Transcription regulation and animal diversity

Michael Levine* & **Robert Tjian***†

* Department of Molecular and Cell Biology, Division of Genetics and Development, Center for Integrative Genomics, and
† The Howard Hughes Medical Institute, University of California, Berkeley, 401 Barker Hall, Berkeley, California 94720, USA

Whole-genome sequence assemblies are now available for seven different animals, including nematode worms, mice and humans. Comparative genome analyses reveal a surprising constancy in genetic content: vertebrate genomes have only about twice the number of genes that invertebrate genomes have, and the increase is primarily due to the duplication of existing genes rather than the invention of new ones. How, then, has evolutionary diversity arisen? Emerging evidence suggests that organismal complexity arises from progressively more elaborate regulation of gene expression.

Comparative genome analyses indicate that increases in gene number do not account for increases in morphological and behavioural complexity. For example, the simple nematode worm, *Caenorhabditis elegans*, possesses nearly 20,000 genes[1] but lacks the full range of cell types and tissues seen in the fruitfly *Drosophila*, which contains fewer than 14,000 genes[2]. Indeed, the revelation that the human genome contains only ~30,000 protein-coding genes precipitated a frenzy of speculation regarding the molecular basis of organismal complexity[3,4].

In principle, there are many ways in which a relatively small number of genes could be exploited so as to generate more complex organisms over evolutionary time. Two mechanisms which have been posited as important sources of complexity are alternative splicing—the production of different RNA species from a given gene during mRNA splicing—and DNA rearrangement, where genes themselves are rearranged during cellular differentiation, as used to generate diversity in mammalian immune systems[3–6]. However, we argue here that physiological and behavioural complexity correlates with the likely number of gene expression patterns exhibited during an animal's life cycle. We discuss two pervasive mechanisms for producing complexity at the genetic level—greater elaboration of *cis*-regulatory DNA sequences, which control the expression of nearby genes, and increased complexity in the multi-protein transcription complexes that regulate gene expression.

An expansion in regulatory complexity

Protein coding sequences account for only a small fraction of a typical metazoan genome; less than 2% in the case of the human genome. It is difficult to estimate the *cis*-regulatory content of metazoan genomes, but it is easy to imagine that as much as a third of the human genome, a remarkable one billion base pairs, controls chromosome replication, condensation, pairing, and segregation, and most importantly for our present discussion, gene expression. Even simple creatures like the sea squirt, *Ciona intestinalis*, have an estimated 10,000–20,000 tissue-specific enhancers[7]. A typical genetic locus in *Drosophila* contains several enhancers (along with other *cis*-regulatory DNAs such as insulators) scattered over an average distance of 10 kilobases (kb) of genomic DNA[8] with transcribed DNA comprising just 2 or 3 kb. By contrast, the regulatory DNAs of unicellular eukaryotes such as yeast are usually composed of short sequences (a few hundred base pairs, bp) located immediately adjacent to the core promoter[9].

Commensurate with this increased complexity in *cis*-control elements, 5–10% of the total coding capacity of metazoans is dedicated to proteins that regulate transcription. These proteins fall into several major classes: First there are the sequence-specific DNA binding proteins that mediate gene-selective transcriptional activation or repression. Second, there are the general but diverse components of large multi-protein RNA polymerase machines

required for promoter recognition and the catalysis of RNA synthesis. Finally, there are the chromatin remodelling and modification complexes that assist the transcriptional apparatus to navigate through chromatin.

The yeast genome encodes a total of ~300 transcription factors; this includes both sequence-specific DNA-binding proteins and subunits of general transcription complexes such as TFIID[9]. In contrast, the genome sequences of *C. elegans* and *Drosophila* reveal at least 1,000 transcription factors in each organism[1,10]. There may be as many as 3,000 transcription factors in humans[4]. It would appear that organismal complexity correlates with an increase in both the ratio and absolute number of transcription factors per genome. Yeast contains an average of one transcription factor per 20 genes, while humans appear to contain one factor for every ten genes. Given the combinatorial nature of transcription regulation, even this twofold increase in the number of factors could produce a dramatic expansion in regulatory complexity, as we discuss below.

Regulatory DNA sequences

A typical yeast transcription unit is presented in Fig. 1a. The regulation of gene expression usually depends on DNA sequences located immediately 5′ of the transcription start site (labelled '+1'). Most core promoters contain a TATA element, which serves
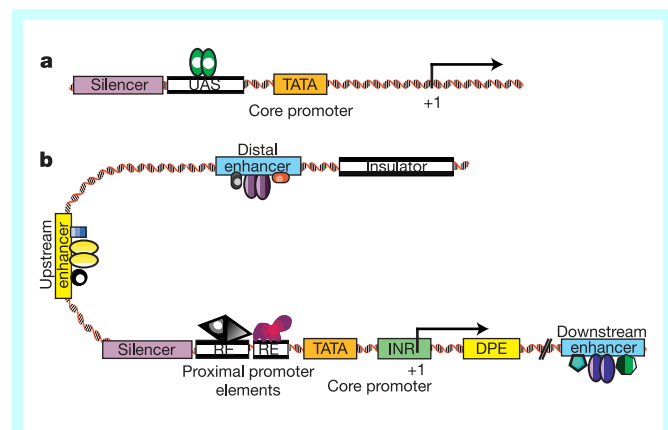
Figure 1 Comparison of a simple eukaryotic promoter and extensively diversified metazoan regulatory modules. **a**, Simple eukaryotic transcriptional unit. A simple core promoter (TATA), upstream activator sequence (UAS) and silencer element spaced within 100–200 bp of the TATA box that is typically found in unicellular eukaryotes. **b**, Complex metazoan transcriptional control modules. A complex arrangement of multiple clustered enhancer modules interspersed with silencer and insulator elements which can be located 10–50 kb either upstream or downstream of a composite core promoter containing TATA box (TATA), Initiator sequences (INR), and downstream promoter elements (DPE).

as a binding site for TBP (TATA-binding protein)[11]. In general, promoters are selected for expression by the binding of TBP to the TATA element. The regulation of TBP binding depends on upstream activating sequences (UAS), which are usually composed of 2 or 3 closely linked binding sites for one or two different sequence-specific transcription factors[12]. A few genes in the yeast genome, such as *HO*, contain distal regulatory sequences[13], but the vast majority contains a single UAS located within a few hundred base pairs of the TATA element.

Metazoan genes contain highly structured regulatory DNAs that direct complex patterns of expression in many different cell types during development (Fig. 1b)[8]. A typical animal gene is likely to contain several enhancers that can be located in 5′ and 3′ regulatory regions, as well as within introns. Each enhancer is responsible for a subset of the total gene expression pattern; they usually mediate expression within a specific tissue or cell type. A typical enhancer is something like 500 bp in length and contains on the order of ten binding sites for at least three different sequence-specific transcription factors, most often two different activators and one repressor[8]. The core promoter is compact and composed of ~60 bp straddling the transcription start site. There are at least three different sequence

Figure 2 The multi-subunit general transcription apparatus: identification of tissue-specific and gene-selective subunits. Diversified metazoan transcription initiation complexes. **a**, The eukaryotic transcriptional apparatus can be subdivided into three broad classes of multi-subunit ensembles that include the RNA polymerase II core complex and associated general transcription factors (TFIIA, -B,-D,-E,-F and -H), multi-subunit cofactors (mediator, CRSP, TRAP, ARC/DRIP, and so on) and various chromatin modifying or remodelling complexes (SWI/SNF, PBAF, ACF, NURF and RSF). **b**, **c**, Metazoan organisms have evolved multiple gene-selective and tissue-specific TFIID-like assemblies by using alternative TAFs (TBP-associated factors such as the ovarian-specific TAF105) as well as TRFs (TBP-related factors such as TRF2 in *Drosophila* and mice) to mediate the formation of specialized RNA polymerase initiation complexes that direct the transcription of tissue-specific and gene-selective programmes of expression.

elements that can recruit the TBP containing TFIID initiation complex: TATA, initiator element (INR) and the downstream promoter element (DPE)[14]. Many genes contain binding sites for proximal regulatory factors located just 5′ of the core promoter. These factors do not always function as classical activators or repressors; instead, they might serve as 'tethering elements' that recruit distal enhancers to the core promoter[15,16]. Finally, insulator DNAs prevent enhancers associated with one gene from inappropriately regulating neighboring genes[17]. These regulatory DNAs, enhancers, silencers and insulators are scattered over distances of roughly 10 kb in fruitflies and 100 kb in mammals.

This elaborate organization of the regulatory DNA permits the detailed control of gene expression. Indeed, a defining feature of metazoan gene regulation is the use of multiple enhancers, silencers and promoters to control the activities of a single transcription unit. Enhancers were initially identified and characterized in mammalian viruses and cultured cells[18]. They were shown to be composed of multiple binding sites for different regulatory proteins. Subsequent analyses in transgenic animals reveal an even greater level of complexity[8]. Metazoan enhancers integrate different regulatory inputs, such as those produced by multiple signalling pathways, to direct stripes, bands, and tissue-specific patterns of gene expression in *Drosophila* embryos and imaginal disks. Similarly organized enhancers are responsible for the restricted expression of the mouse *Pit-1* gene in the anterior pituitary[19], the localized expression of *Hox* genes in rhombomeres of the vertebrate hindbrain[20] and the selective expression of immunoglobulin genes in mammalian B lymphocytes[21].

Tissue-specific enhancers can work over distances of 100 kb or more. There are numerous examples of long-range gene regulation in flies and mammals. For example, the embryonic enhancers that regulate the mouse and human *Igf-2* gene map over 100 kb from the transcription start site[22,23]. The wing margin enhancer that regulates the *Drosophila cut* gene maps at a similar distance[24], while the tissue-specific enhancers that regulate the *Drosophila Decapentaplegic* gene (*Dpp*), and orthologous genes in vertebrates, map far downstream of the transcription unit[25,26]. This type of long-range regulation is not observed in yeast and might be a common feature of genes that play critical roles in morphogenesis and are therefore subject to stringent regulation.
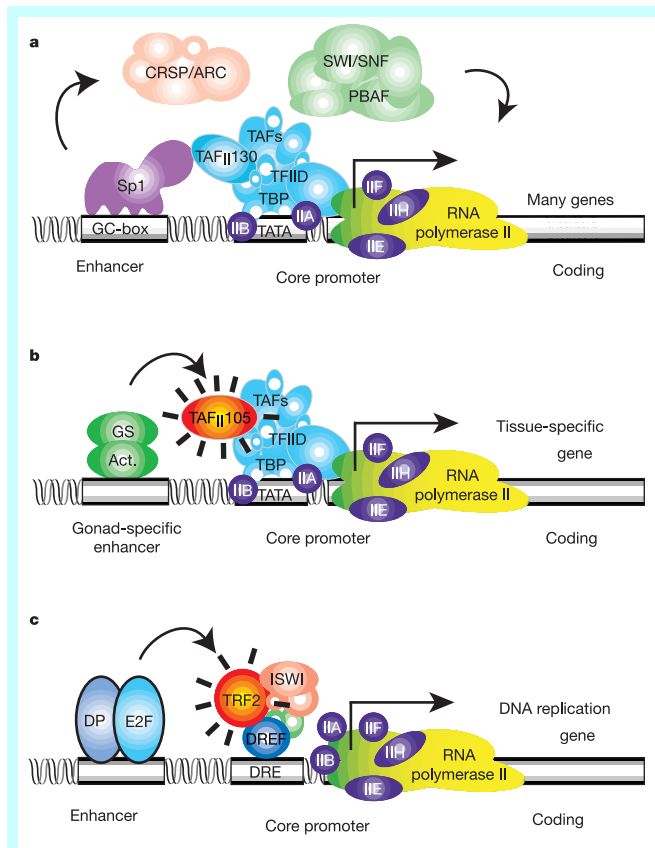
Enhancers generate complex gene expression patterns

Different enhancers can work independently of one another to direct composite patterns of gene expression when linked within a common *cis*-regulatory region. For example, the seven stripes of *even-skipped* expression in the *Drosophila* embryo depend on five separate enhancers; two located 5′ of the transcription start site and three located 3′ of the gene[27,28]. These enhancers function in an autonomous fashion owing to short-range repression: sequence-specific repressors bound to one enhancer do not interfere with the activities of neighbouring enhancers. The repressors must bind within 50–100 bp of an upstream activator or the core promoter in order to inhibit expression[29].

Additional diversity in gene regulation is achieved by the use of multiple promoters for a single gene. For example, the segmentation gene *Hunchback* contains a maternal promoter that is ubiquitously expressed in the germline and a separate zygotic promoter, which mediates restricted expression in the anterior half of early embryos[30].

A potential 'trafficking' problem arises from the fact that *cis*-regulatory DNAs can map far from their target promoters. In some cases the regulatory DNAs actually map closer to the 'wrong' promoter than the proper one[25]. There are at least three underlying mechanisms for ensuring that the right enhancer interacts with the right promoter: insulator DNAs, gene competition, and promoter-proximal tethering elements that recruit distal enhancers.

Insulators are typically 300 bp to 2 kb in length and often contain

clustered binding sites for large zinc finger proteins, such as Su(Hw) and CTCF[17,31]. They selectively block the long-range interaction of a distal enhancer with a proximal target promoter when positioned between the two. Insulators were first identified at gene boundaries, where they prevent cis-regulatory sequences in one gene from inappropriately interacting with neighbouring loci[32]. Insulators have also been identified within complex genetic loci, including the *Bithorax* complex in *Drosophila* and the *Igf-2* locus in mice[31].

Gene competition was first documented in the chicken globin locus, and occurs when a shared enhancer preferentially interacts with just one of two linked promoters[33]. There is emerging evidence that selectivity depends on cis-elements within the core promoter, particularly the TATA sequence, INR and DPE[14] (Fig. 1). Core promoters that lack a TATA sequence usually contain a compensatory DPE element, in order to ensure recognition by the RNA

Figure 3 Diversification of cofactor complexes. Critical components of large multi-subunit cofactor complexes such as mediator and CRSP/ARC have diversified both structurally and functionally, particularly when comparing uni-cellular eukaryotes to their metazoan counterparts. **a**, Yeast mediator versus human cofactor complex. Most of the yeast mediator subunits display very limited, if any, sequence similarity to subunits of human CRSP/ARC although these two cofactor complexes are thought to function in an analogous manner to potentiate RNA Pol II transcription. **b**, *Drosophila* versus human cofactor complex. Even some of the CRSP/ARC subunits have significantly diverged between *Drosophila* and human, although many share conserved regions. We have colour-coded the extent of amino acid sequence similarity between different orthologous co-activator subunits from yeast, *Drosophila* and human. Grey is used to depict yeast subunits with no counterparts in *Drosophila* or humans. White or lightly shaded components represent very little structural conservation (17–20%) while darkly shaded orange and red represent highly conserved (40–80%) subunits.

polymerase II (Pol II) transcription complex. It is possible that some enhancers preferentially activate TATA-containing promoters, while others activate DPE-containing promoters[34,35]. The binding of Pol II depends on an associated multi-subunit complex, TFIID, which is composed of TBP, and TBP-associated factors (TAFs) (Fig. 2A). TFIID binds TATA via TBP, and interacts with DPE and INR through different TAFs[14]. It seems likely that different core promoters can interact with distinct TFIID-like initiator complexes, thus providing a mechanism for matching subsets of enhancers to specific core promoters at complex transcription units (Fig. 2; see below).

There is also evidence that the selection of a particular target promoter by a shared enhancer depends on regulatory factors that bind in promoter-proximal regions, just 5' of the core promoter[36]. It is conceivable that these proximal proteins do not function as classical transcriptional activators or repressors. Instead, they might serve as tethering elements that recruit specific distal enhancers[15,16]. It is easy to imagine a promoter-proximal 'code', whereby one combination of regulatory factors recruits some enhancers, while another combination recruits other enhancers.
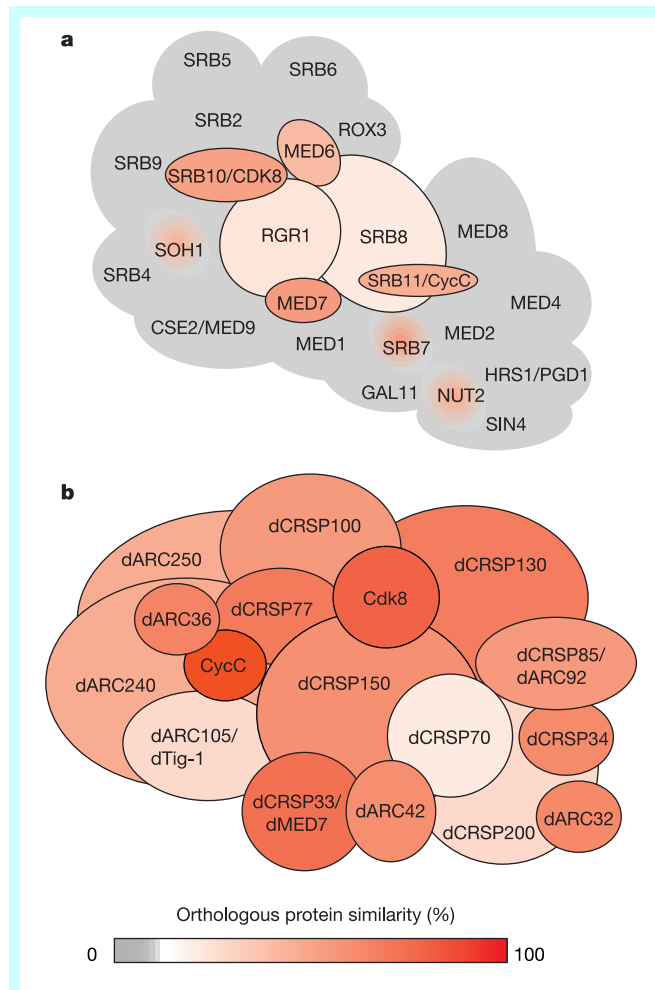
In summary, the *cis*-regulatory DNA of higher metazoans is highly structured and exhibits a modular organization consisting of insulators, silencers, enhancers, and discrete core promoters. We now consider the different protein complexes that interact with these regulatory DNAs. There are three major strategies for regulating the binding and function of the RNA Pol II complex at the core promoter. First, divergent TFIID complexes bind specific sequence elements within the core promoter and recruit Pol II. Second, multi-subunit transcription complexes that are related to the yeast mediator complex also facilitate the binding and function of Pol II. Third, there are enzymatic complexes that remodel or modify (for example, acetylation) chromatin. We discuss each of these pathways of gene activation, with a particular emphasis on diverse TFIID and multi-subunit transcription complexes.

Diversification of core promoter transcription complexes

Multi-subunit transcription complexes are recruited to distal enhancers through interactions with sequence-specific transcriptional activators. These complexes facilitate the binding and function of Pol II at the core promoter (Fig. 2a). Initial studies on the Pol II complex suggested close conservation of the core RNA polymerase, as well as the accessory factors required for promoter recognition and transcription, including TFIIA, -B, -D, -E, -F, and -H (ref. 37). However, it is now apparent that metazoans have evolved multiple related TFIID complexes that can function at distinct promoters through the use of tissue-specific TAFs and TRFs which are not found in yeast.

A tissue-specific mammalian TAF, $TAF_{II}105$, which is related to the ubiquitously expressed human $TAF_{II}130$, operates as part of a unique TFIID complex in follicle cells of the ovary, thereby permitting the selective activation of a small subset of genes[38] (Fig. 2b). Similarly, a testes-specific homologue of $TAF_{II}80$ (Cannonball) is required for spermatogenesis in *Drosophila*[39], and there is emerging evidence that Cannonball may be a component of a complete sperm-specific TFIID complex.

Diverse TFIID complexes have also evolved through the duplication of TBP. There is only one TBP gene and no TRFs in yeast, but there are four TRFs (TRF1–4) in addition to TBP in *Drosophila*[40–46]. TRF1 binds poorly to TATA sequences, but instead initiates transcription at a small group of TATA-less Pol II core promoters. TRF1 is known only in insects, whereas orthologues of TRF2 have also been identified in *C. elegans*, frogs, mice and humans[41–43] (Fig. 2). TRF2 might be essential for mammalian spermiogenesis[44,45] but exerts a more pervasive influence on gene expression during the early embryonic development of *C. elegans*, *Drosophila*, zebrafish, and *Xenopus*[42,43,46]. *Drosophila* TRF2 does not bind TATA but is targeted to distinct core promoters by its interaction with a partner,

DREF[46], which selectively activates genes encoding DNA replication proteins (Fig. 2c). These findings suggest that metazoans have evolved TFIID-related transcription complexes responsible for recognizing distinct core promoters with specific regulatory activities. As mentioned earlier, a number of these genes contain multiple tandem promoters[40,46]. These may be recognized by distinct TFIID-related complexes in different cell types.

Transcription cofactor complexes

The yeast mediator is a multi-subunit co-activator complex that is thought to facilitate the binding and/or function of Pol II at the core promoter[47,48]. While yeast has one such cofactor complex, metazoans contain several related complexes: TRAP, CRSP, ARC/DRIP, SMCC and hMed. These complexes are recruited to the DNA template via interactions with a variety of sequence-specific transcriptional activators, including nuclear receptors such as the vitamin D receptor and the thyroid hormone receptor[49–52]. Like TFIID, these cofactor complexes might serve as bridges between distal activators and the core promoter. However, they might not function solely through the simple recruitment of Pol II, but can also be induced to undergo conformational changes that may be essential for activating transcription[53]. Some of these cofactor complexes could also function in transcriptional repression[52]. In the case of human ARC and CRSP, the smaller CRSP complex can augment transcription *in vitro*, while the larger ARC complex might be involved in repression[53].

Protein chromatography and functional transcription assays are identifying a rapidly expanding number of metazoan cofactor complexes. These represent some of the most dramatic examples of the diversification of general transcription complexes in evolution[53–56]. In contrast to other components of the general transcription machinery such as RNA Pol II subunits, TFIIB, -D, -E, -F and -H, most of the protein subunits that comprise the yeast mediator and metazoan cofactor complexes are not highly conserved (Fig. 3a). In fact, only two of the subunits show significant structural conservation, and these correspond to the smallest subunits (yeast Med6/Med7 and human CRSP33/34)[55–57]. The remaining subunits share almost no discernible sequence similarity save for one or two short stretches of 20–30 amino acid residues (for example, only 10% of the 1,507 residues are conserved between human CRSP 150 and yeast RGR1). Moreover, the low-resolution structure of the yeast mediator complex does not obviously resemble the corresponding human complexes[53,57]. Thus, in contrast to the strong conservation of TFIID and Pol II subunits, cofactor complexes have diversified extensively between eukaryotes, and have expanded among metazoans[58,59].

Diversification of cofactor complexes might also reflect the increasing variety of sequence-specific activators seen in different metazoans. A comparison of the fruitfly and human ARC/CRSP subunits reveals roughly 50% overall similarity in amino acid sequence (Fig. 3b). However, close comparisons of some of the orthologues reveal interesting differences. For example, the *Drosophila* ARC 70, -130 and -230 subunits contain an expansion of glutamine residues, a prevalent feature of sequence-specific activators in *Drosophila*, whereas the human CRSP 70, -130 and -230 subunits possess a broad spectrum of co-activation sequence motifs[55,56].

Even a twofold increase in the number of potential cofactors would result in a substantial increase in the combinatorial control of gene expression. Concomitant with this diversification of cofactor subunits is an increase in the size of gene families encoding sequence-specific transcription factors. Single genes in *Drosophila* are often related to an expanded family of similar genes in humans[60]. Different family members come to acquire diverse activation domains, while retaining highly conserved DNA binding domains. This increase in the complexity of activation domains parallels the expansion and diversification of cofactor subunits.

Chromatin remodelling and modifying complexes

The enzyme complexes that either displace (remodel) nucleosomes (for example, Swi/Snf, Baf/Brm, Acf and Nurf) or covalently modify histones via acetylation (for example, the histone acetyl transferases CBP/p300/pCAF, GCN5) represent another potential source of regulatory diversification during metazoan evolution[61–65]. Although such complexes are found in yeast, there is only limited conservation of orthologous subunits in mammals. For instance, there are a number of *Drosophila* and human ISWI-containing chromatin remodelling complexes such as NURF and RSF[66,67] with no apparent yeast homologues. There is also emerging evidence that remodelling complexes such as the BAFs have diversified in mammals along with the acquisition of specialized cell types. For example, a neuron-specific BAF complex in mammals has no obvious counterpart in fruitflies or nematodes[68]. Moreover, some of the remodelling complexes mediate transcriptional repression rather than activation. For example, the MBD2 protein mediates repression on methylated DNA templates by recruiting NuRD[65], a complex that contains both a Swi/Snf nucleosome displacement activity and histone deacetylase (HDAC) activity. Similarly, in *Drosophila*, the Mi-2 complex initiates the transcriptional silencing of homeotic genes by Polycomb; it contains an ATP-dependent nucleosome displacement factor, as well as an HDAC[69].

Perspective

Rapidly evolving cofactor and chromatin remodelling complexes might be critical for the integration of complex *cis*-regulatory information in metazoans, such as long-range enhancer–promoter interactions. Altogether, the expanded number of metazoan TFIID, cofactor, and chromatin remodelling/modifying complexes appears to be commensurate with the diversification of sequence-specific activators, including nuclear hormone receptors, STATS, SMADs, and promoter-proximal factors such as Sp1, CTF and NTF, all of which are unique to metazoans. This diversification is probably essential for implementing regulation by the highly sophisticated *cis*-regulatory DNAs seen in higher metazoans.

It is unlikely that any single core promoter requires all the transcription complexes we have described. Instead, we suggest that different complexes might be required for regulating distinct *cis*-DNA elements in a temporal and tissue-specific manner. For example, different core promoters might be recognized by diverse TFIID complexes, while enhancers might interact with different cofactor complexes. This type of mixing and matching could generate a huge repertoire of distinct gene expression patterns.

Assembled metazoan genome sequences are now available for nematode worms, fruitflies, mosquitoes, sea squirts, pufferfish, mice and humans. These assemblies provide a powerful foundation for the comparative analysis of gene regulation networks. As we enter the post-genome era it is possible to envision the elucidation of a *cis*-regulatory code, whereby different classes of *cis*-DNAs can be identified by simple sequence analysis. Indeed, computational methods have been used to identify novel enhancers in the *Drosophila* genome based on the clustering of *cis*-regulatory elements[70–74]. Increasingly more powerful methods of comparative genomics should identify many of the changes in *cis*-regulatory DNAs and general transcription complexes underlying animal diversity. □

1. Ruvkun, G. & Hobert, O. The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* **282,** 2033–2041 (1998).
2. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287,** 2185–2195 (2000).
3. Baltimore, D. Our genome unveiled. *Nature* **409,** 814–816 (2001).
4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
5. Graveley, B. R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17,** 100–107 (2001).
6. Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394,** 744–751 (1998).
7. Harafuji, N., Keys, D. N. & Levine, M. Genome-wide analysis of tissue-specific enhancers in the Ciona tadpole. *Proc. Natl Acad. Sci.* **99,** 6802–6805 (2002).

8. Davidson, E. H. *Genomic Regulatory Systems: Development and Evolution* (Academic, New York, 2001).

9. Wyrick, J. J. & Young, R. A. Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.* **12**, 130–136 (2002).

10. Aoyagi, N. & Wassarman, D. A. Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcription regulation. *J. Cell Biol.* **150**, F45–F50 (2000).

11. Struhl, K., Kadosh, D., Keaveney, M., Kuras, L. & Moqtaderi, Z. Activation and repression mechanisms in yeast. *Cold Spring Harb. Symp. Quant Biol* **63**, 413–421 (1998).

12. de Bruin, D., Zaman, Z., Liberatore, R. A. & Ptashne, M. Telomere looping permits gene activation by a downstream UAS in yeast. *Nature* **409**, 109–113 (2001).

13. Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**, 41–48 (1985).

14. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* (in the press).

15. Su, W., Jackson, S., Tjian, R. & Echols, H. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes Dev.* **5**, 820–826 (1991).

16. Calhoun, V. C., Stathopoulos, A. & Levine, M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl Acad. Sci. USA* **99**, 9243–9247 (2002).

17. Burgess-Beusse, B. *et al.* The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl Acad. Sci. USA* **99**, 16433–16437 (2002).

18. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).

19. DiMattia, G. E. *et al.* The Pit-1 gene is regulated by distinct early and late pituitary-specific enhancers. *Dev. Biol.* **182**, 180–190 (1997).

20. Vesque, C. *et al.* Hoxb-2 transcriptional activation in rhombomeres 3 and 5 requires an evolutionarily conserved *cis*-acting element in addition to the Krox-20 binding site. *EMBO J.* **15**, 5383–5396 (1996).

21. Genetta, T., Ruezinsky, D. & Kadesch, T. Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell Biol.* **14**, 6153–6163 (1994).

22. Webber, A. L., Ingram, R. S., Levorse, J. M. & Tilghman, S. M. Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* **391**, 711–715 (1998).

23. Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L. & Tilghman, S. M. An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* **9**, 2079–2089 (1995).

24. Dorsett, D. Distant liaisons: long-range enhancer-promoter interactions in *Drosophila. Curr. Opin. Genet. Dev.* **9**, 505–514 (1999).

25. Merli, C., Bergstrom, D. E., Cygan, J. A. & Blackman, R. K. Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev.* **10**, 1260–1270 (1996).

26. DiLeone, R. J., Russell, L. B. & Kingsley, D. M. An extensive 3′ regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo. *Genetics* **148**, 401–408 (1998).

27. Small, S., Arnosti, D. N. & Levine, M. Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* **119**, 762–772 (1993).

28. Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. & Jaynes, J. B. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* **126**, 2527–2538 (1999).

29. Mannervik, M., Nibu, Y., Zhang, H. & Levine, M. Transcriptional coregulators in development. *Science* **284**, 606–609 (1999).

30. Schroder, C., Tautz, D., Seifert, E. & Jackle, H. Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene hunchback. *EMBO J.* **7**, 2881–2887 (1988).

31. Bell, A. C., West, A. G. & Felsenfeld, G. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447–450 (2001).

32. Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941–950 (1991).

33. Choi, O. R. & Engel, J. D. Developmental regulation of beta-globin gene switching. *Cell* **55**, 17–26 (1988).

34. Ohtsuki, S., Levine, M. & Cai, H. N. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes Dev.* **12**, 547–556 (1998).

35. Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* **15**, 2515–2519 (2001).

36. Perkins, A. C., Gaensler, K. M. & Orkin, S. H. Silencing of human fetal globin expression is impaired in the absence of the adult beta-globin gene activator protein EKLF. *Proc. Natl Acad. Sci.* **93**, 12267–12271 (1996).

37. Roeder, R. G. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. *Cold Spring Harb. Symp Quant. Biol.* **63**, 201–218 (1998).

38. Freiman, R. N. *et al.* Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* **293**, 2084–2087 (2001).

39. Hiller, M. A., Lin, T. Y., Wood, C. & Fuller, M. T. Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.* **15**, 1021–1030 (2001).

40. Holmes, M. & Tjian, R. Promoter selective properties of the TBP-related factor TRF1. *Science* **288**, 867–870 (2000).

41. Rabenstein, M. D., Zhou, S., Lis, J. T. & Tjian, R. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc. Natl Acad. Sci.* **96**, 4791–4796 (1999).

42. Kaltenbach, L., Horner, M. A., Rothman, J. H. & Mango, S. E. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell* **6**, 705–713 (2000).

43. Veenstra, G. J., Weeks, D. L. & Wolffe, A. P. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus. Science* **290**, 2312–2315 (2000).

44. Zhang, D., Penttila, T. L., Morris, P. L., Teichmann, M. & Roeder, R. G. Spermiogenesis deficiency in mice lacking the Trf2 gene. *Science* **292**, 1153–1155 (2001).

45. Martianov, I. *et al.* Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell* **7**, 509–515 (2001).

46. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. & Tjian, R. Promoter Selectivity and target gene identification of a DREF-containing TRF2 complex. *Nature* **420**, 439–445 (2002).

47. Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).

48. Koleske, A. J. & Young, R. A. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466–469 (1994).

49. Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl Acad. Sci. USA* **93**, 8329–8333 (1996).

50. Rachez, C. *et al.* Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* **398**, 824–828 (1999).

51. Gu, W. *et al.* A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* **3**, 97–108 (1999).

52. Akoulitchev, S., Chuikov, S. & Reinberg, D. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature* **407**, 102–106 (2000).

53. Taatjes, D. J., Naar, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, function, and activator-induced conformations of the CRSP co-activators. *Science* **295**, 1058–1062 (2002).

54. Naar, A. M. *et al.* Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**, 828–832 (1999).

55. Ryu, S., Zhou, S., Ladurner, A. G. & Tjian, R. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**, 446–450 (1999).

56. Ito, M. & Roeder, R. G. The TRAP/SMCC/Mediator complex and thyroid hormone receptor function. *Trends Endocrinol. Metab.* **12**, 127–134 (2001).

57. Dotson, M. R. *et al.* Structural organization of yeast and mammalian mediator complexes. *Proc. Natl Acad. Sci. USA* **97**, 14307–14310 (2000).

58. Malik, S. & Roeder, R. G. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* **25**, 277–283 (2000).

59. Glass, C. K. & Rosenfeld, M. G. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121–141 (2000).

60. Bromberg, J. & Chen, X. STAT proteins: signal transducers and activators of transcription. *Methods Enzymol.* **333**, 138–151 (2001).

61. Wang, W. *et al.* Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev.* **10**, 2117–2130 (1996).

62. Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14**, 1553–1577 (2000).

63. Tamkun, J. W. *et al.* Brahma: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561–572 (1992).

64. Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. BRG1 contains a conserved domain of the SWI/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170–174 (1993).

65. Feng, Q. & Zhang, Y. The NuRD complex: linking histone modification to nucleosome remodeling. *Curr. Top. Microbiol. Immunol.* **274**, 269–290 (2003).

66. Tsukiyama, T., Daniel, C., Tamkun, J. & Wu, C. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of thenucleosome remodeling factor. *Cell* **83**, 1021–1026 (1995).

67. LeRoy, G., Loyola, A., Lane, W. S. & Reinberg, D. Purification and characterization of a human factor that assembles and remodels chromatin. *J. Biol. Chem.* **275**, 14787–14790 (2000).

68. Olave, I. A., Reck-Peterson, S. L. & Crabtree, G. R. Nuclear actin and actin-related proteins in chromatin remodeling. *Annu. Rev. Biochem.* **71**, 755–781 (2002).

69. Kehle, J. *et al.* dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**, 1897–1900 (1998).

70. Markstein, M., Markstein, P., Markstein, V. & Levine, M. Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **99**, 763–768 (2002).

71. Berman, B. P. *et al.* Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl Acad. Sci. USA* **99**, 757–762 (2002).

72. Rebeiz, M., Reeves, N. L. & Posakony, J. W. SCORE: A computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. *Proc. Natl Acad. Sci. USA* (in the press).

73. Halfon, M. S., Grad, Y., Church, G. M. & Michelson, A. M. Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res.* **12**, 1019–1028 (2002).

74. Rajewsky, N., Vergassola, M., Gaul, U. & Siggia, E. D. Computational detection of genomic *cis*-regulatory modules applied to body patterning in the early *Drosophila* embryo. *BMC Bioinform.* **3**, 30 (2002).

Correspondence and requests for materials should be addressed to R.T. (jmlim@uclink4.berkeley.edu) or M.L. (mlevine@uclink4.berkeley.edu).