Interestingly, Bks and the peripheral nervous system (PNS) master regulator Senseless [41] also show this aversion in early embryos. Bks is involved in PNS development in the developing eye [44]. Its embryonic role in PNS development is unknown, but these results raise the possibility that it counteracts Sens activity as it does Cad in early embryos.

To further visualize the regulatory interactions among transcription factors, we built an intuitive hierarchy that allows a clear mining of underlying regulatory association between various regulators (**Fig. 4a, Supplemental Fig 24**). This network was constructed using TFs and their target genes from 61 TFs datasets generated by the modENCODE project (pink nodes) and 20 TFs from recently published work [10,11,14] (green and yellow nodes). Specifically, we built a core-hierarchy using a breadth-first search algorithm in a bottom-up fashion. First, the TFs that regulate less than five other TFs formed the bottom layer. Second, we searched the regulators that directly regulated the bottom layer factors and placed them in the second layer. The direct regulators of the second layer factors were then identified as the third layer TFs. This procedure was iterated until all factors were included in the hierarchical network. In total, the network model characterized 835 interactions; 686 were established by TFs mapped in this study (blue edges),125 were derived from previously published data (grey edges), and 24 were auto-regulatory. The network captures many known regulatory interactions, for example Eve regulates *ftz* and *prd* [45]. However, the vast majority of the 686 interactions among TFs from this study represent potential new regulatory relationships.

TFs that regulate one another and also co-regulate sets of target genes are often involved in complex biological processes that require feed forward loops such as observed with segmentation and mesoderm development TFs. To characterize the potential TF regulatory interactions that also involve co-regulation of common targets, we used the observed overlap of their target binding sites (red dashed edges in **Fig. 4a** connect TFs pairs with binding site overlap enrichment z-scores > 10). Examples include pair-rule genes such as Ubx, En, and Dll as well as Brm, Snr1, and Ubx. Overall, TFs with regulatory interactions from the network in 4a also have higher co-binding associations ($p<0.01$). Notably, particularly highly connected nodes in this network also shared a high proportion of target binding sites with connected nodes (Note, for example, the density of associated edges for Trl, Cad, Dl, and Sens).

Finally, to better understand how combinatorial TF binding regulates developmentally dynamic gene expression, we integrated gene expression data from our RNAseq time-course and an independently performed 64 time point expression microarray time course. RNAseq allows for very accurate quantification of gene expression levels, and thus clustering is largely driven by variations in *absolute* transcript levels [46]. A microarray timecourse, by contrast, measures *relative* levels of expression and therefore clustering is driven more by dynamic patterns. We partitioned the expression datasets into 18 and 64 k-means clusters, respectively, which resulted in gene sets with widely varying temporal specificity (**Fig. 4b, c**). For each cluster of genes, we then quantified the enrichment of promoter-proximal binding sites for 90 novel and previously published TF datasets. From the microarray timecourse clustering, five metaclusters were identified. Genes within these metaclusters are most highly expressed at third instar through adulthood (I), first instar through pupal-adult ecdysone pulse (II), early embryos (III), embryogenesis and larval life (IV) and late embryos (V). In both the microarray and RNAseq timecourses, all clusters are significantly associated with a core set of TFs including Sin3A, Ubx, Cad, Sens and Trl. All of these factors have relatively high numbers of binding sites (Table 1), and thus are likely to contribute to the expression of many genes, but this result also likely reflects the involvement of these factors in multiple stages and development processes since other TFs with similar numbers of binding sites are associated with only one or a few developmental expression patterns. For example, all metaclusters are enriched for Trl binding sites except V, which is enriched for Snr1, another Trithorax group gene; this result is consistent with reports that Snr1 has specialized functions [39,47]. Metacluster II is most highly expressed during adult central nervous system development [48] and enriched for several unique factors with known function in neuronal differentiation (Kr, Kni and Jumu) [49]. Metacluster III uniquely is associated with TFs known to have embryonic roles establishing pattern and organogenesis (e.g. Run, Hb, Twi). Notably, many of the co-enrichments within gene expression clusters correspond to co-enrichments (Figs. 3c and 4a), indicating that many of the co-associations of TFs with developmental expression patterns reflect co-binding and coordinate regulation at target sites in the genome.

In summary, these TF binding results defined HOT spots of increased TF complexity and their association with HDACs and open chromatin. Subsequent HOT region subtracted analysis of significantly co-bound TFs and TF networks greatly expands the existing view of potential regulatory interactions among TFs, and it associates specific sets of TFs with specific developmental gene expression patterns. While still a crude map of TF interactions and regulatory potentials, the newly mapped TFs in this study serve to annotate the *Drosophila* genome with over 35,000 unique TF binding regions. Overlaid on the dynamic chromatin map (Fig. 1, 2), they help to further delineate cis-regulatory elements and thus provide starting points for dissecting the expression for thousands of genes. However, clearly only a fraction, approximately 10%, of the total TFs encoded in the genome have been mapped by this and other studies (Table 1, [11,12,14]). These TF data, however, will provide a valuable comparison point for future studies.

Discussion

To produce a regulatory annotation map of the *Drosophila* genome, we generated 313 high-resolution genome-wide datasets that were released to public databases prior to publication. This work is intended to serve as a foundational resource for the *Drosophila* research community. We produced 65 ChIP grade antibodies or epitope-tagged TFs on transgenic BACs [50] and we generated 25 data sets on request using antibodies provided by the *Drosophila* community. As a result, the factors analyzed in the modENCODE Project represent a much more diverse collection of trans-regulators than was previously available. Additionally, while some factors analyzed here are well studied (e.g., Engrailed, Ubx, Eve) with hundreds of publications, several, such as Jumu, Chinmo and GATAe, are referenced in fewer than a dozen publications (**Supplementary Fig. 25**). Interestingly, some of these less studied factors are strongly implicated in regulatory relationships with the highly studied factors. For example, Jumu and Ubx show highly significant co-binding at target sites throughout the genome (Fig. 3c).

Although our results are based on a limited number of analysis methodologies, we have been able to identify or predict thousands of regulatory elements in this and a previous study, including 537 silencers, 2,307 newly annotated promoters, 12,285 candidate enhancers and 7,685 putative insulators ([17]; **Supplementary Tables 6-14**). There have also been several unexpected results from this initial phase of cis-regulatory mapping for the modENCODE Project. For example, we revealed a specific class of unmarked promoters, identified a surprising association of HDAC4a and HDAC1/rpd3 to PREs, and discovered pairs of TFs that systematically avoid binding near each other throughout the genome. Other observations serve as launchpoints for new investigations, such as the apparent swapping of HDAC3 and HDAC6 binding associations in human and *Drosophila*, the nature of H3K27me3 domain dynamics during development and hundreds of newly defined TF co-binding interactions at genomic targets.

As the datasets available from modENCODE grow and the algorithmic approaches for predicting different classes of regulatory elements are refined, we expect to continue to improve the annotations of cis-regulatory elements at higher resolution and with higher accuracy. Expanding the matrix of TF co-binding and co-regulatory patterns also will help to define the regulatory architecture of the genome. Researchers trying to identify the interactions or regulatory functions of their favorite factor can directly use the data and analyses presented in the cis-regulatory map presented here, and they can compare their own ChIP data to help provide context for their results.

Acknowledgements

The project described was supported by U01HG004264 from the National Human Genome Research Institute to KPW. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Human Genome Research Institute or the National Institutes of Health. CDB is supported by a Lilly-Life Sciences Research Foundation fellowship. CAB is supported by a NIH NRSA postdoctoral fellowship. RPA is in part supported by an Isaac Newton Trust award to RW. PL was supported by a grant from the Department of Energy Computational Sciences Graduate Fellowship (DOE CSGF). MEL and DMM work was supported by NHGRI grant U01 HG004279. We thank the FGF at the University of Chicago and the HGAC at Argonne National Laboratory for processing of microarrays and of Illumina sequence. We also thank the many members of the *Drosophila* community who generously contributed to this work by providing reagents. A complete list of community participants is included in the Supplementary Methods.

Figure Legends

Table 1: Summary of datasets produced.

Figure 1: Chromatin dynamics across Drosophila development. (A) Distribution of the number of genes marked (y-axis) by 6 histone modifications of chromatin modifying enzymes (colors), plotted against the number of developmental stages the gene is marked in (x-axis). (B) Pair-wise overlap enrichment between non-TF datasets (block bootstrap enrichment Z-score, from <-5 (blue) to >80 (red)). The RNAseq time course was used to segregate all transcripts into 4 quartiles by FPKM. All factors studied for the chromatin time-course project (marked with 't') have been ordered per factor by developmental stage. (C) Clustering H3K27Me3 domains by temporal dynamics. Domains (columns) are grouped into clusters based on the temporal pattern (y-axis) of Histone mark presence (blue) or absence (white) and are arranged along the x-axis; selected clusters are numbered at top. (D) Use of the classifier to annotate each gene promoter as marked or unmarked (see Methods), as in S. Figure 3. At a precision of 0.95, the number of unmarked active genes increases over time from 13% in embryos to 31% in adult males.

Figure 2: Annotation and prediction of promoters and enhancers with chromatin marks. (A) Prediction of novel promoters. Number of novel CAGE-validated promoter predictions (y-axis) per developmental stage are depicted in grey bars, cumulative total of unique CAGE-validated predictions in black dots. Distribution of H3K4Me3 (grey) and PolII (black) marks relative to gene TSSs depicted in inset. (B) Novel promoter prediction validation. Individual experiments, in triplicate, are represented as a single bar. Mean $\log_{10}$ transformed, normalized luciferase measurements from constructs (x-axis) with inserts in the forward (blue) and reverse (green) orientations (y-axis). Black lines depict standard error. The central portion of the graph depicts the validation of novel promoter predictions based on data from 0-12 hour embryos, while the right depicts validation of novel promoter prediction from Kc cell data. (C) Enrichment of CBP and H3K4me1 (rows) within regions marked by other chromatin modifications, factors, or annotated enhancers (columns). Note that (i) CBP is enriched within all active marks (H3K4me3, H3K27Ac, H3K9Ac, H3K4me1 and PolII) at all stages of development and (ii) early embryo (0-16h) CBP and H3K4me1 marked regions are enriched within H3K27me3 domains and annotated enhancers (right panel). (D) Heatmap depicting fold enrichment of CBP bound regions (columns) at different developmental stages for each of the 20 clusters of TSS-distal regions (rows) grouped by their protein binding profiles. A subset of the clusters shows significant enrichment for CBP at different developmental stages (all values greater or less than zero are significant, p-value < 0.001). (E) Enrichment of enhancer categories (columns) for each of the 20 clusters of TSS-distal regions (rows). Most of the clusters that have significant enrichment for CBP (C2,C3,C4,C5,C7,C9,C18) are also strongly enriched for enhancers. (F) Embryo-specific CBP binding predicts unannotated enhancers. RNA in situs with a Gal4 probe were used to stain embryos transfected with five different enhancer predictions (rows), at four to five different stages (columns). Bottom right panel depicts endogenous expression pattern (RNA *in situ*) of neighboring gene of EO044. (G) Enrichment of enhancer annotations (rows) within the binding sites of each transcription factor (columns). For panels A,C-E gray boxes indicate no overlap.

Figure 3: Transcription factor binding site complexity.

(A) Number of TFBS (left y-axis, black circles) and distribution of genomic annotation classes (right y-axis, colors) as a function of TFBS complexity (x-axis). (B) Heatmaps of TFBS enrichment (color scale, depleted in blue, enriched in red) of TFBS sorted by TF binding site complexity (y-axis) within annotated enhancers (CM: cardiac mesoderm, Ht: heart muscle, SM: somatic muscle; VM: visceral muscle.), HDAC binding sites, early embryo chromatin marks. At right is a heatmap depicting nucleosome density as a function of TFBS complexity.

Figure 4. Transcription factor interactions and associated gene expression patterns.

(A) Hierarchical transcriptional regulatory network defined by TFBS interactions between pairs of TFs in this and published data. Nodes (TFs) identified in this study in pink, those based on two previous studies in green and yellow. Previously identified edges (regulatory interactions) depicted in grey, those derived from this study in blue. Edges connecting factors whose binding sites significantly overlap (block bootstrap Z > 10) are depicted as red dashed lines. (B-C) Gene expression medoids (blue to red) for each of 64 and 18 k-means clusters (y-axis) derived from independent microarray (B) and RNAseq transcription time courses, at each developmental stage (x-axis, labeled by stage). Metaclusters (described in main text) are boxed and labeled in roman numerals.

Supplementary Figures Legends

Supplementary Text 1: Supplementary Methods**.** This document describes all reagents and materials as well as the algorithms and analytic tools used.

Supplementary Table 1. modENCODE datasets generated for this study. The different datasets produced for this study are listed here. They are ordered by groups of coherent production.
Supplementary Table 2. Chromatin time-course datasets. This table indicates for each dataset of the chromatin time-course the number of peaks and their median length in base pairs.
Supplementary Table 3. Promoter validation results. This Table is listing the coordinates of the novel promoters assayed for their activity. The coordinates of each fragment is indicated as well as the result for each orientation tested. "Validated" means that in two out of three independent experiments, the average of the triplicate transfections was greater than 2 standard deviations (SD) above the mean of the negative controls. "Supported" means that only one out of the three independent experiments had the average of the triplicates for that experiment greater than 2 SDs above the mean of the negative controls. "Unsupported" means that none of the experiments had the average of the triplicates greater than 2 SDs above the mean of the negative controls. "Incomplete" means that for that orientation all three experiments have not yet been performed.
Supplementary Table 4. Enhancer validation summary.
Supplementary Table 5. TF complexity percentages. This table indicates for each complexity category the total amount of genome covered, the number of TF associated to each category and the median length of the merged binding sites. It also indicates the number of transcripts associated to each binding region (+/- 1kb from an annotated TSS), their mean RPKM value and the number and percentage of active genes associated to each TF complexity category.
Supplementary Table 6. TSS class annotation at FDR 0.05
Supplementary Table 7. TSS class annotation at FDR 0.1
Supplementary Table 8. Novel promoter prediction based on co-occurence of H3K4me3, PolII and RNA in embryos.
Supplementary Table 9. Novel promoter prediction based on co-occurence of H3K4me3, PolII and RNA in Kc167 cells.
Supplementary Table 10. Insulators Class I.
Supplementary Table 11. Insulators Class II.
Supplementary Table 12. HDAC associated PREs.
Supplementary Table 13. CBP embryo only enhancer predictions.
Supplementary Table 14. CBP driven 20 clusters.
Supplementary Table 15. Insulator validation.

Supplemental Figure 1. Pair-wise overlap enrichment between datasets (block bootstrap enrichment Z-score, from <-5 (blue) to >80 (red)) generated by our group, the BDTNP, and regulatory element predictions. The RNAseq time course was used to segregate all transcripts into 4 quartiles by FPKM. All factors studied for the chromatin time-course project (marked with 't') have been ordered per factor by developmental stage.
Supplementary Figure 2. Example of the distributions of the 8 chromatin marks studied.
These profiles all correspond to ChIP-seq data from the pupal stage. Note the striking difference between the distributions of H3K9me3 and H3K27me3 (in blue) and all other marks. Conversely H3K4me3, H3K9Ac, H3K27Ac and H3K4me1 (purple) all exhibit an occupancy profile similar to that of PolII (red). CBP (green) is also correlated to the RNAseq coverage (orange).
Supplementary Figure 3. Morphology of the TF binding data.
This browser shows different binding site distributions for different factors. While some factors mainly bind narrow peaks (ex. Bab1 and Brm), others mainly bind large domains (ex. Dll and Gro) while still others bind a combination of both (ex. Bks and Chinmo).

Supplementary Figure 4. H3K9me3 defines heterochromatic regions. In our chromatin time-course experiments, H3K9me3 is largely overlapping with H3K27me3 domains. Using peptide competition assays followed by ChIP we were able to demonstrate that this overlap is resulting from an antibody cross-reactivity at this particular locus (data not shown). We detected the real H3K9me3 domains by comparison with HP1, a chromodomain protein recognizing specifically this Histone modification. H3K9me3 is located in large domains at the centromeric end of chromosomes 2L, 2R, 3L and along the chromosome 4.

Supplementary Figure 5. Domains of H3K27me3. (A) Genome browser example of the distribution of the repressive H3K27me3 mark along chromosome 2R? over developmental time starting with embryos (turquoise) and progressing to adults (red) . Most domains appear to be present at all time-points, but a substantial fraction show some stage specificity (starred examples). (B) Genes within H3K27me3 domains have lower RPKM values than genes outside the domains.

Supplementary Figure 6. GO categories analysis of H3K27me3 associated genes. (A) Summary of main enriched GO categories in the different clusters of H3K27me3 domains. (B) Example of GO terms overrepresented (red) or under represented (blue) in a stage specific cluster of H3K27me3 domains.

Supplementary Figure 7. Spatial restriction of H3K27me3 associated genes. (A) Example of Tomancak clusters of similar expression profile from in situ experiments enriched or depleted in H3K27me3 cluster 89. Enriched clusters are enriched for spatially restricted genes while the depleted clusters are enriched for ubiquitously expressed genes. (B,C) Examples of genes within H3K27me3 domains or immediately adjacent showing differences in spatial expression.

Supplementary Figure 8. Percentage of genes associated with each factor conditional upon gene expression status during the time-course. The union of Agilent and Solexa peaks has been used for each factor to assign genes as "marked" or "unmarked" depending to the presence of a specific factor or Histone mark within -1kb to +1kb of the TSS. Genes have also been classified as "active" or "inactive" according to their sequencing coverage in the RNAseq experiments. The distribution of "marked" and "unmarked" genes is represented for (A) H3K4me3 (as in Fig. 1b), (B) H3K4me1, (C) PolII, (D) H3K9Ac, (E) H3K27Ac and (F) CBP. On each graph, for each time point, the genes in red are active and the genes in blue are inactive. The genes in dark color are "marked" while the genes in light color are "unmarked". The red line separates the active genes from the inactive genes at all stages.

Supplementary Figure 9. Building a classifier of gene expression from chromatin marks. (A) Distribution of FPKM estimates for all12 developmental stages. (B) AUC values across a range of FPKM thresholds for models trained on each developmental stage separately. (C) Recovery of marked active genes across a range of FPKM thresholds for models trained on each developmental stage separately (FDR= 0.10). (D) ROC curves for the logistic regression classifier across multiple FPKM values for 12 developmental stages. Line colors correspond to different FPKM thresholds: red = 0.1, green = 0.5, blue = 1.0, cyan = 1.5, magenta = 2.0, black = 2.5.

Supplementary Figure 10. The classifier of gene expression detects an umarked active gene category. (A) Binary classifier outcome transcript distribution. Outcomes defined at FDR = 0.10. MA = marked active, MI = marked inactive, UA = unmarked active, UI = marked inactive. (B) Distribution of FPKM values for binary classier outcomes. **a,** MA vs. MI at FDR = 0.05. **b,** UA vs. UI at FDR = 0.05. **c,** MA vs MI at FDR = 0.10. **d,** UA vs UI at FDR = 0.10.

Supplementary Figure 11. Unmarked active genes have temporally restricted expression patterns. (A) Enrichment of FlyAtlas spatial expression terms for the unmarked active and marked active genes in the Adult male (a similar pattern is observed with adult female). Note that the marked active class is more enriched in tissue specific terms. (B) Predictability of active and inactive transcripts. **a-b,** Predictability of active transcripts is defined as the number of times a transcript is classified as marked (FDR = 0.05 (a), FDR = 0.10 (b)) normalized by the number of stages the transcript is active. **c-d,** Predictability of inactive transcripts is defined as the number of times a transcript is classified as unmarked (FDR = 0.05 (c), FDR = 0.10 (s)) normalized by the number of stages the transcript is inactive.

Supplementary Figure 12. Examples of active genes not associated to H3K4me3. This genome browser view shows the occupancy profile of H3K4me3 (purple) and the RNAseq coverage (orange) around the Trypsin gene complex on the chromosome 2R. Genes associated to H3K4me3 are highlighted. We can observe that they are all expressed at all stages investigated. The Trypsin genes however, as well as the gene *sha* and *nompA* are transiently expressed and do not have H3K4me3 at their promoters.

Supplemental Figure 13. H3K4me3 unmarked, detected genes in Kc cells. (A-B) Seven representative examples of unmarked active genes observed in embryos and synchronized cell culture. (C) qPCR validation of unmarked active genes from synchronized cell culture.

Supplementary Figure 14. HDACs are associated with TSSs, transcribed exons, and PREs. (A-B, D-E) Enrichment of HDAC binding sites (y-axis) around active (A; FPKM > 1) and inactive (D; FPKM < 1) metagenes (x-axis corresponding to 2000 bp upstream and 1000 bp downstream of the TSS and 1000bp upstream and 2000 bp downstream of the TES of genes). Each of five different HDACs is plotted as a seperate color, as labeled. FPKM estimates were derived from pooled RNAseq data from stages E0-4h, E4-8h and E8-12h. HDAC1, 4a, 6 and 11 show a strong enrichment at the TSS and depletion along the gene body. In contrast, HDAC3 shows a strong depletion at the TSS and an enrichment along the gene body. (B, E) Enrichment of 4 different Histone tri-methylations (y-axis) across active (B) and inactive (E) metagenes (x-axis). Note the striking similarity between (i) H3K4me3 and HDAC 1, 4a, 6 and 11 profiles and (ii) H3K36me3 and HDAC3 profiles. In contrast, genes defined as inactive have reduced enrichment of HDACs at the promoter and a depletion along the gene body. Similar differences are also observed for the corresponding Histone tri-methylations. (C) HDAC enrichment (y-axis) is correlated with target gene expression level (x-axis). (F) HDAC enrichment (y-axis) at varying distances from PHO sites (x-axis). HDAC4a and 1 are strongly enriched in the proximity of PHO sites, while HDAC6 and 11 show a moderate enrichment and HDAC3 a strong depletion.

Supplementary Figure 15. Prediction of silencers. Flowchart of silencer prediction from HDAC1 and HDAC4a binding site data. The union of HDAC1 and HDAC4a binding sites (n=6191) was filtered for sites overlapping H3K4me3. Sites within the remaining 2521 sites that overlapped regions of H3K27me3 to predict 537 PREs.

Supplementary Figure 16. Examples of silencers. This IGB browser example is centered around the homeotic gene cluster ANT-C. The PC and PHO data are from [8]. Common binding regions for HDAC1 and HDAC4a are associated with either H3K4me3, GAF or PCL/PHO Binding Regions representing Polycomb Response Elements (blue squares).

Supplementary Figure 17. CBP and H3K4me1 are associated with enhancers.

This IGB browser example represents signal for CBP (green) and H3K4me1(pink) at three different time-points (Adult Male, Pupae and Embryos 0-4h) around the region of *even-skipped* that contains well characterized enhancers (represented by the REDFly track in purple). Also represented are the insulators, the blue dashed line representing Class I gene boundaries. In embryos, the several enhancers within the intergenic region around *eve* are bound by CBP and contain H3K4me1 signal. Note that both signals are not present later during development when these enhancers are not active.

Supplementary Figure 18. Clustering CBP bound regions. (A) Illustration of criteria used to associate experiments with CBP regions. (B) Bayesian information criterion score vs cluster number used in model training. (C) Enrichment of chromatin and PolII profiles within each CBP cluster. The rows of the enrichment map correspond to the CBP clusters 1-20, where the number of regions is indicated in the row label. Columns of the enrichment map correspond to chromatin time course experimental data. Each cell represents the enrichment (red) or depletion (blue) of each experimental binding site set within the binding sites of each CBP cluster.

Supplementary Figure 19. "CBP embryo only" enhancer validation examples. (A) CBP Chip-seq data, across the developmental time course, for genomic regions corresponding to enhancer predictions that were tested in Fig. 4c. (B) Additional examples of tested regions for which reporter expression overlaps aspects of available RNA *in situ* patterns for neighboring genes (known gene expression data from FlyExpress).

Supplementary Figure 20. Insulator validation. (A) Diagram of the insulator validation strategy; A recipient *P* element integrated at 3R:13373664 containing the *even skipped* stripe 2 and 3 enhancer elements separated by an eye-expressed eGFP is used as a substrate for Recombination Mediated Cassette Exchange, replacing the eGFP with a genomic DNA fragment. (B – U) Immunohistochemistry with an anti b-Galactosidase antibody to detect reporter expression. All stage 10/11 embryos are oriented anterior to the left dorsal to the top. (B – E) Control lines: (B) recipient construct showing strong expression in *eve* stripe 2 and 3 territories, with weaker expression in stripe 7 and cephalic territories. (C-D) The characterised *1A2* and *scs* insulators block stripe 3 expression. (E) A negative control spacer fragment from the *eve* locus shows no enhancer blocking activity. (F – N) Class I elements generally show enhancer blocking activity. Each fragment is associate with CTCF and CP190 binding. Strong: (F) 2894, (G) 11628, (H) 4762, (I) 7635, (J) 9220, (K &L) 11742 shows variable activity with some embryos showing strong enhancer blocking (K) and others weak (L). (M & N) 8562 (M) and 11319 (N) are class I elements that show weak enhancer blocking activity. (O & P) Two GAF positive regions show no enhancer blocking activity. (O) Antp1, (P) fab4. (Q – U) Class II elements that bind Su(hw) and CP190 show little or no enhancer blocking activity. (Q) 12404, (R) 8767, (S) 3557, (T) 2738, (U) 4432. See supplementary table 15 for full details of the assayed fragments

Supplementary Figure 21. TF clustering, including HOT spots. Pair-wise enrichment for all transcription factor combinations, including TFBS overlapping HOT regions

Supplementary Figure 22. Hotspot distributions. (A) Distribution of HOT regions (in red) over the genome in relation to GC content (gray scale). (B) Distribution of HOT regions (in red) over the genome in relation to gene density (gray scale). (C) The fraction of HOT regions that overlap with five classes of genomic annotation (5' UTR (dark blue), coding exon (orange), intron (purple), 3'UTR (green), and intergenic(ligght blue)), for each set of merged transcription factor binding sites, binned by complexity (x-axis). From Category 1 to 8, the proportion of intergenic and TSS regions covered increases at the expense of CDS and intron categories. (D) Heatmap depicting the –log transformed Fisher's exact test p-value quantifying the pairwise enrichment between each TF binding site set and merged binding site complexity categories.

Supplementary Figure 23. TFBS interaction vignette.

Supplementary Figure 24. Networks constructed exclusively from Furlong et al. data (A) and BDTNP data (B).

Supplementary Figure 25. Number of references found for each protein that we have studied in either PubMed (blue line) or FlyBase (red line).

References

[1] **Stark, A. et al., Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* 450 (7167), 219 (2007).**

[2] **Ren, B. et al., Genome-wide location and function of DNA binding proteins. *Science* 290 (5500), 2306 (2000); Johnson, L. A., Zhao, Y., Golden, K., and Barolo, S., Reverse-engineering a transcriptional enhancer: A case study in Drosophila. *Tissue Engineering Part A* 14 (9), 1549 (2008); Iyer, V. R. et al., Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409 (6819), 533 (2001).**

[3] **Heintzman, N. D. et al., Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459 (7243), 108 (2009).**

[4] **Mikkelsen, T. S. et al., Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448 (7153), 553 (2007).**

[5] **Celniker, S. E. et al., Unlocking the secrets of the genome. *Nature* 459 (7249), 927 (2009).**

[6] **Schuettengruber, B. et al., Genome regulation by polycomb and trithorax proteins. *Cell* 128 (4), 735 (2007).**

[7] **Schwartz, Y. B. et al., Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet* 6 (1), e1000805.**

[8] **Kwong, C. et al., Stability and dynamics of polycomb target sites in Drosophila development. *PLoS Genet* 4 (9), e1000178 (2008).**

[9] **Mito, Y., Henikoff, J. G., and Henikoff, S., Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37 (10), 1090 (2005); van Steensel, B., Delrow, J., and Henikoff, S., Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* 27 (3), 304 (2001).**

[10] **Li, X. Y. et al., Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* 6 (2), e27 (2008).**

[11] **MacArthur, S. et al., 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 (7), R80 (2009).**

[12] **Zeitlinger, J. et al., Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. *Genes Dev* 21 (4), 385 (2007).**

[13] **Zeitlinger, J. et al., RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* 39 (12), 1512 (2007).**

[14] **Zinzen, R. P. et al., Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462 (7269), 65 (2009).**

[15] **Georlette, D. et al., Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. *Genes Dev* 21 (22), 2880 (2007).**

[16] **Holohan, E. E. et al., CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. *PLoS Genet* 3 (7), e112 (2007); Bushey, A. M., Ramos, E., and Corces, V. G., Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev* 23 (11), 1338 (2009); Adryan, B. et al., Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila. *Genome Biol* 8 (8), R167 (2007).**

[17] **Negre, N. et al., A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* 6 (1), e1000814.**

[18] Arbeitman, M. N. et al., Gene expression during the life cycle of Drosophila melanogaster. *Science* 297 (5590), 2270 (2002); Graveley, B. R. et al., The Developmental Transcriptome of Drosophila melanogaster. *Nature* (in press).

[19] Schotta, G. et al., Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* 21 (5), 1121 (2002).

[20] Smith, E. R. et al., Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. *Mol Cell Biol* 28 (3), 1041 (2008).

[21] Agger, K. et al., UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449 (7163), 731 (2007).

[22] Rugg-Gunn, P. J., Cox, B. J., Ralston, A., and Rossant, J., Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proc Natl Acad Sci U S A* 107 (24), 10783.

[23] Soshnikova, N. and Duboule, D., Epigenetic regulation of vertebrate Hox genes: a dynamic equilibrium. *Epigenetics* 4 (8), 537 (2009).

[24] Schwartz, Y. B. and Pirrotta, V., Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* 8 (1), 9 (2007).

[25] Brookes, E. and Pombo, A., Modifications of RNA polymerase II are pivotal in regulating gene expression states. *EMBO Rep* 10 (11), 1213 (2009); Campos, E. I. and Reinberg, D., Histones: annotating chromatin. *Annu Rev Genet* 43, 559 (2009); Schuettengruber, B. and Cavalli, G., Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* 136 (21), 3531 (2009).

[26] Barrera, L. O. et al., Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* 18 (1), 46 (2008).

[27] Heintzman, N. D. et al., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39 (3), 311 (2007).

[28] Cao, R. et al., Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298 (5595), 1039 (2002); Muller, J. et al., Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111 (2), 197 (2002); Kuzmichev, A. et al., Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16 (22), 2893 (2002); Czermin, B. et al., Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111 (2), 185 (2002); Wang, Z. et al., Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40 (7), 897 (2008); Poux, S., Horard, B., Sigrist, C. J., and Pirrotta, V., The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development* 129 (10), 2483 (2002).

[29] Visel, A. et al., ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457 (7231), 854 (2009).

[30] Barski, A. et al., High-resolution profiling of histone methylations in the human genome. *Cell* 129 (4), 823 (2007).

[31] Kolasinska-Zwierz, P. et al., Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* 41 (3), 376 (2009).

[32] Wang, Z. et al., Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138 (5), 1019 (2009).

[33] Gregoretti, I. V., Lee, Y. M., and Goodson, H. V., Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J Mol Biol* 338 (1), 17 (2004).

[34] Hoskins, R.A. et al., Genome wide analysis of promoter architechure in *Drosophila*. *Genome Res* (in press).

[35] Pfeiffer, B. D. et al., Refinement of Tools for Targeted Gene Expression in Drosophila. *Genetics*.

[36] Moorman, C. et al., Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 103 (32), 12027 (2006).

[37] Deal, R. B. and Henikoff, S., A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev Cell* 18 (6), 1030.

[38] Jimenez, G., Paroush, Z., and Ish-Horowicz, D., Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev* 11 (22), 3072 (1997).

[39] Zraly, C. B. et al., SNR1 is an essential subunit in a subset of Drosophila brm complexes, targeting specific functions during development. *Dev Biol* 253 (2), 291 (2003).

[40] Gebelein, B., McKay, D. J., and Mann, R. S., Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature* 431 (7009), 653 (2004).

[41] Nolo, R., Abbott, L. A., and Bellen, H. J., Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. *Cell* 102 (3), 349 (2000).

[42] Haecker, A. et al., Drosophila brakeless interacts with atrophin and is required for tailless-mediated transcriptional repression in early embryos. *PLoS Biol* 5 (6), e145 (2007).

[43] Rivera-Pomar, R. et al., Activation of posterior gap gene expression in the Drosophila blastoderm. *Nature* 376 (6537), 253 (1995).

[44] Senti, K., Keleman, K., Eisenhaber, F., and Dickson, B. J., brakeless is required for lamina targeting of R1-R6 axons in the Drosophila visual system. *Development* 127 (11), 2291 (2000).

[45] Carroll, S. B. and Vavra, S. H., The zygotic control of Drosophila pair-rule gene expression. II. Spatial repression by gap and pair-rule gene products. *Development* 107 (3), 673 (1989); Manoukian, A. S. and Krause, H. M., Concentration-dependent activities of the even-skipped protein in Drosophila embryos. *Genes Dev* 6 (9), 1740 (1992).

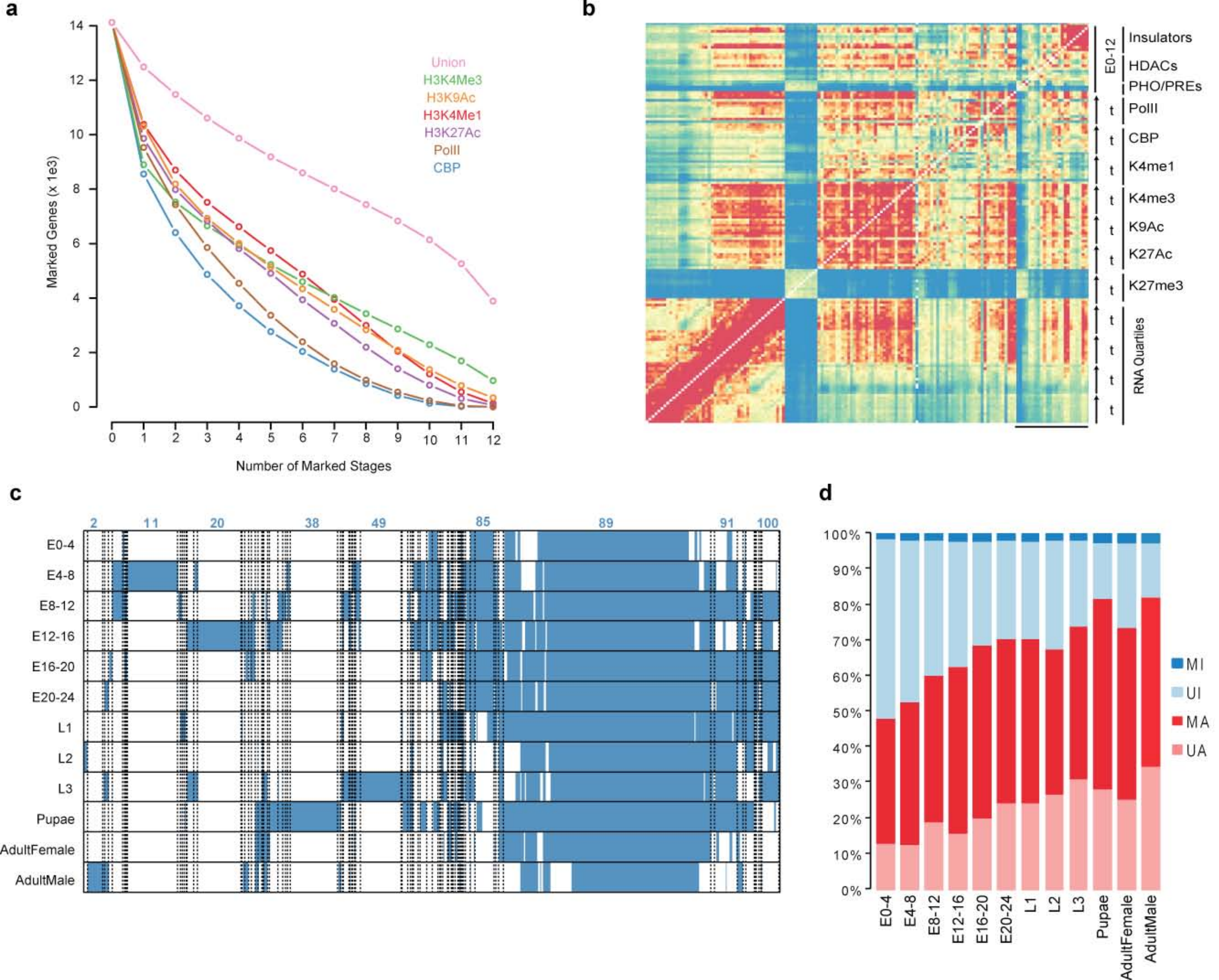
[46] **Marioni, J. C. et al., RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18 (9), 1509 (2008).**

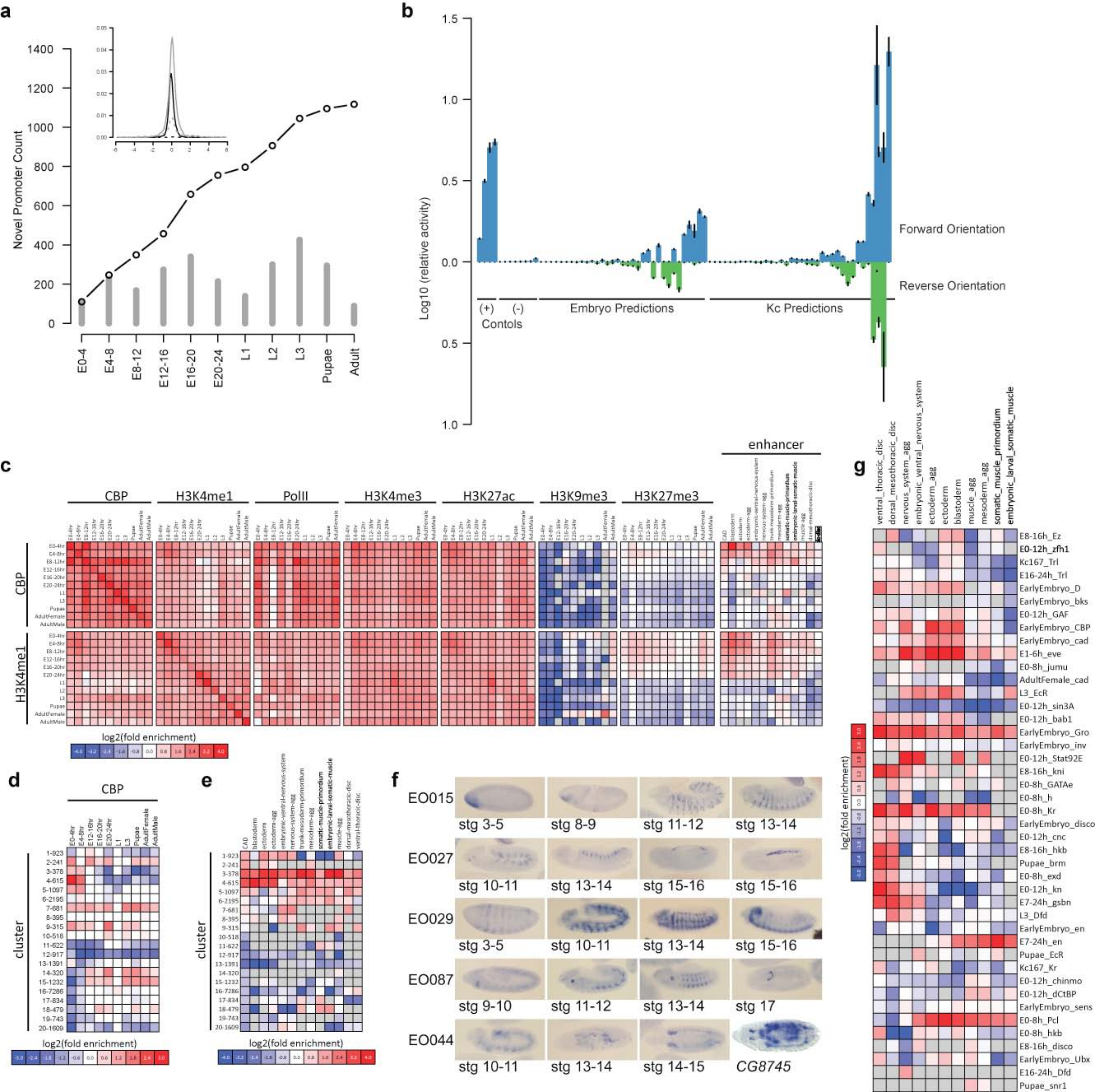
[47] **Marenda, D. R., Zraly, C. B., and Dingwall, A. K., The Drosophila Brahma (SWI/SNF) chromatin remodeling complex exhibits cell-type specific activation and repression functions. *Dev Biol* 267 (2), 279 (2004).**

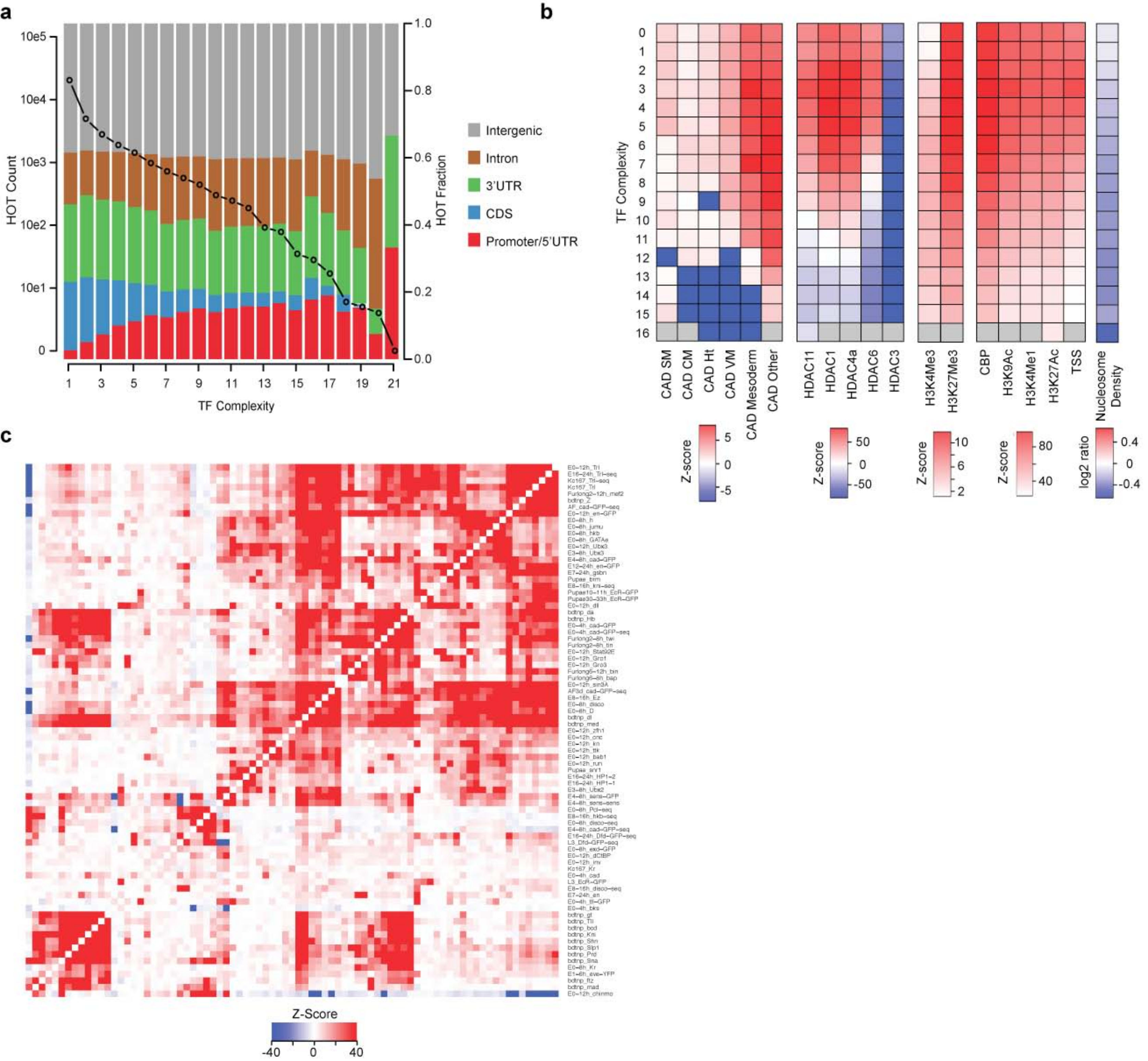
[48] **Truman, J. W., Metamorphosis of the central nervous system of Drosophila. *J Neurobiol* 21 (7), 1072 (1990).**

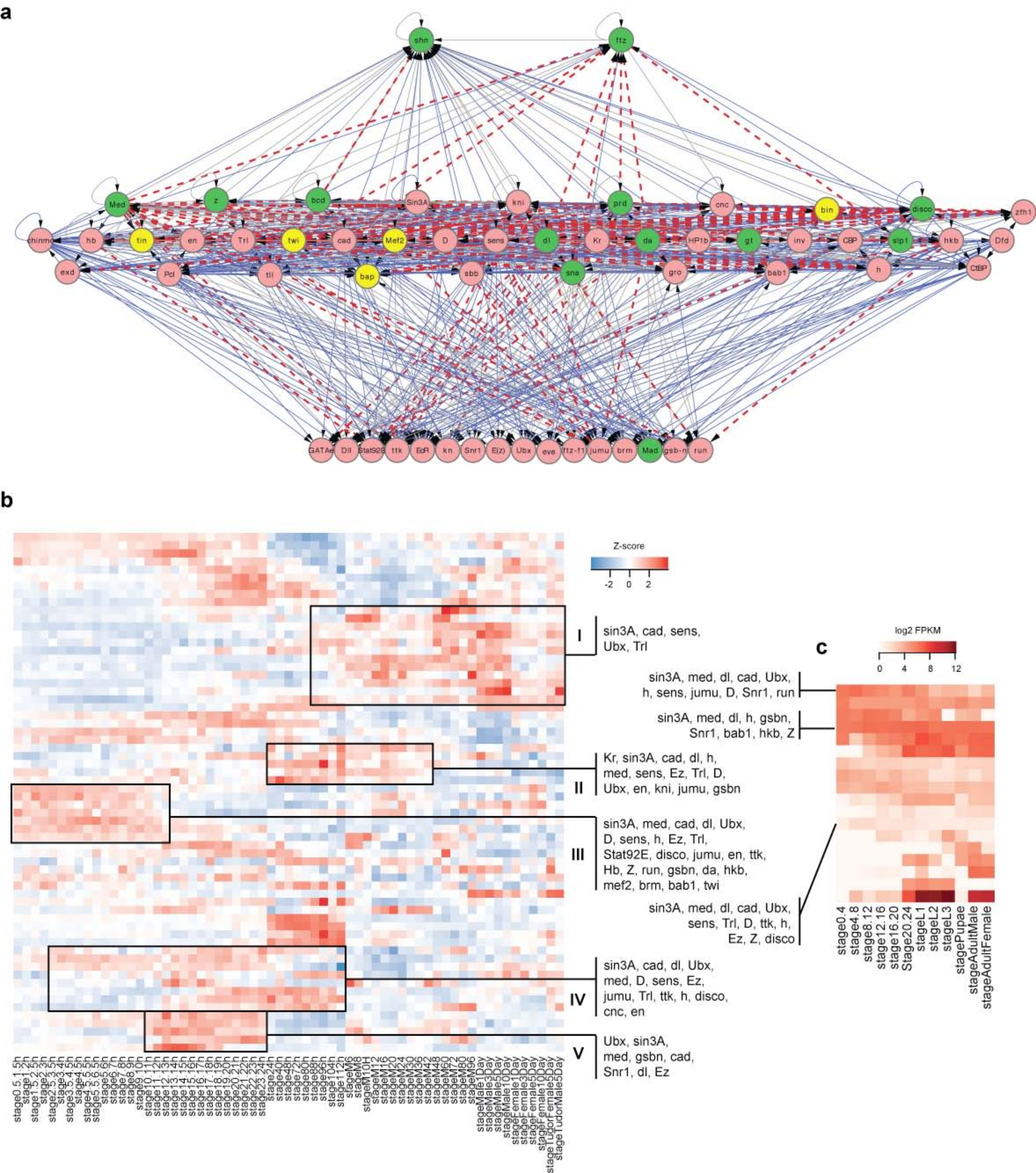
[49] **Brody, T. and Odenwald, W. F., Cellular diversity in the developing nervous system: a temporal view from Drosophila. *Development* 129 (16), 3763 (2002); Cheah, P. Y., Chia, W., and Yang, X., Jumeaux, a novel Drosophila winged-helix family protein, is required for generating asymmetric sibling neuronal cell fates. *Development* 127 (15), 3325 (2000); Parrish, J. Z., Kim, M. D., Jan, L. Y., and Jan, Y. N., Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes Dev* 20 (7), 820 (2006).**

[50] **Venken, K. J. et al., Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nat Methods* 6 (6), 431 (2009).**


a
Union
H3K4Me3
H3K9Ac
H3K4Me1
H3K27Ac
PolII
CBP
Marked Genes (x 1e3)
0
2
4
6
8
10
12
14
Number of Marked Stages
0 1 2 3 4 5 6 7 8 9 10 11 12
b
E0-12
Insulators
HDACs
PHO/PREs
PolII
CBP
K4me1
K4me3
K9Ac
K27Ac
K27me3
RNA Quartiles
t
c
2 11 20 38 49 85 89 91 100
E0-4
E4-8
E8-12
E12-16
E16-20
E20-24
L1
L2
L3
Pupae
AdultFemale
AdultMale
d
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
MI
UI
MA
UA
E0-4
E4-8
E8-12
E12-16
E16-20
E20-24
L1
L2
L3
Pupae
AdultFemale
AdultMale


a
Novel Promoter Count
1400
1200
1000
800
600
400
200
0
E0-4
E4-8
E8-12
E12-16
E16-20
E20-24
L1
L2
L3
Pupae
Adult
b
Log10 (relative activity)
1.5
1.0
0.5
0.0
0.5
1.0
Forward Orientation
Reverse Orientation
(+)
(-)
Contols
Embryo Predictions
Kc Predictions
c
CBP
H3K4me1
PolII
H3K4me3
H3K27ac
H3K9me3
H3K27me3
enhancer
E0-4hr
E4-8hr
E8-12hr
E12-16hr
E16-20hr
E20-24hr
L1
L2
L3
Pupae
AdultFemale
AdultMale
CAD
blastoderm
ectoderm
ectoderm-agg
embryonic-ventral-nervous-system
nervous-system-agg
trunk-mesoderm-primordium
mesoderm-agg
somatic-muscle-primordium
embryonic-larval-somatic-muscle
muscle-agg
dorsal-mesothoracic-disc
ventral-thoracic-disc
log2(fold enrichment)
-4.0
-3.2
-2.4
-1.6
-0.8
0.0
0.8
1.6
2.4
3.2
4.0
d
CBP
cluster
1-923
2-241
3-378
4-615
5-1097
6-2195
7-681
8-395
9-315
10-518
11-622
12-917
13-1391
14-320
15-1232
16-7286
17-834
18-479
19-743
20-1609
log2(fold enrichment)
-3.0
-2.4
-1.8
-1.2
-0.6
0.0
0.6
1.2
1.8
2.4
3.0
e
cluster
log2(fold enrichment)
f
EO015
stg 3-5
stg 8-9
stg 11-12
stg 13-14
EO027
stg 10-11
stg 13-14
stg 15-16
stg 15-16
EO029
stg 3-5
stg 10-11
stg 13-14
stg 15-16
EO087
stg 9-10
stg 11-12
stg 13-14
stg 17
EO044
stg 10-11
stg 13-14
stg 14-15
CG8745
g
ventral_thoracic_disc
dorsal_mesothoracic_disc
nervous_system_agg
embryonic_ventral_nervous_system
ectoderm_agg
ectoderm
blastoderm
muscle_agg
mesoderm_agg
somatic_muscle_primordium
embryonic_larval_somatic_muscle
E8-16h_Ez
E0-12h_zfh1
Kc167_Trl
E16-24h_Trl
EarlyEmbryo_D
EarlyEmbryo_bks
E0-12h_GAF
EarlyEmbryo_CBP
EarlyEmbryo_cad
E1-6h_eve
E0-8h_jumu
AdultFemale_cad
L3_EcR
E0-12h_sin3A
E0-12h_bab1
EarlyEmbryo_Gro
EarlyEmbryo_inv
E0-12h_Stat92E
E8-16h_kni
E0-8h_GATAe
E0-8h_h
E0-8h_Kr
EarlyEmbryo_disco
E0-12h_cnc
E8-16h_hkb
Pupae_brm
E0-8h_exd
E0-12h_kn
E7-24h_gsbn
L3_Dfd
EarlyEmbryo_en
E7-24h_en
Pupae_EcR
Kc167_Kr
E0-12h_chinmo
E0-12h_dCtBP
EarlyEmbryo_sens
E0-8h_Pcl
E0-8h_hkb
E8-16h_disco
EarlyEmbryo_Ubx
E16-24h_Dfd
Pupae_snr1
log2(fold enrichment)
3.0
2.4
1.8
1.2
0.6
0.0
-0.6
-1.2
-1.8
-2.4
-3.0


a
10e5
10e4
10e3
10e2
10e1
0
HOT Count
1.0
0.8
0.6
0.4
0.2
0.0
HOT Fraction
1
3
5
7
9
11
13
15
17
19
21
TF Complexity
Intergenic
Intron
3'UTR
CDS
Promoter/5'UTR
b
TF Complexity
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
CAD SM
CAD CM
CAD Ht
CAD VM
CAD Mesoderm
CAD Other
HDAC11
HDAC1
HDAC4a
HDAC6
HDAC3
H3K4Me3
H3K27Me3
CBP
H3K9Ac
H3K4Me1
H3K27Ac
TSS
Nucleosome Density
Z-score
5
0
-5
Z-score
50
0
-50
Z-score
10
6
2
Z-score
80
40
log2 ratio
0.4
0
-0.4
c
E0–12h_Trl
E16–24h_Trl–seq
Kc167_Trl–seq
Kc167_Trl
Furlong2–12h_mef2
bdtnp_Z
AF_cad–GFP–seq
E0–12h_en–GFP
E0–8h_h
E0–8h_jumu
E0–8h_hkb
E0–8h_GATAe
E0–12h_Ubx3
E3–8h_Ubx3
E4–8h_cad–GFP
E12–24h_en–GFP
E7–24h_gsbn
Pupae_brm
E8–16h_kni–seq
Pupae10–11h_EcR–GFP
Pupae30–33h_EcR–GFP
E0–12h_dll
bdtnp_da
bdtnp_Hb
E0–4h_cad–GFP
E0–4h_cad–GFP–seq
Furlong2–8h_twi
Furlong2–8h_tin
E0–12h_Stat92E
E0–12h_Gro1
E0–12h_Gro3
Furlong5–12h_bin
Furlong6–8h_bap
E0–12h_sin3A
AF3d_cad–GFP–seq
E8–16h_Ez
E0–8h_disco
E0–8h_D
bdtnp_dl
bdtnp_med
E0–12h_zfh1
E0–12h_cnc
E0–12h_kn
E0–12h_ttk
E0–12h_bab1
E0–12h_run
Pupae_snr1
E16–24h_HP1–2
E16–24h_HP1–1
E3–8h_Ubx2
E4–8h_sens–GFP
E4–8h_sens–sens
E0–8h_Pcl–seq
E8–16h_hkb–seq
E0–8h_disco–seq
E4–8h_cad–GFP–seq
E16–24h_Dfd–GFP–seq
L3_Dfd–GFP–seq
E0–8h_exd–GFP
E0–12h_dCtBP
E0–12h_inv
Kc167_Kr
E0–4h_cad
L3_EcR–GFP
E8–16h_disco–seq
E7–24h_en
E0–4h_tll–GFP
E0–4h_bks
bdtnp_gt
bdtnp_Tll
bdtnp_bcd
bdtnp_Kni
bdtnp_Shn
bdtnp_Slp1
bdtnp_Prd
bdtnp_Sna
E0–8h_Kr
E1–6h_eve–YFP
bdtnp_ftz
bdtnp_mad
E0–12h_chinmo
Z-Score
-40
0
40


a
shn
ftz
Med
z
bcd
Sin3A
kni
prd
cnc
bin
disco
zfh1
chinmo
hb
tin
en
Trl
twi
cad
Mef2
D
sens
dl
Kr
da
HP1b
gt
inv
CBP
slp1
hkb
Dfd
exd
Pcl
tll
bap
sbb
sna
gro
bab1
h
CtBP
GATAe
Dll
Stat92E
ttk
EcR
kn
Snr1
E(z)
Ubx
eve
ftz-f1
jumu
brm
Mad
gsb-n
run
b
Z-score
-2 0 2
I
sin3A, cad, sens, Ubx, Trl
II
Kr, sin3A, cad, dl, h, med, sens, Ez, Trl, D, Ubx, en, kni, jumu, gsbn
III
sin3A, med, cad, dl, Ubx, D, sens, h, Ez, Trl, Stat92E, disco, jumu, en, ttk, Hb, Z, run, gsbn, da, hkb, mef2, brm, bab1, twi
IV
sin3A, cad, dl, Ubx, med, D, sens, Ez, jumu, Trl, ttk, h, disco, cnc, en
V
Ubx, sin3A, med, gsbn, cad, Snr1, dl, Ez
stage0.5.1.5h
stage1.2h
stage1.5.2.5h
stage2.3h
stage2.5.3.5h
stage3.4h
stage3.5.4.5h
stage4.5h
stage4.5.5.5h
stage5.6h
stage6.7h
stage7.8h
stage8.9h
stage9.10h
stage10.11h
stage11.12h
stage12.13h
stage13.14h
stage14.15h
stage15.16h
stage16.17h
stage17.18h
stage18.19h
stage19.20h
stage20.21h
stage21.22h
stage22.23h
stage23.24h
stage24h
stage40h
stage48h
stage72h
stage80h
stage88h
stage96h
stage104h
stage112h
stageM6
stageM8
stageM10H
stageM12
stageM16
stageM20
stageM24
stageM30
stageM36
stageM42
stageM48
stageM60
stageM72
stageM80
stageM96
stageMale1Day
stageMale3Day
stageMale5Day
stageMale10Day
stageFemale1Day
stageFemale3Day
stageFemale5Day
stageFemale10Day
stageTudorFemale5Day
stageTudorMale5Day
c
log2 FPKM
0 4 8 12
sin3A, med, dl, cad, Ubx, h, sens, jumu, D, Snr1, run
sin3A, med, dl, h, gsbn, Snr1, bab1, hkb, Z
sin3A, med, dl, cad, Ubx, sens, Trl, D, ttk, h, Ez, Z, disco
stage0.4
stage4.8
stage8.12
stage12.16
stage16.20
Stage20.24
stageL1
stageL2
stageL3
stagePupae
stageAdultMale
stageAdultFemale

A

BS in Stage Factor	E0-4	E4-8	E8-12	E12-16	E16-20	E20-24	L1	L2	L3	Pupae	AdultFemale	AdultMale	Unique # of BS Of Each
CBP	5,072	8,876	329	9,026	6,027	6,042	1,273	-	3,843	8,508	7,735	6,594	17,463
PolII	4,482	6,138	1,333	6,208	12,103	2,413	3,024	5,285	7,143	6,238	2,400	-	15,644
H3K4Me1	8,810	7,403	8,827	14,849	12,610	12,471	3,362	11,290	10,488	10,759	7,669	3,610	21,726
H3K4Me3	7,607	4,160	5,816	5,667	8,558	10,429	6,500	5,743	5,548	5,347	3,841	7,328	11,288
H3K9Ac	6,914	5,243	7,585	8,555	8,603	14,604	4,838	6,907	5,975	5,703	7,512	8,899	21,480
H3K9Me3	974	807	1,440	700	495	449	470	501	315	425	40	466	1,380
H3K27Ac	6,066	5,654	7,720	9,886	12,366	8,674	2,321	6,061	4,774	6,862	7,678	9,725	17,406
H3K27Me3	810	908	1,174	813	422	493	522	428	2,590	441	207	2,202	2,513

B

Factor	Stage / Cell Type	Ab	Platform	NumOfTFBS	Peak Feature
bab1	E0-12h	a-Bab1-SC	Affymetrix v2.0	1,227	punctate
bks	E0-4h	Affymetrix v2.0	a-bks-MM	1,706	broad
brm	Pupae	Affymetrix v2.0	a-brm-AD	270	punctate
cad	E4-8h	Affymetrix v2.0	KWG-GFP	1,070	punctate
cad	AF(3d)	Solexa	KWG-GFP	3,457	punctate
cad	E0-4h	Affymetrix v2.0	a-cad55-JR	724	punctate
cad	E0-4h	Solexa	KWG-GFP	2,207	punctate
cad	E4-8h	Solexa	KWG-GFP	5,626	punctate
cad	AF	Solexa	KWG-GFP	7,579	punctate
cad	E0-4h	Affymetrix v2.0	KWG-GFP	1,700	punctate
chinmo	E0-12h	Affymetrix v2.0	a-chinmo-EB	7,054	broad
cnc	E0-12h	Affymetrix v2.0	KW0-CNC	699	punctate
CtBp	E0-12h	Affymetrix v2.0	KW0-dCtBP766	4,947	punctate
D	E0-8h	Affymetrix v2.0	KW3-D-D2	2,979	punctate
Dfd	E16-24h	Affymetrix v2.0	KW0-dCtBP766	581	punctate
Dfd	L3	Solexa	KWG-GFP	3,159	punctate
disco	E0-8h	Solexa	KW3-disco-D2	1,723	punctate
disco	E8-16h	Solexa	KW3-disco-D2	616	punctate
disco	E0-8h	Affymetrix v2.0	KW3-disco-D2	2,628	punctate
Dll	E0-12h	Affymetrix v2.0	a-dll-SC	75	broad
E(z)	E8-16h	Agilent 1M	KW4-E(z)-D2	1,927	punctate
EcR	Pupae	Solexa	KWG-GFP	483	mixed
EcR	L3	Solexa	KWG-GFP	603	mixed
EcR	Pupae	Solexa	KWG-GFP	508	mixed
en	E0-12h	Affymetrix v2.0	KWG-GFP	3,568	punctate
en	E7-24h	Affymetrix v2.0	a-end300	286	punctate
en	E12-24h	Affymetrix v2.0	KWG-GFP	1,502	punctate
eve	E1-6h	Agilent 1M	KWG-GFP	1,738	punctate
exd	E0-8h	Affymetrix v2.0	KWG-GFP	4,483	punctate
GATAe	E0-8h	Affymetrix v2.0	KW4-GATAe-D1	901	punctate
Gro	E0-12h	Affymetrix v2.0	KW0-GRO	626	broad
Gro	E0-12h	Affymetrix v2.0	KW0-GRO	1,338	broad
gsb-n	E7-24h	Affymetrix v2.0	a-gsbn-FM	765	punctate
h	E0-8h	Affymetrix v2.0	KW3-h-D1	1,944	punctate
hkb	E8-16h	Solexa	KW3-hkb-D1	1,623	punctate
hkb	E0-8h	Affymetrix v2.0	KW3-hkb-D1	1,269	punctate
HP1b	E16-24h	Affymetrix v2.0	a-HP1-Covance	3,396	mixed
HP1b	E16-24h	Affymetrix v2.0	a-HP1-Abcam	6,967	mixed
inv	E0-12h	Affymetrix v2.0	KW0-INV7657	3,222	mixed
jumu	E0-8h	Affymetrix v2.0	KW3-jumu-D2	943	punctate
kn	E0-12h	Affymetrix v2.0	KW0-KN7697	792	punctate
kni	E8-16h	Solexa	KW3-kni-D2	662	punctate
Kr	Kc167	Affymetrix v2.0	KW3-Kr-D2	2,809	punctate
Kr	E0-8h	Affymetrix v2.0	KW3-Kr-D2	869	punctate
Pl	E0-8h	Solexa	KW4-Pcl-D2	2,457	punctate
run	E0-12h	Affymetrix v2.0	KW0-RUN7659	333	punctate
sens	E4-8h	Affymetrix v2.0	a-sens-HB	11,773	mixed
sens	E4-8h	Affymetrix v2.0	KWG-GFP	16,070	mixed
Sin3A	E0-12h	Solexa	a-Sin3A-RC	4,046	punctate
Snr1	Pupae	Affymetrix v2.0	a-snr1-AD	280	mixed
Stat92E	E0-12h	Affymetrix v2.0	a-STAT92E-EB	105	punctate
tll	E0-4h	Affymetrix v2.0	KWG-GFP	97	punctate
Trl	E3-8h	Affymetrix v2.0	GAF3558	6,438	mixed
Trl	Kc167	Affymetrix v2.0	KW3-Trl-D2	7,692	mixed
Trl	E16-24h	Solexa	KW3-Trl-D2	5,195	mixed
Trl	Kc167	Solexa	KW3-Trl-D2	3,842	mixed
ttk	E0-12h	Affymetrix v2.0	KW0-TTK7691	384	punctate
Ubx	E3-8h	Affymetrix v2.0	KW0-UBX7701	729	mixed
Ubx	E3-8h	Affymetrix v2.0	a-Ubx2-MK	161	mixed
Ubx	E0-12h	Affymetrix v2.0	KW0-UBX7701	1,300	mixed
zfh1	E0-12h	Affymetrix v2.0	KW0-ZFH17684	895	punctate

C

Factor	Stage / Cell Type	NumOfTFBS
HDAC1	E0-12h	4,468
HDAC11	E0-12h	2,301
HDAC3	E0-12h	2,588
HDAC4a	E0-12h	5,960
HDAC6	E0-12h	4,983
CTCF_N_Kc	Kc	2,024
CTCF_N	E0-12h	2,534
CTCF_N_S2	S2	2,254
CTCF_C	E0-12h	3,156
CP190	E0-12h	6,654
Mod(modg4)	E0-12h	3,062
Trl	E0-12h	3,906
BEAF-32	E0-12h	4,711
su(Hw)-1	E0-12h	3,422
su(Hw)-2	E0-12h	3,632