

REPORT

Dynamic CRM occupancy reflects a temporal map of developmental progression

Bartek Wilczyński¹ and Eileen EM Furlong*

Department of Genome Biology, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

¹ Present address: Institute of Informatics, University of Warsaw, Warsaw, Poland

* Corresponding author. Department of Genome Biology, European Molecular Biology Laboratory (EMBL), Myerhofstrasse 1, Heidelberg 69117, Germany. Tel.: +49 6221 387416; Fax: +49 6221 387166; E-mail: furlong@embl.de

Received 18.12.09; accepted 30.4.10

Development is driven by tightly coordinated spatio-temporal patterns of gene expression, which are initiated through the action of transcription factors (TFs) binding to *cis*-regulatory modules (CRMs). Although many studies have investigated how spatial patterns arise, precise temporal control of gene expression is less well understood. Here, we show that dynamic changes in the timing of CRM occupancy is a prevalent feature common to all TFs examined in a developmental ChIP time course to date. CRMs exhibit complex binding patterns that cannot be explained by the sequence motifs or expression of the TFs themselves. The temporal changes in TF binding are highly correlated with dynamic patterns of target gene expression, which in turn reflect transitions in cellular function during different stages of development. Thus, it is not only the timing of a TF's expression, but also its temporal occupancy in refined time windows, which determines temporal gene expression. Systematic measurement of dynamic CRM occupancy may therefore serve as a powerful method to decode dynamic changes in gene expression driving developmental progression.

Molecular Systems Biology 6:383; published online 22 June 2010; doi:10.1038/msb.2010.35

Subject Categories: chromatin and transcription; development

Keywords: *cis*-regulatory modules; developmental networks; dynamic binding; temporal regulation; transcription factor binding

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar license to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

Introduction

Transcriptional networks, which serve as the first point of control for gene expression, encompass large numbers of transcription factors (TFs) that bind to specific DNA motifs within *cis*-regulatory modules (CRMs) (Lee *et al*, 2002). Genome-wide ChIP studies are beginning to reveal extensive patterns of TF occupancy in a number of developmental contexts, including *Drosophila* (Sandmann *et al*, 2006, 2007; Zeitlinger *et al*, 2007; Li *et al*, 2008; Liu *et al*, 2009; MacArthur *et al*, 2009; Zinzen *et al*, 2009), mouse (Lagha *et al*, 2008; Vokes *et al*, 2008) and fish (Morley *et al*, 2009). Although this information is an essential first step to decipher developmental networks, a static map of TF binding does not reflect the dynamic nature of gene expression nor predict network behavior. Elucidating the dynamic properties of *cis*-regulatory networks is therefore central to understanding how temporal expression states arise.

Regulating the time span of a TF's expression is one obvious means to restrict its activity and therefore regulate dynamic

gene expression. This can be achieved through the transient expression of the TF itself at precise stages of development (Azpiazu and Frasch, 1993; Lin *et al*, 1997; Bruneau *et al*, 1999; Duan *et al*, 2001; Goldman and Arbeitman, 2007), facilitating the regulation of sequential cascades of expression or by modulating the half-life of the TFs mRNA, as suggested in yeast (Jothi *et al*, 2009). Studies during metazoan development indicate an additional more complex regulatory mechanism where an individual TF can regulate distinct temporal groups of CRMs in more refined temporal windows than their expression suggests (Gaudet and Mango, 2002; Gaudet *et al*, 2004; Sandmann *et al*, 2006, 2007; Jakobsen *et al*, 2007). However, the extent and functional consequence of dynamic CRM occupancy within global developmental networks remains unknown.

A number of computational methods have been used to infer temporal regulation of gene expression, which are mainly based on the specific enrichment of TF motifs in differentially expressed groups of genes (Gao *et al*, 2004; Das *et al*, 2006; Wilczyński *et al*, 2006). Although these approaches can

successfully identify some key TFs (Suzuki *et al*, 2009), global inference of motif activities does not indicate if or when an individual site is occupied, making it difficult to link the timing of a TF's input to the temporal aspect of the target gene's expression. Even with information on motif occupancy, it is difficult to infer regulation as, for example, only ~50% of TF binding is estimated to be functional in yeast (Gao *et al*, 2004; Ucar *et al*, 2009). In higher eukaryotes, the overlap between TF-binding data and regulated gene expression is even lower, typically in the range of 10–50% (Sandmann *et al*, 2006, 2007; Vokes *et al*, 2008; Krejci *et al*, 2009). Therefore, although TFs occupy a vast array of motifs *in vivo*, a large fraction of these are either redundant or not functional under the conditions tested. Here, we used a time course of TF binding to more directly assess the prevalence of dynamic TF occupancy on developmental CRMs and asked to what extent temporal CRM occupancy correlates globally with temporal patterns of gene expression and developmental progression.

Results and discussion

In a recent study, we described the integration of ChIP-on-chip data for 15 developmental conditions to construct a global atlas of mesodermal CRMs, which was used to predict their spatio-temporal activity (Zinzen *et al*, 2009). The five TFs examined (Twist, Tinman, Mef2, Biniou, and Bagpipe) were selected based on their central role within the genetic network controlling *Drosophila* mesoderm development and their tissue-specific expression, avoiding confounding signals from other tissues (Thisse *et al*, 1987; Leptin, 1991; Azpiazu and Frasch, 1993; Lilly *et al*, 1994; Nguyen *et al*, 1994; Bour *et al*, 1995; Zaffran *et al*, 2001). The ChIP experiments for four TFs (excluding Bagpipe) were conducted in time courses spanning at least three consecutive stages of development. The time points examined cover the first stages of mesoderm development when the cells are pluripotent (2–4 h), through their specification into different muscle primordia (6–8 h) and the initiation of tissue differentiation (8–12 h) (Supplementary Figure 1). These data represent the only genome-wide time course of TF-CRM occupancy during embryonic development to date. Here, we used this unique temporal aspect to perform a systematic analysis of the dynamic properties of CRM occupancy during metazoan development.

We first assessed the temporal occupancy of each TF independently, focusing on regions bound by a TF in at least two consecutive time points (see Materials and methods). As these criteria will eliminate CRMs bound at only a single time point, it provides a very stringent set of combinatorially bound modules (Supplementary Table 1). Even within this conservative definition, unsupervised clustering of binding profiles revealed extensive temporal dynamics (Figure 1A), confirming our earlier findings (Sandmann *et al*, 2006, 2007; Jakobsen *et al*, 2007), while extending the analysis genome wide. All TFs examined target three broad classes of CRMs with different temporal occupancy (Figure 1A; Supplementary Figure 2): early bound modules, having TF binding during the first two time points for a given TF, but not later, continuous, occupied at all time points (representing about 50% of CRMs bound by a TF) and late, having binding only at the last two time points and not at earlier stages of development. Therefore, each TF

occupies ~50% of its targeted CRMs in a transient manner; being bound either at early or late stages of development (Supplementary Figure 2). Defining transient occupancy is inherently difficult due to potential false negatives, which can lead to a misclassification of continuous binding. However, measuring the quantitative signal for all CRMs revealed very low levels of occupancy on 'early' CRMs at late developmental stages and conversely a low-binding signal for 'late' CRMs at early stages of development (Figure 2A; Supplementary Figure 3), demonstrating that transient binding is not an effect of thresholding of the ChIP signal (see Materials and methods).

As these four TFs form part of an interconnected regulatory network driving mesoderm specification, the temporally bound regions for each TF merge into 2813 non-redundant CRMs (Supplementary Table 1). Investigating the degree of dynamic TF binding on these combinatorially bound CRMs revealed even more dramatic temporal occupancy (Figure 1B). In all, 83% (2353) of developmental regulatory modules show a transient-binding pattern for at least one of the TFs involved. In some cases, all TFs follow the same temporal pattern, suggesting that the CRM is only accessible at specific stages of development (Figure 1C, CRM #659, #153, #895, #38). For other CRMs, one or two TFs can provide distinct transient inputs, whereas another factor remains constitutively bound, highlighting their inherent complexity of regulation (Figure 1C, CRM #147, #81). This thereby uncovers pervasive temporal regulation within a more refined temporal window than the TFs expression pattern predicts.

To determine how the temporal occupancy patterns arise, we assessed if variation in the binding motifs for the TFs themselves can explain their temporal binding, as shown in *Caenorhabditis elegans* (Gaudet and Mango, 2002) and yeast (Buck and Lieb, 2006). As TF consensus binding sites are typically based on a small number of known sites from *in vitro* approaches, they are often poor predictors of TF occupancy. To obtain a more biological representation of the sequence preferences for each TF, we used the large number of *in vivo* bound regions to optimize the TF-binding site (TFBS) for each TF using all ChIP-bound regions (Zinzen *et al*, 2009) and separately in the early and late-bound CRMs (Supplementary Figure 4B). As expected from earlier studies (Li *et al*, 2008), the ChIP-bound regions are highly enriched in the optimized TFBS for the bound TFs (Figure 2B; Supplementary Figure 4). However, there is no significant difference in the enrichment (Figure 2B; Supplementary Figure 4), or apparent quality (Figure 2C; Supplementary Figures 4B and 5), of the TF-binding motifs in different temporal classes of CRMs. Assessing the contribution of multiple low quality, presumed weak affinity, sites to the strength of the ChIP signal using a quasi-biophysical model (TRAP; Roeder *et al*, 2007) revealed a similar trend (Supplementary Figure 6). Therefore, temporal dynamics in TF binding cannot be readily explained by the affinity of the binding sites for the TFs themselves. Conversely, many occupied CRMs contain a high-quality binding site for a given TF and yet are not bound by that factor (Figure 2C, blue dots; Supplementary Figures 5 and 6). Collectively, these results show that although motif presence is generally necessary, it is not sufficient to predict TF binding. To further confirm this, we examined the enrichment of conserved motifs within different temporal classes of CRMs, as these motifs are

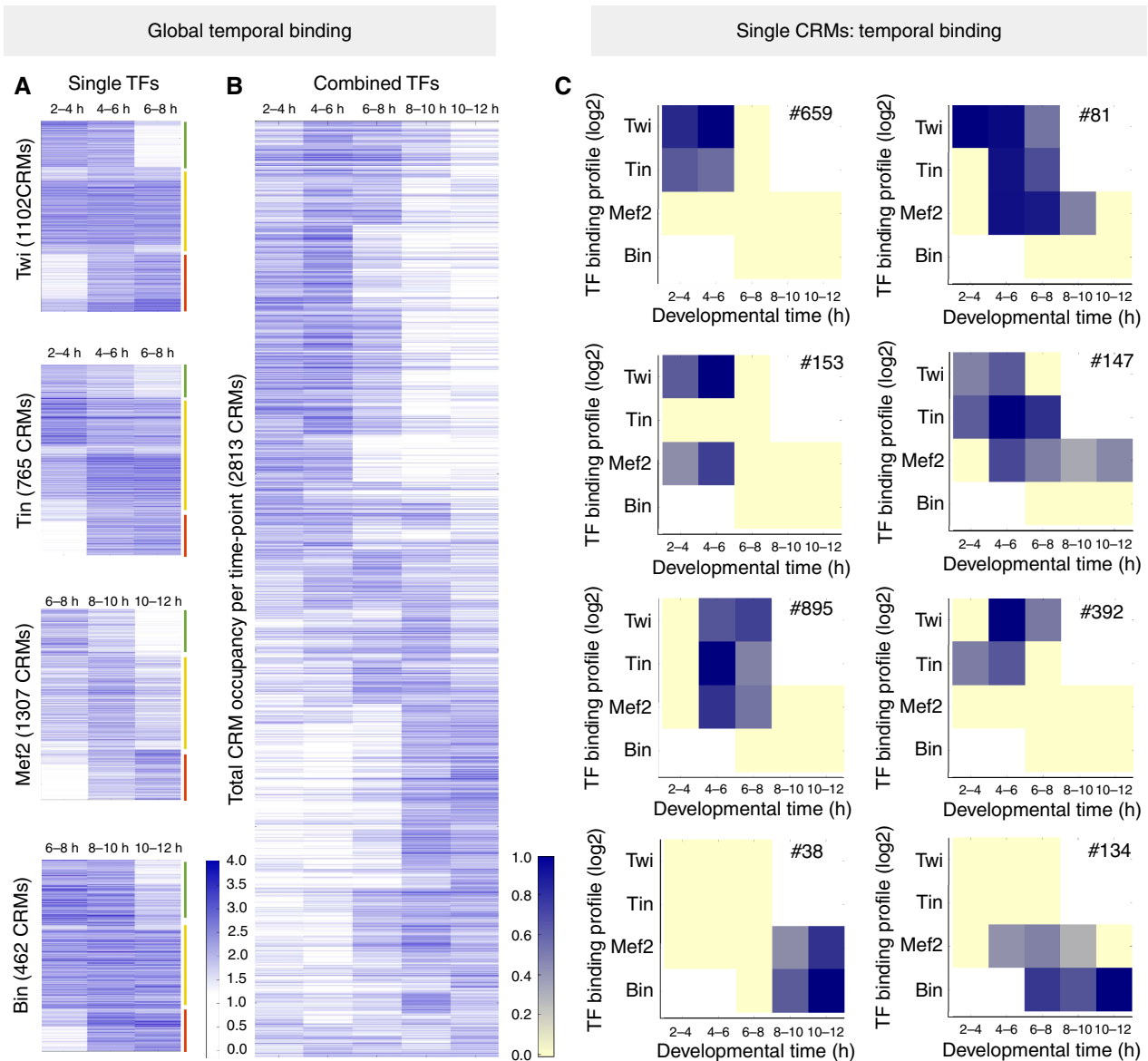


Figure 1 Transcription factors occupy CRMs in highly dynamic patterns. **(A)** *k*-Means clustering of binding profiles of four TFs (Twi, Tin, Mef2, and Bin) shows emerging temporal-binding patterns, which are similar in all four cases. Each row represents a TF-bound region, each column a time window of development. The intensity of blue represents the level of ChIP enrichment (log₂ of the peak height). Colored vertical bars represent early (green), continuous (yellow), and late (red) bound regions. **(B)** Hierarchical clustering of total CRM occupancy, using the average TF-binding enrichment at a given time point, for all 2813 unique CRMs. **(C)** Representative examples of individual CRMs from (B), illustrating that different temporal-binding patterns can occur in the same locus either in a coordinated (CRMs #659, 153, 895, 38) or cascading (CRMs #81, 147, 392, 134) manner. Blue represents the level of ChIP enrichment, yellow represents no binding.

more likely to be the functionally relevant sites. The TF motifs are more highly conserved in bound CRMs compared with non-bound regions (Figure 2d), as reported earlier (Li *et al*, 2007; Zinzen *et al*, 2009). However, there is no significant difference in the enrichment of conserved motifs between the temporal classes (Figure 2D). One potential mechanism of temporal CRM occupancy is the cooperative recruitment of TF binding by other TFs, as observed in a number of well-characterized enhancers (Ip *et al*, 1992; Woodard *et al*, 1994; Broadus *et al*, 1999; Halfon *et al*, 2000; Li *et al*, 2000; Lee and Frasch, 2005). Indeed, additional TF motifs are enriched in specific temporal classes of CRMs (Supplementary Table 2).

For example, the Twist motif is enriched in Tinman early bound CRMs ($P < 0.001$) but not in late-bound modules ($P = 0.77$). This differential motif enrichment is in concordance with the high frequency of temporal cobinding by these TFs to the same CRMs (Supplementary Figure 7). In summary, the presence of a high-affinity TF motif within a CRM is a poor predictor of TF occupancy at any given stage of development, even if the motif is conserved, which has important implications for sequence-based global CRM predictions and quantitative modeling of CRM activity.

The high degree of temporal TF binding suggests tight regulation, which implies functional importance. To assess

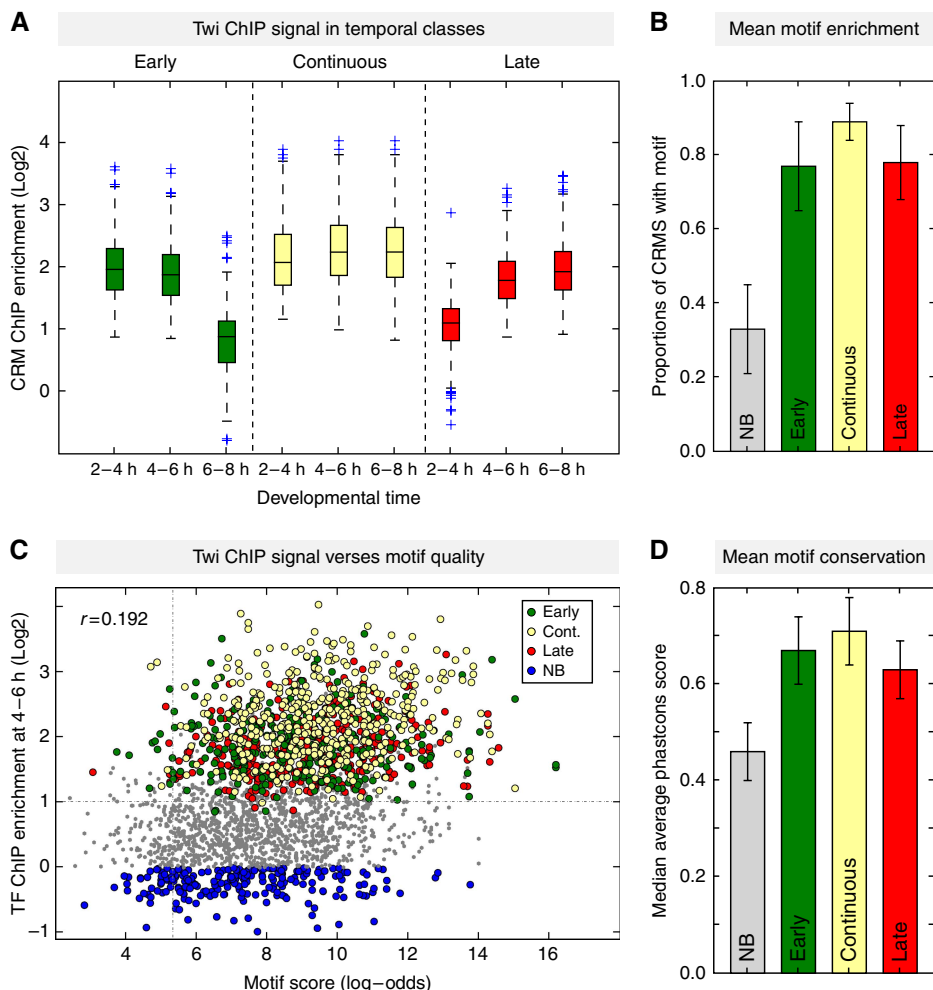


Figure 2 Relationship between TF binding, motif quality, and temporal occupancy. **(A)** The distribution of quantitative-binding signal (log₂ ChIP peak height) for different temporal classes of CRMs bound by the TF Twist shows that there is a clear distinction between early, continuous, and late CRMs. All TFs are shown in Supplementary Figure 3. **(B)** Average enrichment of all motifs in CRM classes defined by binding of respective TFs. The appropriate motif is significantly enriched in bound CRMs compared with non-bound CRMs (NB). There is no difference in motif enrichment between the three temporal classes. **(C)** Quantitative TF binding (ChIP enrichment) is only weakly correlated (Pearson's $r=0.192$) with motif strength as measured by the PWM log-odds score. Exemplified using Twi binding at 4–6 h, where each dot represents a single CRM (all TFs shown in Supplementary Figures 5 and 6). A number of CRMs with no detectable Twist binding (blue dots) contain high-affinity motifs. **(D)** Motifs in bound CRMs are significantly more conserved than in non-bound CRMs, but there is no significant difference in motif conservation between temporal classes. NB, non-bound CRMs.

this, we examined the relationship between temporal CRM occupancy and temporal patterns of gene expression and developmental progression. As all four factors are transcriptional activators, we compared the timing of TF binding (both on and off) to the time of activation of the associated target genes, using two independent data sets for gene expression. First, using *in situ* hybridization data from the Berkeley *Drosophila* Genome Project database (BDGP; Tomancak *et al*, 2002), which provide an accurate measure of expression timing in the tissue of interest (Figure 3A and B). Second, using a microarray-based developmental time course with finer temporal resolution, restricting our analysis to tissue-specific genes (Supplementary Figure 8B). Both approaches show a significant correlation between the timing of TF binding and gene expression: For early bound CRMs, for example, the largest proportion of target genes are activated during early stages of development, reflecting the transient occupancy of their CRMs at these stages (Figure 3A). Similarly, for CRMs

that are occupied during mid-embryogenesis (6–8 h) or later (10–12 h), the majority of their target genes are activated during the respective developmental stages (Figure 3A). This trend holds true for each TF taken separately (Supplementary Figure 8A) and for CRMs occupied very transiently at a single time point, despite potentially containing more noise due to false positives (Supplementary Figure 9). The correlation between dynamic TF binding and the timing of gene expression during development is much higher than what has been observed in yeast (Ni *et al*, 2009), which is remarkable given that developmental genes typically have multiple CRMs, adding another level of complexity.

We next investigated whether genes with temporally regulated CRMs are linked to developmental progression. If this is the case, they should have different biological functions that reflect cellular processes specific to defined developmental stages. To assess this, we analyzed the enrichment of different functional categories (using GO terms; Ashburner

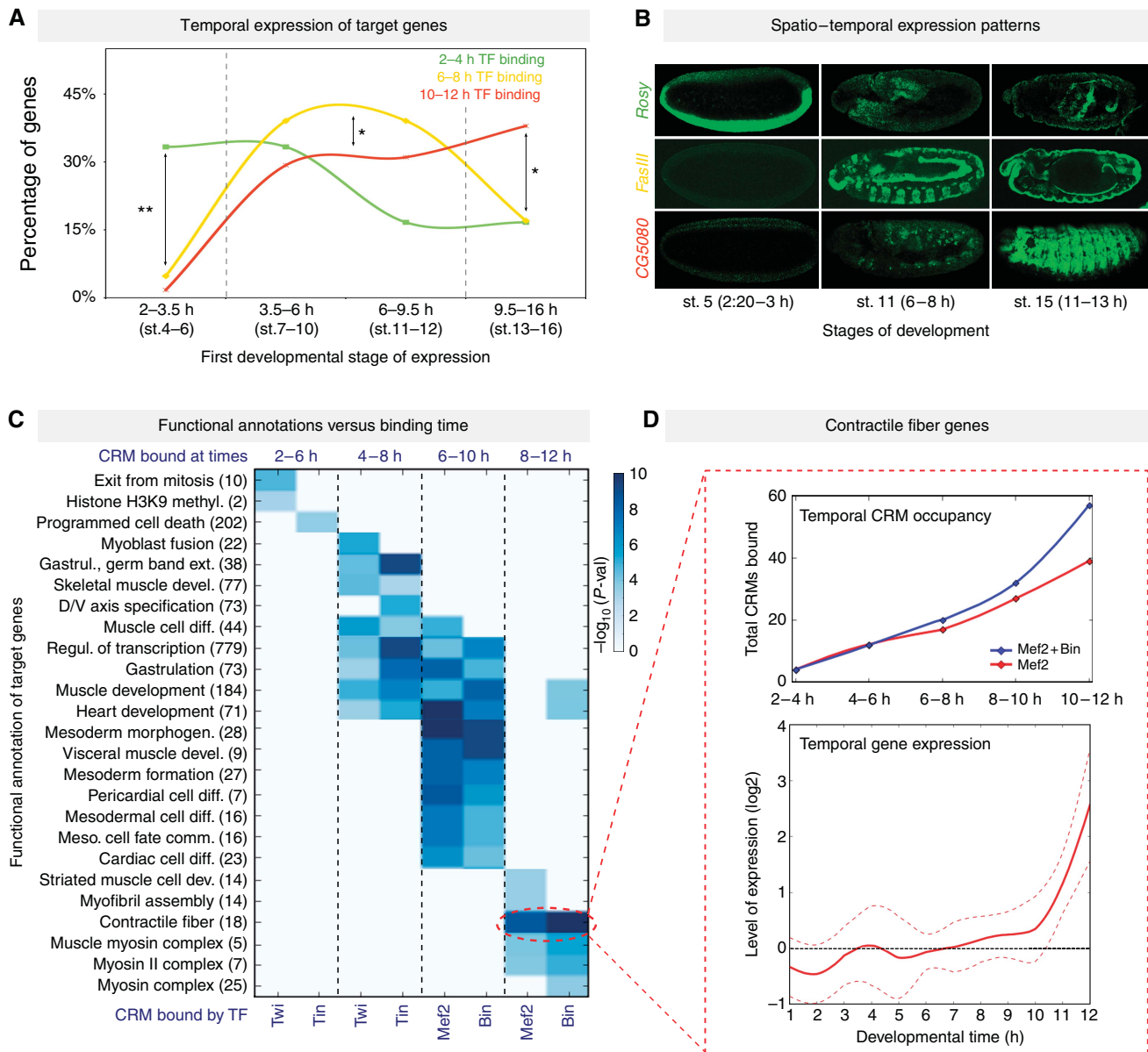


Figure 3 Temporal TF binding correlates with temporal gene expression and function. **(A)** Proportion of target genes with CRMs bound at different time points (green, early; yellow, mid-stages; red, late stages), divided into groups based on the first stage that they are expressed in the embryo. In all, 33% of genes with CRMs bound at 2-4 h initiate expression at 2-3.5 h, whereas an additional 33% initiate expression at 3.5-6 h. Significantly enriched classes peaking in different time points are marked with asterisks (Fisher's exact test $*P < 0.05$, $**P < 0.001$). **(B)** Representative examples of expression patterns of genes targeted by temporally bound CRMs. The only expression of the *rosy* gene in mesodermal and/or muscle occurs at early stages in development, the expression of *fasIII* initiates in visceral and somatic muscle at mid-embryogenesis, whereas *CG5080* is strongly expressed in somatic muscle at late stages. **(C)** Genes with occupied enhancers early in development have different functions than those that are bound by the same TF at later stages. GO categories (with the number of genes indicated in brackets) showing differential enrichment between target genes grouped by temporal-binding classes. The blue shade corresponds to the $-\log_{10}$ of the P -value (Benjamini-Hochberg corrected). For the full set of GO terms see Supplementary Table 3. **(D)** The upper panel displays the number of CRMs bound by Mef2 and/or Biniou out of the 69 CRMs associated with 18 muscle contractile genes. Bottom panel shows the temporal expression pattern of the 18 genes, which initiate at 8-9 h of development, reaching high levels of expression by 11-12 h, reflecting the temporal occupancy of their CRMs. Solid line represents mean expression and dashed lines represent 1 s.d. above and below the mean.

et al., 2000) of genes associated with temporally bound CRMs. In all, 137 GO terms are differentially enriched between genes with early and late-bound CRMs for one or more TF (see Supplementary Table 3). Many of these (41 classes, Figure 3C) are relevant for mesoderm or muscle development and provide a temporal gene activity map for the development of these tissues. For example, Twist and Tinman bind to CRMs of genes involved in mitosis, apoptosis and chromatin remodeling at

very early stages, whereas at later stages of development, these TFs target CRMs of genes essential for gastrulation, transcription and mesoderm morphogenesis (Figure 3C). Similarly, Mef2 and Biniou occupy CRMs of genes involved in mesodermal cell fate commitment and muscle development during mid-embryogenesis, whereas at later stages, they regulate a battery of genes involved in muscle differentiation, including muscle contractile proteins (Figure 3C). Genes

coding for *contractile fiber* proteins are a striking example of the tight correlation between temporal enhancer occupancy and the timing of target gene function. Of the 21 genes in this functional class, 19 are associated with 69 CRMs from our ChIP atlas. In all, 18 contractile genes (95%) have CRMs bound by Mef2 and/or Biniou exclusively at later stages of development (Figure 3D, upper panel). The expression of these genes is tightly synchronized, initiating at ~8 h in development, which mirrors the temporal occupancy of their corresponding CRMs (Figure 3D).

As the temporal aspects of TF occupancy have not been explored globally in other developmental contexts, we cannot compare our findings to other metazoans; however, we can compare to global studies of dynamic responses in yeast. Studies examining the response of TFs to different stimuli in yeast revealed two types of TFs: those that remain bound continuously (termed permanent hubs (Luscombe *et al*, 2004) or condition-invariant regulators (Harbison *et al*, 2004)), and TFs that only bind in a specific condition (Luscombe *et al*, 2004). What we observe within a developmental network is quite different: here, the same TF has the inherent ability to act as a 'permanent' or a 'transient' factor, depending on the context of the CRM it is interacting with. This is not a rare property for selected hubs, but rather occurs in all TFs examined in time series to date, albeit a small number. Interestingly, the proportions of CRMs in each temporal class are very similar among the four TFs, suggesting that this may be a general feature. More recent studies revealed similar dynamic-binding patterns for a number of TFs involved in yeast response to high-salt conditions (Ni *et al*, 2009) and glucose depletion (Buck and Lieb, 2006). Although different TFs regulated target genes that are significantly functionally different, the correlation between dynamic-binding patterns of yeast TFs and gene expression or gene function was more elusive (Ni *et al*, 2009). This comparison suggests that the timing of CRM occupancy may be more tightly coupled to temporal aspects of gene expression in developmental networks, perhaps due to the importance of precise spatio-temporal expression and the irreversible nature of embryonic development.

The composition of regulatory networks describing developmental progression will naturally change over time, as key TFs often have transient and specific expression. Our results show an additional layer of complexity: TFs occupy CRMs in smaller time windows than their temporal expression predicts. As this differential CRM occupancy is occurring within the same cells at the same stages of development, this is unlikely to be controlled by phosphorylation events; the data rather supports a model of extensive cooperative binding (Broadus *et al*, 1999), chromatin remodeling (Buck and Lieb, 2006), or a combination of both. Given the conservative manner in which we have classified temporal binding, the degree of dynamic CRM occupancy is likely to be even more pervasive. These temporal changes in network connectivity are highly correlated with temporal patterns of gene expression and most likely have an extensive function in developmental regulatory networks.

The explosive increase in global TF occupancy data in recent years has revealed that TFs bind to thousands (Sandmann *et al*, 2007; Li *et al*, 2008) or tens of thousands (Chen *et al*, 2008; Vokes *et al*, 2008) of individual sites at any one

condition. This raises a new challenge to distinguish which of these binding events are functionally relevant. Here, we show that dynamic TF occupancy is highly correlated with two functional aspects of target gene expression, its timing and gene function, which clearly contrasts to what is expected from random non-functional sites. Dynamic TF occupancy may therefore provide a mechanism to distinguish functional from non-functional binding events and can thereby help to decode dynamic changes in gene expression driving developmental progression.

Materials and methods

Defining temporal CRM classes

The ChIP data set contained 8008 CRMs constructed from ChIP peaks measured over 15 developmental conditions (Zinzen *et al*, 2009): five TFs (Twist, Tinman, Mef2, Bagpipe, and Biniou) measured during five time points (2–4 h, 4–6 h, 6–8 h, 8–10 h, and 10–12 h). To define temporal occupancy, we selected all CRMs bound by at least one TF for at least two consecutive time points. This resulted in a set of 2813 CRMs containing 1102 CRMs bound by Twist, 765 CRMs bound by Tin, 1307 CRMs bound by Mef2 and 462 CRMs bound by Biniou (Supplementary Table 1). Bagpipe was assayed only for a single time point and was therefore removed from the data set. As these TFs constitute a transcriptional hierarchy during mesoderm development, each factor is only expressed during a subset of the whole time series. Therefore, to define temporal CRM classes for Twist (Twi) and Tinman (Tin), the following three time points were used: 2–4 h, 4–6 h and 6–8 h. For Biniou (Bin) and Myocyte enhancer factor 2 (Mef2), the time points 6–8 h, 8–10 h, and 10–12 h were used. CRMs bound during the two earlier times and not at the third were called 'early'; those bound at the second and third time point but not at the first are called 'late.' Remaining CRMs were continuously bound.

Clustering analysis using the quantitative ChIP signal

The quantitative ChIP signal was calculated for each CRM at each experimental condition (TF × time) as a maximum average value of probe intensities (Log_2) over a sliding window of size 200 bp (the minimal CRM length). The signals were quantile normalized (Bolstad *et al*, 2003) to make them comparable between conditions (Supplementary Table 1). The quantitative CRM profiles were clustered using the *k*-means method as implemented in the Biopython library (de Hoon *et al*, 2004).

Motif analysis

The Biopython library (Cock *et al*, 2009) was used to find all occurrences of known *Drosophila* TFBS motifs (125 motifs from JASPAR (Sandelin *et al*, 2004) and five PWMs for the TFs examined (Zinzen *et al*, 2009)). The threshold for motif occurrence was set based on the information content of the position weight matrix (Hertz and Stormo, 1999). Enrichment of a Motif in a CRM class was assessed using *P*-values of a binomial distribution estimated from the set of all CRMs (Supplementary Table 2). Motif conservation in a CRM class was measured by average median phastCons (Siepel *et al*, 2005) score of the best PWM hit in each CRM. NestedMICA (Down *et al*, 2007) was used to discover motifs *de novo* in all early and late CRM classes (Supplementary Table 2). Quasi-biophysical motif-binding scores for all CRMs were calculated using TRAP (Roeder *et al*, 2007).

Gene expression analysis

To define putative direct target genes, each CRM was assigned to the nearest gene (based on the distance between the middle of a CRM to the nearest gene boundary, 3' or 5'). *In situ* annotations were obtained

from the BDGP database (Tomancak *et al.*, 2002). For each gene, we assessed its first stage of expression in the mesoderm or muscle, representing the first time point where the gene could be regulated by one of the four TFs examined. The timing of TF binding is significantly globally correlated to the timing of the target gene's expression (bootstrapped Kendall $P < 1e-5$). All expression classes show also significant (Fisher exact $P < 0.05$) enrichment for CRMs bound at corresponding time points (Figure 3a). Genes annotated as maternally or ubiquitously expressed or expressed earlier in non-mesoderm related tissue were excluded to avoid confounding expression patterns, as they are not expected to be regulated by mesodermal CRMs.

Functional annotations using gene ontology

GO term (Ashburner *et al.*, 2000) enrichment for target genes associated with temporal CRM classes was calculated using the Ontologizer software (Grossmann *et al.*, 2007) (annotations were obtained from <http://www.GeneOntology.org> on 17 September 2009). Terms enriched for at least one class ($P < 0.05$, Benjamini–Hochberg corrected) are listed in Supplementary Table 3. In the interest of clarity, only terms relevant for development are displayed in Figure 3C.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (<http://www.nature.com/msb>).

Acknowledgements

We thank Paul Bertone and Ewan Birney for comments on the manuscript and Przemyslaw Biecek for insightful comments on statistical analyses. We are very grateful to all members of the Furlong laboratory, in particular to Charles Girardot and Mikhaïl Spivakov, for discussions and critical input. This work was supported by a Human Frontiers Science Program grant awarded to EF.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29

Azpiazu N, Frasch M (1993) Tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev* **7**: 1325–1340

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185–193

Bour BA, O'Brien MA, Lockwood WL, Goldstein ES, Bodmer R, Taghert PH, Abmayr SM, Nguyen HT (1995) *Drosophila* MEF2, a transcription factor that is essential for myogenesis. *Genes Dev* **9**: 730–741

Broadus J, McCabe JR, Endrizzi B, Thummel CS, Woodard CT (1999) The *Drosophila* beta FTZ-F1 orphan nuclear receptor provides competence for stage-specific responses to the steroid hormone ecdysone. *Mol Cell* **3**: 143–149

Bruneau BG, Logan M, Davis N, Levi T, Tabin CJ, Seidman JG, Seidman CE (1999) Chamber-specific cardiac expression of Tbx5 and heart defects in Holt-Oram syndrome. *Dev Biol* **211**: 100–108

Buck MJ, Lieb JD (2006) A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nat Genet* **38**: 1446–1451

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh Y-H, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung W-K *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106–1117

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423

Das D, Nahle Z, Zhang MQ (2006) Adaptively inferring human transcriptional subnetworks. *Mol Syst Biol* **2**: 2006 0029

de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* **20**: 1453–1454

Down TA, Bergman CM, Su J, Hubbard TJP (2007) Large-scale discovery of promoter motifs in *Drosophila melanogaster*. *PLoS Comput Biol* **3**: e7

Duan H, Skeath JB, Nguyen HT (2001) *Drosophila* Lame duck, a novel member of the Gli superfamily, acts as a key regulator of myogenesis by controlling fusion-competent myoblast development. *Development* **128**: 4489–4500

Gao F, Foat BC, Bussemaker HJ (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics* **5**: 31

Gaudet J, Mango SE (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295**: 821–825

Gaudet J, Muttumu S, Horner M, Mango SE (2004) Whole-genome analysis of temporal gene expression during foregut development. *PLoS Biol* **2**: e352

Goldman TD, Arbeitman MN (2007) Genomic and functional studies of *Drosophila* sex hierarchy regulated gene expression in adult head and nervous system tissues. *PLoS Genet* **3**: e216

Grossmann S, Bauer S, Robinson PN, Vingron M (2007) Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics* **23**: 3024–3031

Halfon MS, Carmena A, Gisselbrecht S, Sackerson CM, Jimenez F, Baylies MK, Michelson AM (2000) Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* **103**: 63–74

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J-B, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104

Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**: 563–577

Ip YT, Park RE, Kosman D, Yazdanbakhsh K, Levine M (1992) Dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev* **6**: 1518–1530

Jakobsen JS, Braun M, Astorga J, Gustafson EH, Sandmann T, Karzynski M, Carlsson P, Furlong EEM (2007) Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev* **21**: 2448

Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Babu MM (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol Syst Biol* **5**: 294

Krejci A, Bernard F, Housden BE, Collins S, Bray SJ (2009) Direct response to Notch activation: signaling crosstalk and incoherent logic. *Sci Signal* **2**: ra1

Lagha M, Kormish JD, Rocancourt D, Manceau M, Epstein JA, Zaret KS, Relaix F, Buckingham ME (2008) Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program. *Genes Dev* **22**: 1828–1837

Lee HH, Frasch M (2005) Nuclear integration of positive Dpp signals, antagonistic Wg inputs and mesodermal competence

- factors during *Drosophila* visceral mesoderm induction. *Development* **132**: 1429–1442
- Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne J-B, Volkert TL, Fraenkel E, Gifford DK *et al* (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**: 799–804
- Leptin M (1991) twist and snail as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev* **5**: 1568–1576
- Li C, Kapitskaya MZ, Zhu J, Miura K, Segraves W, Raikhel AS (2000) Conserved molecular mechanism for the stage specificity of the mosquito vitellogenic response to ecdysone. *Dev Biol* **224**: 96–110
- Li L, Zhu Q, He X, Sinha S, Halfon M (2007) Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol* **8**: R101
- Li X-Y, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Hendriks CLL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP *et al* (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6**: e27 EP
- Lilly B, Galewsky S, Firulli AB, Schulz RA, Olson EN (1994) D-MEF2: a MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during *Drosophila* embryogenesis. *Proc Natl Acad Sci USA* **91**: 5662–5666
- Lin Q, Schwarz J, Bucana C, Olson EN (1997) Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**: 1404–1407
- Liu Y-H, Jakobsen JS, Valentin G, Amarantos I, Gilmour DT, Furlong EEM (2009) A systematic analysis of tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. *Dev Cell* **16**: 280–291
- Luscombe NM, Madan Babu M, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**: 308–312
- MacArthur S, Li X-Y, Li J, Brown J, Chu HC, Zeng L, Grondona B, Hechmer A, Simirenko L, Keranen S, Knowles D, Stapleton M, Bickel P, Biggin M, Eisen M (2009) Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80
- Morley RH, Lachani K, Keefe D, Gilchrist MJ, Flicek P, Smith JC, Wardle FC (2009) A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation. *Proc Natl Acad Sci USA* **106**: 3829–3834
- Nguyen HT, Bodmer R, Abmayr SM, McDermott JC, Spoerel NA (1994) D-mef2: a *Drosophila* mesoderm-specific MADS box-containing gene with a biphasic expression profile during embryogenesis. *Proc Natl Acad Sci USA* **91**: 7520–7524
- Ni L, Bruce C, Hart C, Leigh-Bell J, Gelperin D, Umansky L, Gerstein MB, Snyder M (2009) Dynamic and complex transcription factor binding during an inducible response in yeast. *Genes Dev* **23**: 1351–1363
- Roider HG, Kanhere A, Manke T, Vingron M (2007) Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* **23**: 134–141
- Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucl Acids Res* **32**: D91–D94
- Sandmann T, Girardot C, Brehme M, Tongprasit W, Stolc V, Furlong EEM (2007) A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev* **21**: 436
- Sandmann T, Jensen LJ, Jakobsen JS, Karzynski MM, Eichenlaub MP, Bork P, Furlong EEM (2006) A temporal map of transcription factor activity: Mef2 directly regulates target genes at all stages of muscle development. *Dev Cell* **10**: 797–807
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050
- Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwiercz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, Katayama S, Schroder K, Carninci P, Tomaru Y, Kanamori-Katayama M, Kubosaki A, Akalin A, Ando Y, Arner E, Asada M *et al* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* **41**: 553–562
- Thisse B, el Messal M, Perrin-Schmitt F (1987) The twist gene: isolation of a *Drosophila* zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Res* **15**: 3439–3453
- Tomancak P, Beaton A, Weiszmann R, Kwan E, Shu S, Lewis SE, Richards S, Ashburner M, Hartenstein V, Celniker SE, Rubin GM (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* **3**: RESEARCH0088
- Ucar D, Beyer A, Parthasarathy S, Workman CT (2009) Predicting functionality of protein-DNA interactions by integrating diverse evidence. *Bioinformatics* **25**: i137–i144
- Vokes SA, Ji H, Wong WH, McMahon AP (2008) A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* **22**: 2651–2663
- Wilczyński B, Hvidsten TR, Kryshchavych A, Tiurny J, Komorowski J, Fidelis K (2006) Using local gene expression similarities to discover regulatory binding site modules. *BMC Bioinformatics* **7**: 505
- Woodard CT, Baehrecke EH, Thummel CS (1994) A molecular mechanism for the stage specificity of the *Drosophila* prepupal genetic response to ecdysone. *Cell* **79**: 607–615
- Zaffran S, Kuchler A, Lee HH, Frasch M (2001) binou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in *Drosophila*. *Genes Dev* **15**: 2900–2915
- Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M (2007) Whole-genome ChIP-chip analysis of dorsal, twist, and snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* **21**: 385
- Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM (2009) Combinatorial transcription factor binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65–70



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.