LETTER

doi:10.1038/nature09990

A *cis*-regulatory map of the *Drosophila* genome

Nicolas Nègre[1]*, Christopher D. Brown[1]*, Lijia Ma[1]*, Christopher Aaron Bristow[2]*, Steven W. Miller[3]*, Ulrich Wagner[4]*, Pouya Kheradpour[2], Matthew L. Eaton[5], Paul Loriaux[6], Rachel Sealfon[2], Zirong Li[4], Haruhiko Ishii[3], Rebecca F. Spokony[1], Jia Chen[7], Lindsay Hwang[4], Chao Cheng[8,9,10], Richard P. Auburn[11], Melissa B. Davis[1], Marc Domanus[1], Parantu K. Shah[12], Carolyn A. Morrison[1], Jennifer Zieba[1], Sarah Suchy[1], Lionel Senderowicz[1], Alec Victorsen[1], Nicholas A. Bild[1], A. Jason Grundstad[1], David Hanley[7], David M. MacAlpine[5], Mattias Mannervik[13], Koen Venken[14], Hugo Bellen[14], Robert White[15], Mark Gerstein[8,9], Steven Russell[11], Robert L. Grossman[1,7,16], Bing Ren[4,17], James W. Posakony[3], Manolis Kellis[2] & Kevin P. White[1]

Systematic annotation of gene regulatory elements is a major challenge in genome science. Direct mapping of chromatin modification marks and transcriptional factor binding sites genome-wide[1,2] has successfully identified specific subtypes of regulatory elements[3]. In *Drosophila* several pioneering studies have provided genome-wide identification of Polycomb response elements[4], chromatin states[5], transcription factor binding sites[6–9], RNA polymerase II regulation[8] and insulator elements[10]; however, comprehensive annotation of the regulatory genome remains a significant challenge. Here we describe results from the modENCODE *cis*-regulatory annotation project. We produced a map of the *Drosophila melanogaster* regulatory genome on the basis of more than 300 chromatin immunoprecipitation data sets for eight chromatin features, five histone deacetylases and thirty-eight site-specific transcription factors at different stages of development. Using these data we inferred more than 20,000 candidate regulatory elements and validated a subset of predictions for promoters, enhancers and insulators *in vivo*. We identified also nearly 2,000 genomic regions of dense transcription factor binding associated with chromatin activity and accessibility. We discovered hundreds of new transcription factor co-binding relationships and defined a transcription factor network with over 800 potential regulatory relationships.

To reveal chromatin, promoter and enhancer domains in the genome, we performed a developmental time course of six histone modifications, the *Drosophila* CREB binding protein (CBP) and RNA polymerase II (PolII) across twelve stages of embryonic, larval, pupal and adult development (Supplementary Table 1 and Supplementary Figs 1–2; see Supplementary Methods). We used whole animals to generate the maximum number of chromatin marks across the genome. We identified 506,001 chromatin-associated features covering 101 megabases (Mb) (86.99%) of the non-repetitive genome. To relate these chromatin features to gene activity, we quantified transcript levels by high-throughput complementary DNA sequencing (RNA-seq) with the same biological samples used for chromatin immunoprecipitation (ChIP). Additionally, we mapped 38 functionally diverse transcription factors in different developmental stages and cell types. A total of 155,048 transcription factor binding sites (TFBSs) were identified, including 35,096 unique TFBSs. Of these, 93.76% overlap at least one chromatin feature. We noted that although the majority of factors are bound in discrete regions, several are distributed in larger domains (Supplementary Table 1 and Supplementary Fig. 3). We also characterized the binding distributions of five Histone deacetylases (HDACs), identifying a total of 19,937 HDAC binding sites mapping to 7,692 unique genomic locations. Of these, 99.25% overlap with at least one chromatin feature, and 94.58% overlap with at least one TFBS. All data from this study are available at http://www.modencode.org and http://www.cistrack.org.

For each chromatin mark, very few target genes showed either repressive or activating marks across all of development; most genes were within dynamically marked regions (Supplementary Fig. 4a). We observed three major patterns of chromatin mark distributions, corresponding to active promoters (H3K4me3, H3K9ac, H3K27ac), repressive states and silencers (H3K27me3, PHO/Polycomb response elements (PREs)), and enhancers (CBP, H3K4me1) (Supplementary Fig. 4b).

The first pattern, represented by the activating histone modifications H3K4me3, H3K9ac and H3K27ac, was strongly associated with transcription start sites (TSSs) and was positively correlated with gene expression levels[11] (Supplementary Figs 4b, 5a, 7). Although the enrichment of activating histone modifications was quite marked, we note that a substantial fraction of genes (34%) were expressed but lacked H3K4me3 marks at their annotated TSS (Supplementary Figs 8–12). Regions marked by each activating modification also significantly overlapped with class I insulators, PolII binding sites, and a large fraction of TFBSs (Supplementary Fig. 1).

The second type of pattern, repressive chromatin marks H3K9me3 and H3K27me3, showed a distribution of large domains throughout development (Supplementary Figs 2, 6, 13). As expected on the basis of polytene chromosomes *in situ* data, H3K9me3 marks localized to ~20 developmentally stable domains primarily at centromeres[12]. H3K27me3 marks, in contrast, were remarkably dynamic (Fig. 1a). Dynamic domains may be due to changes in specific cell populations during development or the active addition and removal of H3K27me3 marks. Previous studies have implicated H3K27me3 dynamics in the regulation of homeotic genes, in the differentiation of stem cells, and in developmental processes in vertebrates[13]. We found between 123 and 438 discrete domains present at the developmental stages assayed, each with an average length of ~70 kb (Supplementary Fig. 13a and Supplementary Table 2). One-thousand two-hundred and sixty-four genes were associated with H3K27me3 in at least one stage of development, with 397

[1]Institute for Genomics and Systems Biology, Department of Human Genetics, The University of Chicago, 900 East 57th Street, Chicago, Illinois 60637, USA. [2]Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA. [3]Division of Biological Sciences/CDB, University of California San Diego, La Jolla, California 92093, USA. [4]Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. [5]Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. [6]Signaling Systems Laboratory, Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [7]National Center for Data Mining, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, Illinois 60607, USA. [8]Program in Computational Biology & Bioinformatics, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [9]Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [10]Department of Computer Science, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [11]Department of Genetics and Cambridge Systems Biology Centre, University of Cambridge, Cambridge CB2 3EH, UK. [12]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [13]Department of Developmental Biology, Wenner-Gren Institute, Arrhenius Laboratories E3, Stockholm University, S-106 91 Stockholm, Sweden. [14]Department of Molecular and Human Genetics, BCM, Houston, Texas 77030, USA. [15]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, UK. [16]Department of Medicine, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA. [17]Department of Cellular and Molecular Medicine, Institute of Genomic Medicine & Moores Cancer Center, 9500 Gilman Drive, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

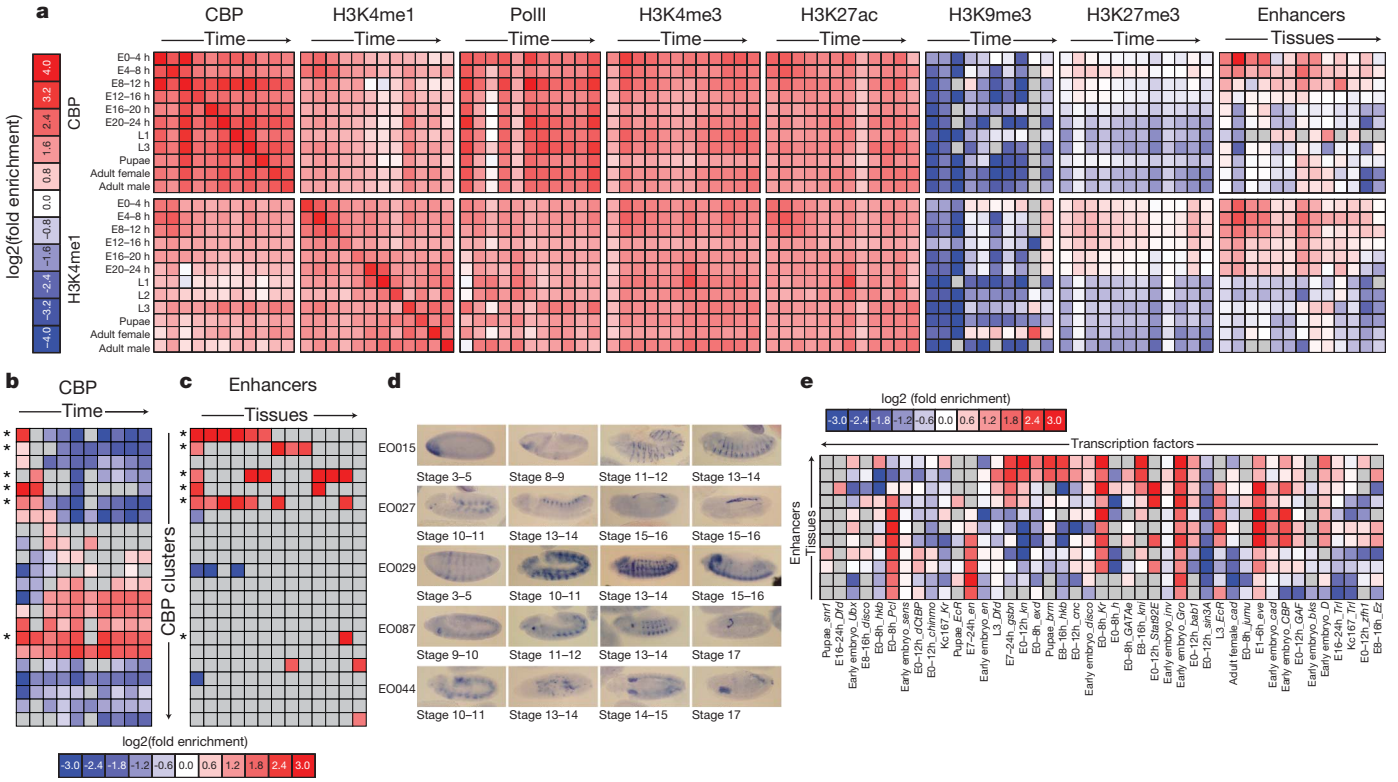

Figure 1 | Chromatin dynamics across *Drosophila* development. **a**, Enrichment of CBP and H3K4me1 (rows) within regions marked by other chromatin modifications, factors, or annotated enhancers (columns). Note that (1) CBP is enriched within all active marks (H3K4me3, H3K27ac, H3K9ac, H3K4me1 and PolII) at all stages of development, and (2) early embryo (0–16 h) CBP- and H3K4me1-marked regions are enriched within H3K27me3 domains and annotated enhancers (right panel). **b**, Heatmap depicting fold enrichment of CBP-bound regions (columns) at different developmental stages for each of the 22 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. **c**, Enrichment of enhancer categories (columns) for each of the 22 clusters of TSS-distal regions (rows). Many clusters enriched for CBP binding in early development are also strongly enriched for enhancers (rows marked with an asterisk). **d**, Embryo-specific CBP binding predicts unannotated enhancers. RNA *in situ* hybridizations with a Gal4 probe were used to stain transgenic embryos representing five different enhancer predictions (rows), at four to five different stages (columns). EO044 overlaps the known expression pattern for the neighbouring gene, *CG8745* (FlyExpress Database). **e**, Enrichment of enhancer annotations (rows) within the binding sites of each transcription factor (columns). For panels **a**, **b**, **c** and **e** grey boxes indicate no overlap. For panels **a**, **b**, **c** and **e** all values greater or less than zero are significant, false discovery rate (FDR) < 0.01.

genes (31%) in domains present in all stages of development and 867 genes (69%) in dynamic domains (Fig. 1a). Stable H3K27me3 domains corresponded to those reported in embryos and tissue culture cells[4], and were enriched for genes involved in development, transcription and segmentation. However, identification of stage-specific H3K27me3 domains revealed previously unappreciated H3K27me3 targets, including genes that control apoptosis, regulation of growth and neurotransmitter transport (Supplementary Fig. 14). We found that stable H3K27me3 domains are highly enriched for genes that exhibit stage- and tissue-specific expression, and are depleted for ubiquitously expressed genes (Supplementary Fig. 15).

H3K4me1-marked and CBP/p300-bound regions form a third, intermediate class of genomic elements known to be associated with active enhancers[3,14] (Fig. 1a). They were also associated with active promoters, activating histone marks and transcription factor binding sites (Supplementary Fig. 1). H3K4me1 and CBP were bound broadly across TSSs, typically positioned 1–2 kb upstream and downstream of the TSS, consistent with previous observations[11] (Supplementary Figs 2, 4b). Accordingly, these patterns were very dynamic across development (Supplementary Fig. 4a).

To characterize the regulators of chromatin mark dynamics, we determined the genome-wide distribution pattern of all five known *Drosophila* HDACs (HDAC1 (also known as RPD3), HDAC3, HDAC4, HDAC6 and HDACX (also known as HDAC11)). All five HDACs are enriched at active promoters, and enrichment is correlated with target gene expression level (Supplementary Fig. 16). HDAC4 and HDAC1/RPD3 binding sites also mark PREs. HDAC1 and HDAC4 binding sites are frequently found within H3K27me3 repressive domains (Supplementary Figs 17, 18), and are significantly enriched at embryonic PHO (a PcG recruiter protein)-bound regions (Supplementary Fig. 16f). Of the 537 HDAC1 and 4a binding sites that overlap H3K27me3 but not H3K4me3 (Supplementary Fig. 16), 149 overlap with 350 previously described[15] embryonic PHO sites (Supplementary Figs 17, 18). HDAC3 is primarily associated with transcribed, H3K36me3 marked exons[16] (Supplementary Fig. 16a, d).

Using the dynamic chromatin signatures and RNA-seq data, we sought next to systematically annotate *cis*-regulatory elements. To identify novel promoters, we identified coincident H3K4me3, PolII and RNA signals at least 1,000 base pairs away from any annotated TSS (see Supplementary Methods). In each developmental stage we found several hundred such regions (average, 485; range, 179–885), resulting in a total of 2,307 novel promoter predictions; 1,117 of which are supported by modENCODE cap analysis of gene expression (CAGE) data from embryos[17] (Supplementary Fig. 5a). We subjected 110 novel promoter predictions to biological validation using a luciferase reporter assay in Kc167 cells. Seventy-five of these 110 predicted promoters (68%) yielded significant luciferase activity in at least one orientation, with 26 displaying bi-directionality (Supplementary Fig. 5b and Supplementary Table 3). Together, the CAGE and reporter assay data indicate that a high proportion of these novel promoter predictions indeed correspond to previously unannotated TSSs.

To identify additional putative *cis*-regulatory elements on a genome-wide scale, we examined two signatures of enhancers, H3K4me1 and CBP/p300[11,14]. CBP and H3K4me1 are significantly enriched within

several classes of known enhancers from the CRM activity database (CAD)[9]. 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 (Supplementary Fig. 19). Given that CBP can be recruited to enhancers by bound transcription factors, we sought to further support the functional relevance of CBP-bound regions by examining clusters of co-occupancy with other transcription factors. Several CBP clusters are bound by transcription factors known to interact physically with CBP, such as Bicoid, Dorsal (DL) and Trithorax-like (TRL (also known as GAF)); whereas other clusters are enriched for known enhancers (Fig. 1e) and are strongly enriched in K3K4me1 and the repressive mark H3K27me3 (Supplementary Fig. 20). In total, 14,450 distinct putative CBP-bound *cis*-regulatory elements were identified across the genome (Supplementary Table 18).

To validate the ability of CBP binding data to accurately identify *cis*-regulatory elements, we tested 33 putative enhancer sequences using reporter assays in transgenic *Drosophila*. We focused on putative enhancers that have dynamic CBP association during embryogenesis (Supplementary Table 4). Thirty of the 33 predicted enhancers produce specific reporter expression patterns (Fig. 1d and Supplementary Fig. 21). We also selected a set of putative insulator elements[10], and we 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 Centrosomal protein 190 kDa (CP190) + CTCF (class I), CP190 + suppressor of Hairy wing (Su(Hw)) (class II) and TRL[10]. 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 TRL sites nor any of the five Su(Hw) sites we tested blocked enhancer–promoter interactions in this assay (Supplementary Fig. 22). These results support a role for CTCF in insulator activity *in vivo*, but indicate that other proteins that have classically been associated with insulator activity are not strictly linked to this function.

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 analysed the patters of 38 diverse transcription factors we mapped at various developmental stages. We compared our data with the CAD database (Fig. 1e) and observed that many factors are specifically enriched in particular enhancer classes. For example, Engrailed (EN) binding sites are enriched within mesothoracic disc enhancers, whereas Knirps (KNI), Tailless and Schnurri (SHN) are enriched within blastoderm enhancers. Indeed, enhancers are usually characterized by multiple transcription factors binding in concert to target genomic DNA. We used a Gaussian kernel density estimation across the binding profiles of 38 transcription factors mapped in early embryos in this and two previously published studies[7,9] to define a 'transcription factor complexity' score based on the number and proximity of contributing transcription factors (see Supplementary Methods). Of 38,562 unique binding sites mapped by the 38 transcription factors, 38.3% are bound by more than two factors. Of the unique binding sites, 5.2% (1,962) are bound by more than eight factors (Supplementary Table 5 and Supplementary Figs 23, 24) and are considered high occupancy target (HOT) regions. Although HOT regions have been observed in *Caenorhabditis elegans*[18] and human (ENCODE project, unpublished results), their function in gene regulation is unknown[7,19].

Regions of higher complexity are weakly associated with more highly expressed genes ($r^2 = 0.19$), indicating that low-complexity binding sites are associated with more restricted expression patterns. Interestingly, annotated enhancers, CBP, activating histone marks including H3K4me1, and HDAC1, 4, 6 and 11 are most significantly enriched within low-to moderate- complexity category (CC) regions (CC2–CC8) (Fig. 2b). 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[16] (Fig. 2b and Supplementary Fig. 24), are depleted from moderate-to-high-complexity regions (>CC4). As expected, transcription factor complexity is inversely correlated with nucleosome enrichment[20] (Fig. 2b). Interestingly, when compared to our enhancer validations and negative controls that were selected independent of HOT region determination, there seems to be no obvious relationship between enhancer activity and HOT regions; 13 validated enhancers overlap with HOT regions but so did several sequences that give no enhancer activity (Fig. 1d, Supplementary Table 4 and data not shown). Taken together, these results indicate that HOT regions are primarily associated with open chromatin but that they do not always demarcate *cis*-regulatory elements.

The existence of HOT regions complicates the interpretation of transcription factor co-occurance. For example, pair-wise clustering of TFBSs resulted in very large groups of co-occurring transcription factors, revealing few specific relationships (Supplementary Fig. 23). However, TFBS clustering performed on HOT-subtracted TFBSs reveals structure that is otherwise obscured when HOT regions are included (Fig. 3). For example, binding sites from different stages assayed for the same transcription factor (for example, TRL, Ultrabithorax (UBX), Ecdysone receptor (ECR)), known interactors (for example, Tinman (TIN) with Twist (TWI) and Biniou (BIN) with Bagpipe (BAP)), and technical replicates (for example, GRO) are more tightly clustered in the HOT-subtracted data. Transcription factors known to physically interact with one another at specific enhancers were significantly associated throughout the genome. For example, the co-repressor complex of GRO and EN and the *Drosophila* SWI/SNF chromatin remodelling complex components Brahma (BRM) and Snf5-related 1 (SNR1) show significant co-binding ($z > 20$). Co-binding enrichment genome-wide was also observed for transcription factors that are known to bind independently to particular enhancers, such as 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[21]. 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 significance levels equal to or greater than those of known interactions. For example, while the previously reported mesodermal transcription factor data set[9] (TIN, TWI, BIN, BAM, MEF2) all have high overlap with one another as expected, these factors also all show highly significant overlap with GRO, CAD and EN. Many other notable overlap pairs were identified, including the Ecdysone receptor with TRL,

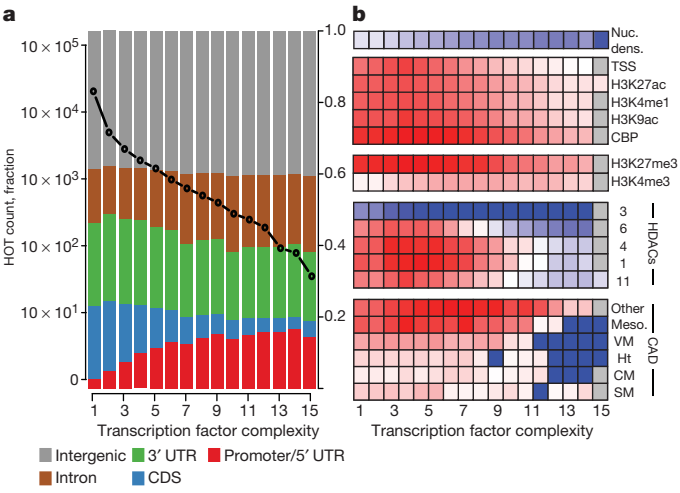

Figure 2 | Transcription factor binding site complexity. **a**, Number of TFBSs (left *y*-axis, black circles) and distribution of genomic annotation classes (right *y*-axis, colours) as a function of TFBS complexity (*x*-axis). **b**, Enrichment (colour scale: depleted in blue, enriched in red) of TFBSs sorted by complexity (*x*-axis) within annotated enhancers (CM, cardiac mesoderm; Ht, heart muscle; SM, somatic muscle; Meso., mesoderm; VM, visceral muscle), HDAC binding sites, early embryo chromatin marks. At the top is a heatmap depicting nucleosome density (Nuc. dens.) as a function of TFBS complexity.

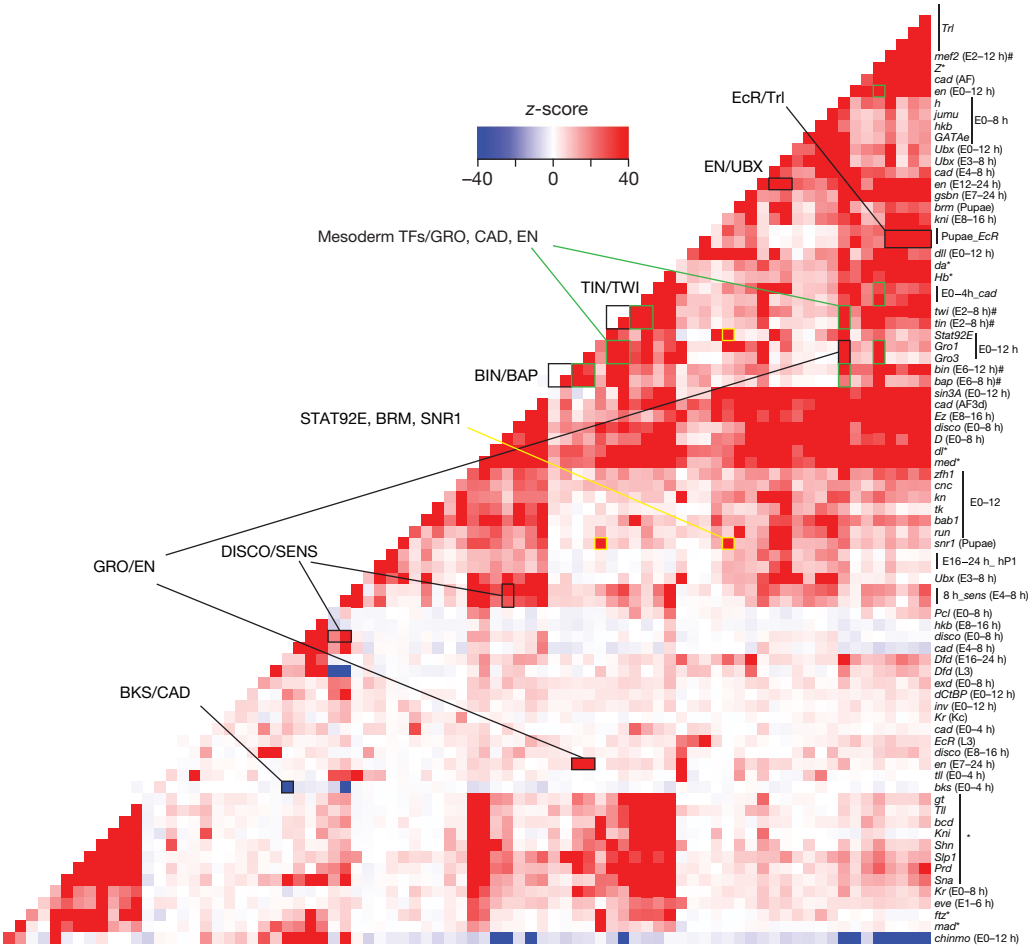

Figure 3 | Transcription factor binding site overlap. Pairwise TFBS enrichment/depletion (colour-coded by *z*-score). TFBS data sets labelled at right. Asterisks indicate data from the BDTNP consortium[6,7]; hashes denote data from the Furlong laboratory[9]. Selected interactions described in the text are highlighted.

the peripheral nervous system master regulator Senseless (SENS)[22] with the axon guidance transcription factor Disconnected, and the Jak/Stat signalling pathway transcription factor STAT92E with the chromatin remodelling complex factors BRM and SNR1—all potential new connections between well-studied regulatory pathways or mechanisms. In total there are 831 highly significant positive pair-wise co-binding interactions in Supplementary Fig. 25 (*z*-score >20; bright red in Supplementary Fig. 25c), most of which are previously undescribed.

Although most significantly associated transcription factor pairs did show positive overlaps, we observed a few instances of highly significant negative associations (shown in blue, Fig. 3). One of the most anti-correlated pairs of transcription factors is Brakeless (BKS) and CAD. 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[23]. In contrast, CAD activates *kni* and *gt* in the same embryonic domain[24]. 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 (Supplementary Fig. 26). The biologically opposing roles of these two transcription factors seem 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 transcription factor co-occupancy has not previously been observed in a metazoan genome.

To visualize the regulatory interactions among transcription factors, we built an intuitive hierarchy representing transcription factor regulatory associations (Supplementary Figs 25, 27; see Supplementary Methods). This network was constructed using 61 transcription factor data sets generated by the modENCODE project (pink nodes) and 20 transcription factors from recently published work[6,7,9] (green and yellow nodes). Specifically, we built a core hierarchy using a breadth-first search algorithm in a bottom-up fashion. Transcription factors that regulate fewer than five transcription factors formed the bottom layer whereas transcription factors that directly regulated the bottom layer factors form the second layer. In total, the network model characterized 835 interactions; 686 were established by transcription factors mapped in this study (blue edges), 125 were derived from previously published data (grey edges), and 24 were auto-regulatory[6,7,9] (Supplementary Fig. 25). Components of the network derived from modENCODE-mapped transcription factors capture many known regulatory interactions; for example, EVE regulates *ftz* and *prd*. However, the vast majority of the 686 transcription factor interactions represent new putative regulatory relationships.

Transcription factors involved in widespread target co-binding and feed-forward loops are also likely to be involved in regulating common patterns of expression. To understand better how combinatorial transcription factor binding regulates developmentally dynamic gene expression, we analysed gene expression data from our RNA-seq time course and an independently performed 64-stage-developmental microarray expression time course. We partitioned the expression data sets into 18 and 64 *k*-means clusters, respectively, which resulted in gene sets with widely varying temporal specificity (Supplementary Fig. 25b). For each cluster of genes, we then quantified the enrichment of promoter-proximal binding sites for 90 modENCODE and previously published transcription factor data sets. From the microarray time-course clustering, five metaclusters were identified. Genes within these metaclusters are most highly expressed at third instar to adulthood (I), first instar to pupal-adult ecdysone pulse (II), early embryos (III), embryogenesis and larval life (IV) and late embryos (V). In both the microarray and RNA-seq time courses, most clusters are significantly associated with a core set of transcription factors including SIN3A, UBX, CAD, SENS and TRL. Interestingly, all metaclusters are enriched for TRL binding sites except V, which is enriched for SNR1, another Trithorax group gene; consistent

with reports that SNR1 has specialized functions[25]. Metacluster II is most highly expressed during adult central nervous system development[26] and enriched for several neuronal differentiation factors (Kruppel, KNI and Jumeau)[27]. Metacluster III uniquely is associated with embryonic patterning and organogenesis transcription factors (for example, Runt, Hunchback, TWI). Notably, many of the transcription factor co-enrichments within gene expression clusters correspond to binding site and regulatory co-enrichments (Fig. 3 and Supplementary Fig. 25), indicating that many of the co-associations of transcription factors with developmental expression patterns reflect co-binding and coordinate regulation at target sites in the genome.

In summary, we generated a draft regulatory annotation map of the *Drosophila* genome from 313 genome-wide data sets that identify or predict thousands of regulatory elements, including 537 silencers, 2,307 newly annotated promoters, 14,450 candidate CBP-bound *cis*-regulatory elements, 7,685 putative insulators[10] and 35,000 unique TBFSs that were bound by one or more transcription factor (Supplementary Tables 6–16 and Supplementary Fig. 28). The transcription factor binding results defined HOT regions of increased transcription factor complexity and their association with HDACs and open chromatin. Subsequent analysis of significantly co-bound transcription factors and transcription factor networks with HOT-subtracted data greatly expands the existing view of regulatory interactions among transcription factors and associates specific sets of transcription factors with specific developmental gene expression patterns. Several unexpected results arose from this initial phase of the modENCODE Project. For example, we revealed a specific class of unmarked promoters, identified a surprising association of HDAC4 and HDAC1/RPD3 to PREs, and discovered pairs of transcription factors that systematically avoid binding near each other throughout the genome. We expect that the results from modENCODE will serve as launch points for many new investigations, and that additional novel insights about the functional consequences of the patterns we describe here will emerge as others in the community engage with these data.

METHODS SUMMARY

ChIP experiments were performed on whole *Drosophila melanogaster* animals from the following developmental stages: embryonic stages 0–4 h, 4–8 h, 12–16 h, 16–20 h, 20–24 h, larval stages L1, L2 and L3, pupal stage and adult male. The biological material was homogenized in 1.8% of formaldehyde. The cross-linked chromatin was sonicated to an average size of 500 bp. Pre-cleared chromatin extract was incubated overnight at 4 °C with the specific antibody and immunoprecipitated. ChIP material was hybridized either on custom Agilent tiling microarrays or on Affymetrix Tiling arrays. For ChIP-seq, standard protocols for Illumina Genome Analysers were used. The software packages used for peak detection were MACS, Peakseq, HGGSEG, CisGenome, MAT and HMMseg where appropriate. For RNA-seq experiments, total RNA was extracted from the same material used for ChIP and processed according to Illumina standard protocols. All methods and scripts used for the analysis of the data are described in Supplementary Information and are available on request. Transgenic assays for promoters, insulators and enhancers are described in Supplementary Information.

Received 23 September 2010; accepted 2 February 2011.

1. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290,** 2306–2309 (2000).
2. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316,** 1497–1502 (2007).
3. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).
4. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735–745 (2007).
5. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143,** 212–224 (2010).
6. Li, X. Y. *et al.* Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6,** e27 (2008).
7. 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,** R80 (2009).
8. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39,** 1512–1516 (2007).
9. Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462,** 65–70 (2009).
10. Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6,** e1000814 (2010).
11. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
12. Schotta, G. *et al.* Central role of *Drosophila* SU(VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21,** 1121–1131 (2002).
13. Agger, K., Christensen, J., Cloos, P. A. & Helin, K. The emerging functions of histone demethylases. *Curr. Opin. Genet. Dev.* **18,** 159–168 (2008).
14. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
15. Kwong, C. *et al.* Stability and dynamics of polycomb target sites in *Drosophila* development. *PLoS Genet.* **4,** e1000178 (2008).
16. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41,** 376–381 (2009).
17. Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21,** 182–192 (2011).
18. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330,** 1775–1787 (2010).
19. Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103,** 12027–12032 (2006).
20. Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19,** 460–469 (2009).
21. Gebelein, B., McKay, D. J. & Mann, R. S. Direct integration of *Hox* and segmentation gene inputs during *Drosophila* development. *Nature* **431,** 653–659 (2004).
22. Nolo, R., Abbott, L. A. & Bellen, H. J. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102,** 349–362 (2000).
23. Haecker, A. *et al.* *Drosophila* brakeless interacts with atrophin and is required for tailless-mediated transcriptional repression in early embryos. *PLoS Biol.* **5,** e145 (2007).
24. Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. & Jackle, H. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* **376,** 253–256 (1995).
25. 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,** 291–308 (2003).
26. Truman, J. W. Metamorphosis of the central nervous system of *Drosophila*. *J. Neurobiol.* **21,** 1072–1084 (1990).
27. Parrish, J. Z., Kim, M. D., Jan, L. Y. & Jan, Y. N. Genome-wide analyses identify transcription factors required for proper morphogenesis of *Drosophila* sensory neuron dendrites. *Genes Dev.* **20,** 820–835 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by U01HG004264 from the National Human Genome Research Institute to K.P.W. and also funded by the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Human Genome Research Institute (NHGRI) or the National Institutes of Health (NIH). C.D.B. is supported by a Lilly-Life Sciences Research Foundation fellowship. C.A.B. is supported by a NIH NRSA postdoctoral fellowship. R.P.A. is in part supported by an Isaac Newton Trust award to R.W. P.L. was supported by a grant from the Department of Energy Computational Sciences Graduate Fellowship (DOE CSGF). M.E.L. and D.M.M. work was supported by NHGRI grant U01 HG004279. We thank the Functional Genomics Facility at the University of Chicago and the High-Throughput Genome Analysis Core at Argonne National Laboratory for processing of microarrays and of Illumina sequence. We thank T.-R. Li, J. D. Lambert, S. Rifkin, T. Herreman, C. Mason, L. Sun and Z. Gauhar for producing the developmental expression microarray data. We also thank the many members of the *Drosophila* community who contributed to this work by providing reagents. A complete list of community participants is included in the Supplementary Methods.

Author Contributions N.N., L.S. and K.P.W. designed and produced modENCODE antibodies; N.N., Z.L., H.I., R.F.S., M.B.D., C.A.M., J.Z., S.S. and M.D. performed the ChIP-chip and ChIP-seq experiments; R.F.S., K.V., H.B. and A.V. produced the GFP-tagged transcription factor *Drosophila* lines; S.W.M., H.I., L.H. and R.P.A. performed the validation experiments of promoters, enhancers and insulators; N.N., P.K.S., N.A.B., A.J.G., D.H. and R.L.G. performed the primary analysis and organized the data of ChIP-chip and ChIP-seq experiments; N.N., C.D.B., L.M., C.A.B., U.W., P.K., M.L.E., P.L., R.S., J.C., C.C., P.K.S., D.M.M. and M.G. analysed the data; M.M. contributed to reagents; N.N., C.D.B., L.M., C.A.B., S.W.M., R.P.A., R.W., S.R., B.R., M.G., J.W.P., M.K. and K.P.W. wrote the paper; H.B., R.W., S.R. (silencer/insulator analysis), R.L.G. (informatics), B.R. (chromatin data and promoter validation), J.W.P. (enhancer/promoter validation), M.K. (data analysis) and K.P.W. (project director) supervised the work.

Author Information Data were deposited at GEO under accession numbers GSE23537, GSE15292, GSE20000, GSE16245, GSE25955, GSE25964, GSE25956, GSE25957, GSE25958, GSE25959, GSE25960, GSE25961, GSE25962 and GSE25963. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.P.W. (kpwhite@uchicago.edu).