Genomic regulatory regions: insights from comparative sequence analysis

Gregory M Cooper* and Arend Sidow*†

Comparative sequence analysis is contributing to the identification and characterization of genomic regulatory regions with functional roles. It is effective because functionally important regions tend to evolve at a slower rate than do less important regions. The choice of species for comparative analysis is crucial: shared ancestry of a clade of species facilitates the discovery of genomic features important to that clade, whereas increased sequence divergence improves the resolution at which features can be discovered. Recent studies suggest that comparative analyses are useful for all branches of life and that, in the near future, large-scale mammalian comparative sequence analysis will provide the best approach for the comprehensive discovery of human regulatory elements.

Addresses
*Department of Genetics, Stanford University, Stanford, CA 94305-9010, USA
†Department of Pathology, Stanford University, Stanford, CA 94305-5324, USA
Correspondence: Arend Sidow; e-mail: arend@stanford.edu

Current Opinion in Genetics & Development 2003, **13**:1–7

This review comes from a themed issue on
Genomes and evolution
Edited by Evan Eichler and Nipam Patel

0959-437X/$ – see front matter
© 2003 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.gde.2003.10.001

Abbreviation
MARs matrix attachment regions

Introduction

The comprehensive discovery and characterization of functional elements currently represent important goals for biomedical research. A sign of the emerging consensus in the biology community that this goal is now within reach has been its recent elevation to one of the grand challenges for genomics research in the near future [1•]. Comparative sequence analysis, which features prominently in this review, will be necessary to achieve this goal.

Functional elements are defined as those portions of a genome that contribute to an organism's progress through life and reproduction. Most mobile elements and pseudogenes, for example, do not fit this definition. Functional elements can be categorized grossly into coding and noncoding classes. Although coding regions comprise no more than 1.5% of the human genome, the minimal amount of the human genome estimated to be under evolutionary constraint is ~5% [2]. Assuming a general relationship between the presence of constraint and the presence of function [3,4], this suggests that a substantial fraction of the human genome consists of noncoding functional elements. Although some of these may be noncoding RNAs [5], a large number are likely to be regulatory elements with roles in modulating gene expression and chromatin organization, among other processes.

It is on these elements that we focus in this review, by briefly characterizing several types of known regulatory element and by discussing recent examples of regulatory elements that have been discovered along various branches of life. In particular, we review examples that provide general insight into the structure and function of regulatory elements or into the location and function of human regulatory elements.

Classes of regulatory element

Functional assays provide the basis for defining regulatory elements and, vice versa, better definitions of regulatory elements permit more refined assays. Improved definitions will ultimately provide the basis for the high-throughput methods that will be necessary to discover and to characterize comprehensively the regulatory elements in the human genome. Below we discuss several types of known regulatory element and the ways in which they are currently defined. Two main classes of regulatory element can be distinguished as those that control gene expression and those that function in chromatin organization.

Control of gene expression: promoters, enhancers and silencers

The best understood noncoding functional elements are promoters. A promoter is defined as the segment of DNA located immediately proximal to the transcriptional start site of a gene that is involved in initiating transcription. Because of this definition, the identification of promoters is dependent on knowledge of the locations of transcriptional start sites, which is consistently being improved for all sequenced genomes. Recently, predictions of the promoters for 10,000 human genes were made [6•]. For verification, 152 DNA segments corresponding to these predictions were randomly selected and examined in a luciferase reporter assay. Nearly 90% of the predicted promoters were found to function, in that they were able to promote expression in several types of cell. This

ARTICLE IN PRESS

confirms that a nearly comprehensive identification, along with basic functional characterization, of human promoters will be feasible in the near future.

Enhancers are regulatory elements that upregulate gene expression by sequence-specific positioning of transcriptional activators. Enhancers can function independently of position and orientation, although they are generally located within hundreds of kilobases of their target genes. Although no attempt has been made so far to characterize enhancer elements comprehensively in an organism, the genetically simple and experimentally tractable sea squirts of the *Ciona* genus [7–10] may provide just such an opportunity. Recently, >100 random genomic fragments were assayed for their ability to direct expression in a specific manner in *Ciona* tadpoles; at least five of these seemed to contain authentic tissue-specific enhancers, thereby providing promise for larger efforts to characterize enhancers in the future [11•].

Silencers are elements that are capable of repressing transcription. Many silencers are found near the promoter of their target gene, but there are various other subclasses [12]. Although less is known about silencers than about enhancers, silencers are clearly important in the regulation of gene expression. Consider the recent sequencing of the chicken gene encoding CD4, in which a functionally characterized human silencer [13] appears to be conserved in the chicken genome [14•]. This level of distant conservation suggests that this silencer has a fundamental role in controlling gene expression.

Control of chromatin organization: insulators and matrix attachment regions

Insulator elements are barriers that separate domains within chromatin and confine the actions of regulatory elements to their appropriate targets. They can block the action of enhancers as well as prevent the spread of chromatin condensation from nearby regions [15]. The existence of these functional elements partially satisfies intuitive curiosity concerning how enhancers can function so specifically and yet so distantly, in terms of nucleotide position, from their target genes. Recent work is advancing our understanding of these elements and facilitating the derivation of molecular mechanisms for their activity [16]. These advances include several comparative studies that highlight the importance of insulators among vertebrates in the *Hoxd* gene complex [17•] and among mammals in the β-globin locus [18•].

Another group of regulatory elements that are likely to have important roles in chromatin organization includes matrix attachment regions (MARs). These regions may mediate binding to the nuclear matrix and may have key roles in the higher-order organization of eukaryotic nuclei [19]. Their importance has been highlighted by the striking discovery that as much as 11% of the noncoding regions that are conserved between human and mouse may be due to the presence of MARs [20•].

Of course, the distinction of regulatory elements into those controlling gene expression and those involved in chromatin structure is not definitive, given the correlation between chromatin structure and transcriptional activity. Indeed, regulatory elements function and interact through numerous mechanisms by virtue of the activities of the bound proteins. Overlaps in functional capacity, such as working associations between enhancers, insulators and MARs, have been already reported [16]. In addition, evidence suggests that the commonly used definition of enhancers is somewhat arbitrary. At least some enhancers function as 'anti-silencers'; that is, they prevent regional chromatin condensation without necessarily having any direct effect on the rate of transcription [21]. A great challenge of systems biology will be to understand regulatory elements in light of their multipotent capacities and their participation in complex regulatory networks [22]. Simple discovery and basic functional annotation on a larger scale, however, will be a prerequisite for this.

Discovery and functional tests of regulatory elements

Aligning orthologous genomic sequence from different species, coupled with finding regions of conservation, is a powerful method for discovering functional elements. The basic principle behind the method is found in standard molecular evolutionary theory: mutations in functional DNA are likely to be deleterious and therefore will be selected against [23], resulting in a reduced rate of evolution in functional elements [3,4,24].

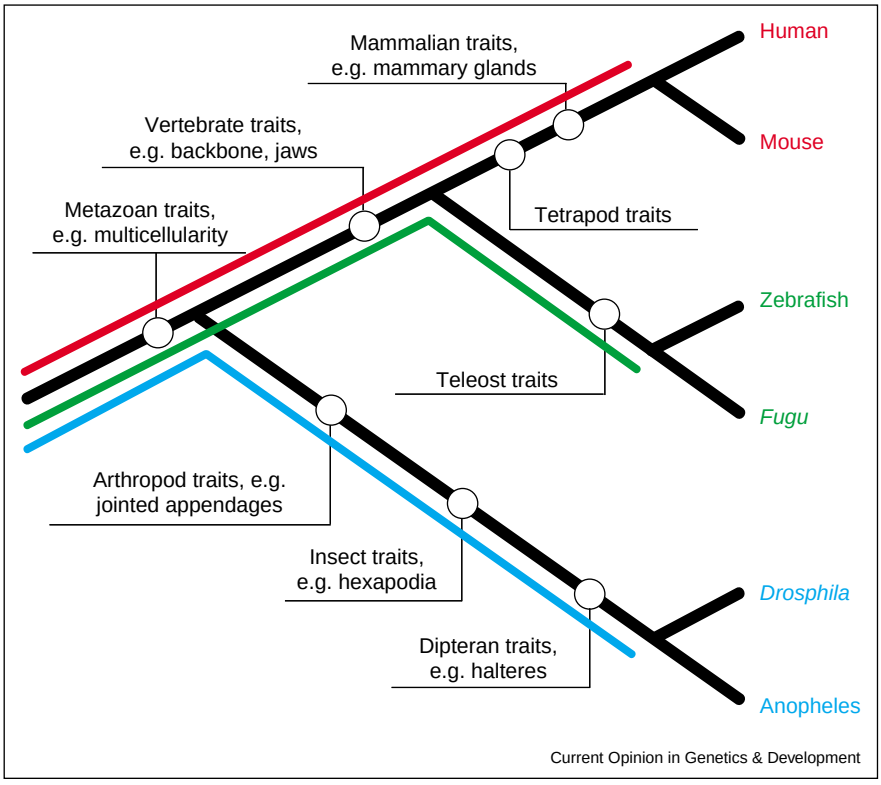
The two most important factors affecting the results of a comparative analysis are, first, the amount of divergence being captured and, second, the phylogenetic scope of the aligned sequences [25]. The amount of divergence affects the power and resolution of the analyses; the scope, which is defined as the narrowest taxonomic group that encompasses all analyzed sequences, affects the applicability of conclusions and the generality of results.

The importance of scope

A given scope captures the biology that is shared among the compared species because of a common line of descent (Figure 1, note colored lines). A dipteran scope, such as one that includes *Drosophila* and *Anopheles*, can be used to find elements necessary for features that were present in their common dipteran ancestor, including traits that evolved before, for example, the diversification of hexapods, arthropods and metazoa. Alternatively, consider a placental mammal scope, such as one including human and mouse: functional elements that might be captured in this scope include not only those that evolved before the diversification of placental mammals, but also

ARTICLE IN PRESS

Figure 1



The importance of scope and the impact of shared ancestry on comparative sequence analysis. The tree describing the relationships among six actively studied genomes is drawn in black (not to scale). Each colored line indicates the phylogenetic scope that applies to each pair of species at the terminal nodes: red line, placental mammal scope; green line, teleost scope; blue line, dipteran scope. Overlaps of the colored lines indicate shared ancestry and capture traits shared by the indicated scopes and, by implication, shared functional elements. For example, whereas placental mammal traits are specific to human and mouse in this tree, vertebrate traits are common to human, mouse, zebrafish and *Fugu*, but exclude insects. As such, a comparative analysis using human and mouse will capture not only placental mammal functional elements, but also vertebrate functional elements. Open circles and associated text show various traits that exemplify the major animal clades and the branch of the tree on which they arose.

those that evolved before the evolution of tetrapods, vertebrates and metazoa.

Note that another consequence of scope is that, barring lineage-specific losses of elements, a conclusion drawn applies to all of the analyzed species. For example, a comparative analysis using human, mouse and rat sequences not only will annotate the human genome, but will also provide data relevant to the mouse and rat genomes, allowing their simultaneous annotation. Comparative analyses on scopes that do not include human but are centered on model organisms such as worms or flies, do not directly annotate the human genome but are valuable for facilitating experimentation in those models.

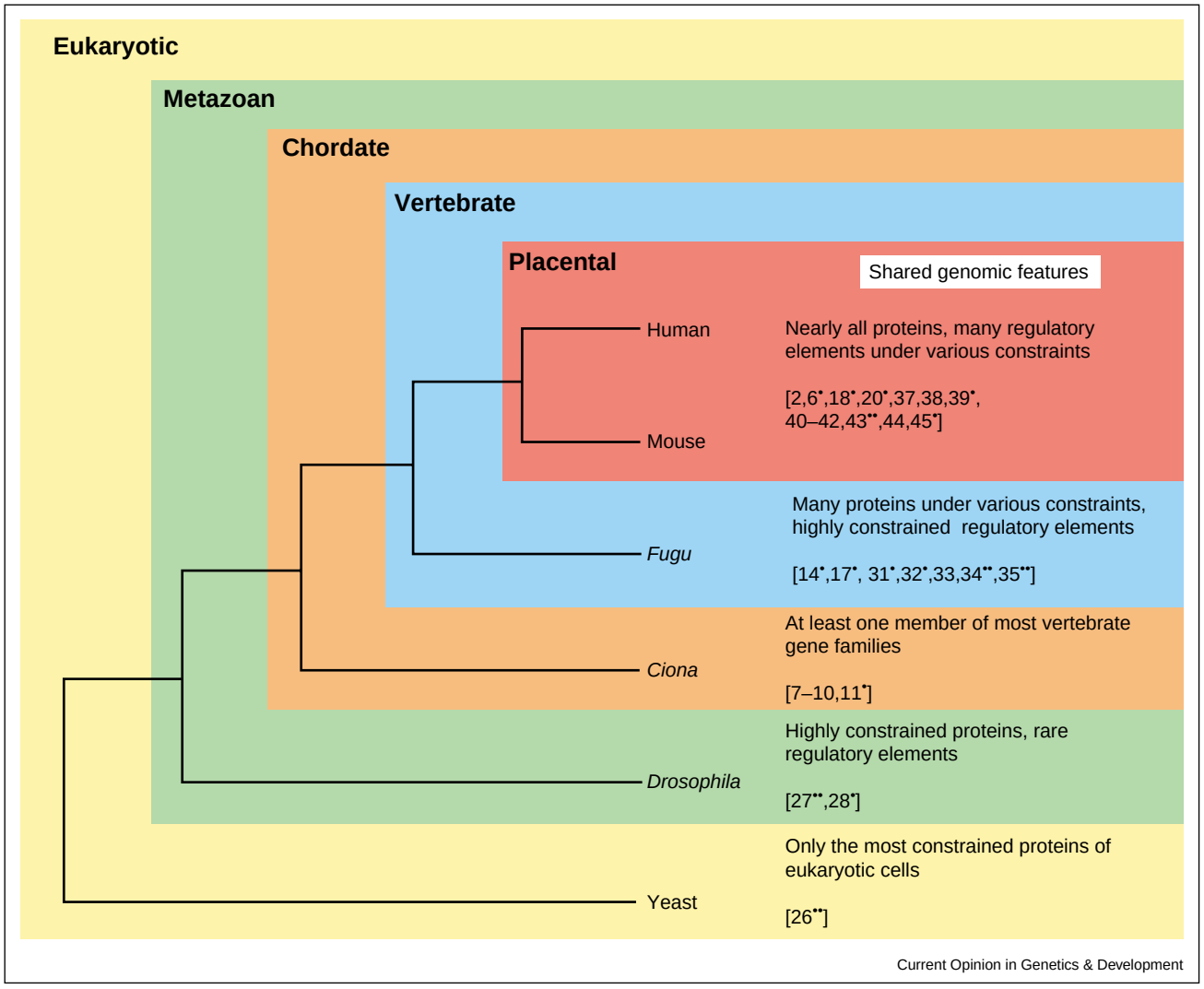
Below we review recent discoveries of regulatory elements that are organized according to the phylogenetic scope employed by the study. Furthermore, we highlight the importance of these discoveries in terms of the insights that they may provide into the biology of human functional elements (Figure 2).

Nonvertebrate model organism scopes

A few recent papers highlight the power of comparative analyses in taxa that are distant from humans. The recent genomic comparison of four species of yeast, constituting the relatively narrow fungal scope of *Saccharomyces sensu strictu*, represents a landmark among these efforts [26••]. By coupling the discovery of conserved noncoding elements with informatic approaches to discover sequence motifs, Kellis *et al.* [26••] were able to generate insight into the organization of regulatory elements and the ways in which they might function combinatorially to control gene expression.

Another excellent example of the utility of model organism scopes is a recent comparison of the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*, from which a model for identifying noncoding RNAs was

Figure 2



Phylogenetic scopes relative to human, depicted around a tree connecting human to other actively studied organisms. Sequentially broader scopes that include relevant species are indicated by nested and colored boxes, beginning with placental mammals. For example, the placental mammal scope (red) that includes human and mouse in this tree, is a subset of the vertebrate scope (blue), which also includes *Fugu*. References discussed in the text and pertinent to each scope are provided in the appropriate boxes. General classes of functional element that are shared among all members of a given scope are also provided.

obtained; this model was then successfully applied to the identification of vertebrate noncoding RNAs [27••]. This underscores the utility of such scopes for understanding general features of metazoan biology.

Model organism scopes have been also applied to the elucidation of general evolutionary mechanisms. For example, the extensive molecular, genetic and developmental resources of *Drosophila*, coupled with the existence of many drosophilids with varying and easily scored phenotypes such as pigmentation, will enable the determination of regulatory changes that cause known evolutionary differences [28•].

Vertebrate scopes

The most divergent phylogenetic scope in which direct sequence comparisons for the detection of human regulatory elements will be generally successful includes vertebrate species. A genome-wide vertebrate scope for comparisons will be facilitated by the recent completion of the genomes of human, mouse and the pufferfish *Fugu rubripes* [2,29,30], and the anticipated sequencing of the rat, chicken, frog and zebrafish genomes. The use of several distant vertebrate species for discovering human regulatory elements is exemplified by the recent study of the *SCL* gene associated with stem cell leukemia [31•].

ARTICLE IN PRESS

Other similar studies have been carried out and provide equally exciting promise. For example, comparison of the *Dlx* gene cluster between fish and mammals detected enhancers that are important for regulating *Dlx* expression in the forebrain [32•]. This is particularly interesting when coupled with the recent characterization of *Dlx* mouse mutants, which has suggested that modification of *Dlx* expression played an important role in the vertebrate acquisition of jaws [33].

Another intriguing example is the discovery of expression control elements near the *Hoxd* cluster in the mouse genome. These elements coordinate expression between genes in the cluster and evolutionarily unrelated genes that flank the cluster, and are highly conserved in the *Fugu* genome [34••]. Finally, enhancers that regulate the expression of chicken *Sox2*, a gene that is involved in the specification of neural cell fates, were discovered through functional assays and were found to correspond to blocks of sequence that are conserved among human, mouse and chicken [35••].

Mammalian scopes

Large-scale alignments of mammalian genomic sequences can be made with high accuracy [36]. Although the degree of sequence similarity among mammals has the drawback that simple pairwise comparisons are generally insufficient for the precise discovery of regulatory elements, such comparisons can succeed in significantly narrowing down the regions in which to search [2,20•].

Many comparisons of human and mouse sequences have been used to estimate the abundance of regulatory elements or to identify and to test specific elements [37,38,39•,40,41]. The completion of more mammalian genome sequences such as rat and dog, in addition to other as-yet-unsequenced mammalian genomes, will provide substantially more comparative power in the future. These improved analyses are likely to discover regulatory elements more precisely and with lower false-positive rates, as well as to provide an estimate of the distributions of the most important positions in the discovered elements [25].

Given the balance between the large amount of shared biology and the sequence diversity among mammals, especially among placental mammals, it is likely that the most abundant and comprehensive discoveries of human regulatory elements will be made through comparisons of several mammalian genomes. This will remain true until the cost of whole-genome sequencing efforts decreases considerably and the large-scale sequencing of primate genomes becomes feasible.

Ultimately, an important aim will be to have a complete list of functional elements in the human genome, including those that evolved in primates after their divergence from other mammals. Such data will provide key clues by which to determine the changes that define humans as a species [42]. Research efforts have been already made with this goal in mind. For example, a recent study has used deep primate sequence data sets, from over a dozen species, to find primate-specific functional elements [43••]. In another example, the evolutionary history of a regulatory element of the gene encoding human interferon-γ has been dissected by using primate sequence comparisons, and this has shown that transposable elements may be involved in mobilizing regulatory elements to new locations in the genome [44].

Studies that are focused specifically on humans also provide, and benefit from, information about regulatory elements. For example, a single-nucleotide polymorphism near the programmed cell death 1 (*PDCD1*) gene is associated with susceptibility to systemic lupus erythematosus. The single-nucleotide polymorphism alters a binding site for a transcription factor and thus is likely to represent a regulatory change that confers disease susceptibility [45•]. This demonstrates the benefits of characterizing noncoding functional elements for directly improving our understanding of human evolution, variability and disease.

Conclusions

Although much success has been achieved in identifying and characterizing genomic regulatory regions in all branches of life, the exciting fact remains that these achievements represent only the tip of the iceberg. Consider that even though much attention in the modern era has been focused on coding genes and on the proteins that they produce, many eukaryotic proteins — not to mention the expanded and more complex protein repertoires of plants and animals — still await functional characterization. Regulatory elements, which substantially outnumber proteins, have been even more elusive and less studied.

This implies a vast, systematic ignorance about the biological world that we seek to understand. Comparative sequence analysis will be needed to eliminate this ignorance in the long term. We anticipate that larger and cheaper volumes of sequencing, improved molecular biology and genetic tools, enhanced computational resources and advances in the study of molecular evolution will facilitate discovery in this emerging, interdisciplinary field.

Comparative sequence analysis leverages biological variation, in the form of neutral evolution, to annotate functional elements. In turn, the annotation of these elements will be necessary to understand functional variation in populations and evolution. Whether we study cell biology, genetics, development, evolution or human disease, it is ultimately this variation that intrigues and motivates us as biologists.

ARTICLE IN PRESS

Acknowledgements

GM Cooper is a Howard Hughes Medical Institute Pre-doctoral Fellow. A Sidow acknowledges support from the Stanford BioX interdisciplinary program and the National Institutes of Health.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Collins FS, Green ED, Guttmacher AE, Guyer MS: **A vision for the future of genomics research**. *Nature* 2003, **422**:835-847.
This paper provides a perspective on the future of genomics research, including goals for research, and comments on the methodology necessary to achieve those goals. The importance of the identification and characterization of all human functional elements, including regulatory elements, is discussed.

2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwal R, Ainscough R, Alexandersson M, An P *et al.*: **Initial sequencing and comparative analysis of the mouse genome**. *Nature* 2002, **420**:520-562.

3. Hardison RC: **Conserved noncoding sequences are reliable guides to regulatory elements**. *Trends Genet* 2000, **16**:369-372.

4. Sidow A: **Sequence first. Ask questions later**. *Cell* 2002, **111**:13-16.

5. Eddy SR: **Computational genomics of noncoding RNA genes**. *Cell* 2002, **109**:137-140.

6. • Trinklein ND, Aldred SJ, Saldanha AJ, Myers RM: **Identification and functional analysis of human transcriptional promoters**. *Genome Res* 2003, **13**:308-312.
The authors predict the promoters of ~10,000 human genes. They test a random selection of 152 of these predictions, confirming that 89% are capable of promoting expression in a cell-culture luciferase reporter assay for at least two types of cell. These cell types may not reflect the genes' native expression domains. This suggests that promoters may function promiscuously in general, and that other factors are necessary for precise spatiotemporal control of gene expression.

7. Di Gregorio A, Corbo JC, Levine M: **The regulation of forkhead/HNF-3β expression in the *Ciona* embryo**. *Dev Biol* 2001, **229**:31-43.

8. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM *et al.*: **The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins**. *Science* 2002, **298**:2157-2167.

9. Keys DN, Levine M, Harland RM, Wallingford JB: **Control of intercalation is cell-autonomous in the notochord of *Ciona intestinalis***. *Dev Biol* 2002, **246**:329-340.

10. Satoh N: **The ascidian tadpole larva: comparative molecular development and genomics**. *Nat Rev Genet* 2003, **4**:285-295.

11. • Harafuji N, Keys DN, Levine M: **Genome-wide identification of tissue-specific enhancers in the *Ciona* tadpole**. *Proc Natl Acad Sci USA* 2002, **99**:6802-6805.
Random genomic fragments from *Ciona*, ~2 kb in size, are generated and tested for their ability to direct tissue-specific expression in *Ciona* larvae. Eleven fragments produce localized expression, and at least five authentic enhancers are identified, suggesting an enhancer density of 1 per 20–50 kb of genomic DNA. The results may provide the basis for much larger future screens.

12. Ogbourne S, Antalis TM: **Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes**. *Biochem J* 1998, **331**:1-14.

13. Donda A, Schulz M, Burki K, De Libero G, Uematsu Y: **Identification and characterization of a human CD4 silencer**. *Eur J Immunol* 1996, **26**:493-500.

14. • Koskinen R, Salomonsen J, Tregaskes CA, Young JR, Goodchild M, Bumstead N, Vainio O: **The chicken CD4 gene has remained conserved in evolution**. *Immunogenetics* 2002, **54**:520-525.
The authors sequence the CD4 gene in chicken and report that the gene is conserved between mammals and chicken. In addition, a functional human silencer [13] also seems to be conserved in chicken.

15. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, Simpson M, West A, Felsenfeld G: **The insulation of genes from external enhancers and silencing chromatin**. *Proc Natl Acad Sci USA* 2002, **99(Suppl 4)**:16433-16437.

16. West AG, Gaszner M, Felsenfeld G: **Insulators: many functions, many mechanisms**. *Genes Dev* 2002, **16**:271-288.

17. • Kmita M, Tarchini B, Duboule D, Herault Y: **Evolutionary conserved sequences are required for the insulation of the vertebrate *Hoxd* complex in neural cells**. *Development* 2002, **129**:5521-5528.
Hox genes play crucial roles in controlling vertebrate morphology, and their expression is precisely regulated in developing vertebrate organisms. Genomic organization of *Hox* genes as clusters, with a specific ordering of genes within the clusters, is important for achieving this precise regulation. The authors discover genomic sequences that are necessary for insulating the *Hoxd* cluster from external influences, including one genomic element that is conserved among mouse, chicken and zebrafish.

18. • Farrell CM, West AG, Felsenfeld G: **Conserved CTCF insulator elements flank the mouse and human β-globin loci**. *Mol Cell Biol* 2002, **22**:3820-3831.
The authors report the discovery of conserved genomic regions that flank the β-globin loci in human and mouse. These regions contain binding sites for CTCF, a protein known to be important for enhancer-blocking insulator activity, and are capable of blocking enhancer activity *in vitro*.

19. Cremer T, Cremer C: **Chromosome territories, nuclear architecture and gene regulation in mammalian cells**. *Nat Rev Genet* 2001, **2**:292-301.

20. • Glazko GV, Koonin EV, Rogozin IB, Shabalina SA: **A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions**. *Trends Genet* 2003, **19**:119-124.
The authors use a previous study [46] in which large stretches of intergenic sequence were aligned between human and mouse and strongly conserved segments were identified. They correlate those segments with MARs, which are predicted by programs that scan for sequence motifs indicative of MARs. The authors conclude that as much as 11% of the strongly conserved segments consist of MARs; they also note that many of the conserved MARs precede the 5′ ends of genes, which may be a consequence of a role in regulating transcription.

21. Martin DI: **Transcriptional enhancers—on/off gene regulation as an adaptation to silencing in higher eukaryotic nuclei**. *Trends Genet* 2001, **17**:444-448.

22. Davidson EH: *Genomic Regulatory Systems: Development and Evolution*. San Diego, CA: Academic Press; 2001.

23. Kimura M: *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press; 1983.

24. Sumiyama K, Kim CB, Ruddle FH: **An efficient *cis*-element discovery method using multiple sequence comparisons based on evolutionary relationships**. *Genomics* 2001, **71**:260-262.

25. Cooper GM, Brudno M, Program NC, Green ED, Batzoglou S, Sidow A: **Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes**. *Genome Res* 2003, **13**:813-820.

26. •• Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements**. *Nature* 2003, **423**:241-254.
The authors perform a thorough analysis on alignments among four species of yeast, generating precise estimates of *Saccharomyces* open reading frames and discovering many putative regulatory elements. Furthermore, they use informatic approaches to find shared sequence motifs in conserved noncoding regions and use these to generate hypotheses regarding gene networks and combinatorial control of gene expression.

27. •• Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: **Vertebrate microRNA genes**. *Science* 2003, **299**:1540.
Using a model to discover microRNAs on the basis of comparative sequence analysis between *C. elegans* and *C. briggsae*, the authors

ARTICLE IN PRESS

are able to generate predictions for the presence of microRNAs in the human, mouse and *Fugu* genomes. Between 200 and 255 such microRNAs are predicted to exist in the human genome.

28. Wittkopp PJ, Vaccaro K, Carroll SB: **Evolution of yellow gene regulation and pigmentation in *Drosophila***. *Curr Biol* 2002, **12**:1547-1556.
An excellent example of the power of *Drosophila* cross-species analysis, this paper analyzes genomic changes that may underlie the divergence of pigmentation pattern among drosophilids. The authors suggest that both coding and noncoding changes are likely to have had roles in this phenotypic divergence.

29. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A *et al.*: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes***. *Science* 2002, **297**:1301-1310.

30. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.

31. Göttgens B, Barton LM, Chapman MA, Sinclair AM, Knudsen B, Grafham D, Gilbert JG, Rogers J, Bentley DR, Green AR: **Transcriptional regulation of the stem cell leukemia gene (*SCL*) — comparative analysis of five vertebrate *SCL* loci**. *Genome Res* 2002, **12**:749-759.
SCL is an important gene involved in vertebrate hematopoiesis. The authors report the characterization of the zebrafish *SCL* locus, and compare it to the *Fugu*, chicken, human and mouse sequences. A regulatory region is found to contain several conserved motifs that are essential for regulating *SCL* expression in hematopoietic cells.

32. Ghanem N, Jarinova O, Amores A, Long Q, Hatch G, Park BK, Rubenstein JL, Ekker M: **Regulatory roles of conserved intergenic domains in vertebrate *Dlx* bigene clusters**. *Genome Res* 2003, **13**:533-543.
Genomic elements that are conserved between fish and mammalian species near the *Dlx* bigene clusters are discovered. Moreover, they are functionally characterized as enhancers that are capable of directing expression to the forebrain. This result may shed light on the genomic changes that accompanies the acquisition of vertebrate characteristics.

33. Depew MJ, Lufkin T, Rubenstein JL: **Specification of jaw subdivisions by *Dlx* genes**. *Science* 2002, **298**:381-385.

34. Spitz F, Gonzalez F, Duboule D: **A global control region defines a chromosomal regulatory landscape containing the *HoxD* cluster**. *Cell* 2003, **113**:405-417.
The authors report the discovery of a cluster of enhancer elements that are conserved between mammals and *Fugu*. This cluster is capable of coordinating expression between *Hoxd* genes and nearby genes that are evolutionarily unrelated. This includes a newly identified gene that the authors name *lunapark*.

35. Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H: **Functional analysis of chicken *Sox2* enhancers highlights an array of diverse regulatory elements that are conserved in mammals**. *Dev Cell* 2003, **4**:509-519.
Several enhancers of *Sox2*, a gene involved in the specification of neural cell identities, are functionally detected in chickens. These enhancers correspond precisely to blocks of strong sequence conservation among human, mouse and chicken.

36. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S: **LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA**. *Genome Res* 2003, **13**:721-731.

37. Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA: **Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons**. *Science* 2000, **288**:136-140.

38. Qiu Y, Cavelier L, Chiu S, Yang X, Rubin E, Cheng JF: **Human and mouse ABCA1 comparative sequencing and transgenesis studies revealing novel regulatory sequences**. *Genomics* 2001, **73**:66-76.

39. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson BJ, Flegel V, Bucher P, Jongeneel CV *et al.*: **Numerous potentially functional but non-genic conserved sequences on human chromosome 21**. *Nature* 2002, **420**:578-582.
This comparison between human and mouse sequence detects a large amount of noncoding conserved sequence. In addition, a thorough analysis is performed to ensure that these conserved blocks are neither protein coding nor RNA coding, confirming the presence of many important regulatory elements distributed throughout the mammalian genome.

40. Kondrashov AS, Shabalina SA: **Classification of common conserved sequences in mammalian intergenic regions**. *Hum Mol Genet* 2002, **11**:669-674.

41. Mural RJ, Adams MD, Myers EW, Smith HO, Miklos GL, Wides R, Halpern A, Li PW, Sutton GG, Nadeau J *et al.*: **A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome**. *Science* 2002, **296**:1661-1671.

42. Carroll SB: **Genetics and the making of *Homo sapiens***. *Nature* 2003, **422**:849-857.

43. Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM: **Phylogenetic shadowing of primate sequences to find functional regions of the human genome**. *Science* 2003, **299**:1391-1394.
Comparisons of many primate sequences can and will yield exciting insights into the functional elements that are important to distinguish humans and their closest relatives from other mammals. This paper represents a striking example of this.

44. Ackerman H, Udalova I, Hull J, Kwiatkowski D: **Evolution of a polymorphic regulatory element in interferon-γ through transposition and mutation**. *Mol Biol Evol* 2002, **19**:884-890.

45. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, Brookes AJ, Tentler D, Kristjansdottir H, Grondal G *et al.*: **A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans**. *Nat Genet* 2002, **32**:666-669.
Relatively few mutations in regulatory DNA (other than in splice sites) that are responsible for human genetic diseases have been reported; this paper is one such example.

46. Shabalina SA, Ogurtsov AY, Kondrashov VA, Kondrashov AS: **Selective constraint in intergenic regions of human and mouse genomes**. *Trends Genet* 2001, **17**:373-376.