

from an IgG1-637-treated monkey induced strong AChR degradation, in contrast to serum from a monkey injected with the IgG4 antibody (Fig. 3B). The latter serum, moreover, prevented AChR degradation by IgG1-637 in an in vitro mixing experiment (Fig. 3C). Finally, serum from an animal treated with both antibodies induced only moderate degradation. The IgG4-637 antibody, which induced AChR degradation before injection (Fig. 3A), apparently acquired non-cross-linking and protective activity in vivo.

Antibody function is regulated by posttranslational modifications. Fucosylation in the Fc domain, for example, affects immune cell activation by modulating affinity for Fc receptors (21). Sialylation has been shown to endow IgG with anti-inflammatory activity (22), but that cannot explain our findings, as none of the antibodies used contained sialic acid. Dynamic Fab arm exchange therefore represents a novel type of posttranslational modification, which serves as an additional mechanism for generating anti-inflammatory activity. The mechanism by which IgG4 Fab arm exchange occurs in vivo likely requires the reducing environment in blood or at cell surfaces to facilitate the breaking of inter-heavy chain disulfide bonds located in the hinge region. Indeed, the addition of reducing compounds, such as GSH, to purified IgG4 alone was sufficient to induce in vitro Fab arm exchange. GSH, present in all cell types, may well perform this role in vivo, and so additional cofactors, chaperones, or receptors, as hypothesized for PDI and FcRn previously (10), may therefore not be essential. An important second antibody heavy chain interface is located between the C<sub>H</sub>3 domains, which we show to be critically involved in Fab arm exchange. Elucidating the contribution of specific C<sub>H</sub>3-domain amino acid contacts to the mechanism of this reaction requires further investigation.

We show that IgG4 molecules acquire two distinct Fab arms by Fab arm exchange, which, when derived from polyclonal plasma IgG4, are (usually) directed against unrelated antigens. Fab arm exchange, furthermore, is dynamic, and combinations of certain specific Fab arms are therefore only expected to exist transiently. IgG4 molecules thereby lose their ability to cross-link antigen and to form immune complexes under most conditions. IgG4, often induced by chronic antigen stimulation, then may interfere with immune complex formation by other antibody isotypes and may dampen inflammatory reactions. In specific immunotherapy with allergen in allergic rhinitis, for example, increases in allergen-specific IgG4 levels indeed correlate with clinical responses (7, 23). A first proof of concept using monoclonal antibodies indicates that the formation of non-cross-linking IgG4 antibodies in vivo provided protection against a pathogenic antibody in experimental autoimmune MG.

Our results have an impact on immunotherapy with IgG4 monoclonal antibodies. The in vivo instability and dynamics of IgG4

introduce unpredictability, which is undesirable for human immunotherapy. Future studies should address the contribution of IgG4 Fab arm exchange to in vivo activity of therapeutic IgG4 monoclonal antibodies.

In summary, antibodies of the IgG4 isotype are shown to be dynamic molecules, undergoing Fab arm exchange in vivo and in vitro. The ability to engage in Fab arm exchange appears to be an inherent feature of IgG4 that involves the third constant domain in addition to the hinge region and that only requires a reducing environment to be activated. This novel protein modification challenges the commonly accepted one antibody-one antigen paradigm and redefines our thinking about the role of IgG4 in antibody-mediated immunity and the application of IgG4 monoclonal antibodies to immunotherapy.

#### References and Notes

1. F. M. Burnet, *Aust. J. Sci.* **20**, 67 (1957).
2. G. M. Edelman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **63**, 78 (1969).
3. E. O. Saphire *et al.*, *J. Mol. Biol.* **319**, 9 (2002).
4. J. Schuurman, G. J. Perdok, A. D. Gorter, R. C. Aalberse, *Mol. Immunol.* **38**, 1 (2001).
5. J. S. van der Zee, P. van Swieten, R. C. Aalberse, *Clin. Exp. Immunol.* **64**, 415 (1986).
6. R. Jefferis, in *The Human IgG Subclasses*, F. Shakib, Ed. (Pergamon Press, Oxford, 1990), pp. 15–30.
7. J. S. van der Zee, P. van Swieten, R. C. Aalberse, *J. Immunol.* **137**, 3566 (1986).
8. R. A. Margni, R. A. Binaghi, *Annu. Rev. Immunol.* **6**, 535 (1988).
9. J. Schuurman *et al.*, *Immunology* **97**, 693 (1999).
10. R. C. Aalberse, J. Schuurman, *Immunology* **105**, 9 (2002).
11. B. J. Scallon *et al.*, *J. Immunother.* **29**, 351 (2006).
12. Materials and methods are available as supporting material on Science Online.
13. R. C. Aalberse, J. Schuurman, R. van Ree, *Int. Arch. Allergy Immunol.* **118**, 187 (1999).
14. For calculations, see supporting online material. A maximum of 50% of IgG4-Betv1 half-molecules are

- incorporated in bispecific antibodies if equal amounts of IgG4 are mixed and if exchange is stochastic. The term stochastic indicates that exchange occurs without preference and results in complete mixing at equilibrium.
15. We prefer to describe the classical bivalency of IgG as monospecific cross-linking activity.
16. R. P. Junghans, C. L. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5512 (1996).
17. J. L. Teeling *et al.*, *Blood* **104**, 1793 (2004).
18. W. K. Bleeker *et al.*, *J. Immunol.* **173**, 4699 (2004).
19. M. Losen *et al.*, *Brain* **128**, 2327 (2005).
20. Y. F. Graus *et al.*, *J. Immunol.* **158**, 1919 (1997).
21. R. L. Shields *et al.*, *J. Biol. Chem.* **277**, 26733 (2002).
22. Y. Kaneko, F. Nimmerjahn, J. V. Ravetch, *Science* **313**, 670 (2006).
23. A. Nanda *et al.*, *J. Allergy Clin. Immunol.* **114**, 1339 (2004).
24. We thank P. de Heer, J. Scheik, A. Ortiz Buijsse, A. Verwilligen, and G. Perdok for technical assistance; H. de Vrieze for providing patient sera; J. Werkman for his help with the generation of the IgG mutants; K.V. Toyka, H. Brok, and B. 't Hart for their help with the rhesus monkey model; J. Bakker for graphics; and S. Ruuls for critically reviewing the manuscript. This work was supported by grants from the Prinses Beatrix Fonds, L'Association Française contre les Myopathies, a Marie Curie Fellowship, project grant QLG3-CT-2001-00225 of the European Community program "quality of life and management of living resources," the European Union Sixth Framework Programme (FP6) MYASTAID LSMM-CT-2006-037833, Joghem van Loghem Foundation, and Genmab. Genmab is an international biotechnology company that creates and develops human antibodies for immunotherapy. Genmab has partnered antibody products with Amgen, Roche, and GlaxoSmithKline. Authors specified below have a financial interest in Genmab. J.S., W.K.B., L.W., T.V., J.G.J.W., and P.W.H.I.P. have stock and/or warrants; L.A.A. and R.C.A. received consulting fees.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5844/1554/DC1  
Materials and Methods  
Figs. S1 to S6  
Table S1  
References

3 May 2007; accepted 8 August 2007  
10.1126/science.1144603

## Functional Architecture and Evolution of Transcriptional Elements That Drive Gene Coexpression

Christopher D. Brown,<sup>1</sup> David S. Johnson,<sup>1</sup> Arend Sidow<sup>1,2,\*</sup>

Transcriptional coexpression of interacting gene products is required for complex molecular processes; however, the function and evolution of cis-regulatory elements that orchestrate coexpression remain largely unexplored. We mutagenized 19 regulatory elements that drive coexpression of *Ciona* muscle genes and obtained quantitative estimates of the cis-regulatory activity of the 77 motifs that comprise these elements. We found that individual motif activity ranges broadly within and among elements, and among different instantiations of the same motif type. The activity of orthologous motifs is strongly constrained, although motif arrangement, type, and activity vary greatly among the elements of different co-regulated genes. Thus, the syntactical rules governing this regulatory function are flexible but become highly constrained evolutionarily once they are established in a particular element.

Gene products that are involved in the same molecular process must be coordinately expressed. Transcriptional coexpression is achieved by regulatory proteins and

their target cis-regulatory elements that promote gene transcription in overlapping spatiotemporal distributions (1–7). The function of a cis-element is encoded in its molecular architecture: a nucle-

otide sequence with instantiations (motifs) of the binding sites (motif types) for one or more transcription factors arranged with functionally significant motif combinations, orientations, or spacing. We address the molecular architecture of cis-elements driving gene coexpression and how the sequence and function of such elements evolves.

Because motifs are the functional units within a cis-element, analysis of a cis-element's molecular architecture and evolution requires an experimental system that allows quantification of each motif's activity, in a sufficiently large number of coexpressed genes whose functions have been maintained throughout evolution. The urochordate *Ciona* harbors such a system in the form of 19 genes that are coexpressed in the 36 muscle cells of the developing embryo (6). Of these 19 genes, 17 function in the same macromolecular complex, underscoring the requirement for tight coexpression. The genes include six single-copy loci from *C. savignyi* and their six orthologs in the sister species, *C. intestinalis*:  $\alpha$ -tropomyosin 1 (AT1),  $\alpha$ -tropomyosin 2 (AT2), myosin binding protein (MBP), troponin I (TI), troponin T (TT), and creatine kinase (CK). The remaining seven genes comprise two or three paralogs each of the multicopy gene families muscle actin (MA), myosin light chain (MLC), and myosin regulatory light chain (MRLC) from *C. savignyi* (fig. S1). The three motifs that mediate muscle-specific transcription in *Ciona* in general, and of these loci in particular, are the cyclic adenosine 5'-monophosphate response element (CRE) (6, 8), the MyoD motif (9–11), and the Tbx6 motif (12) (fig. S2).

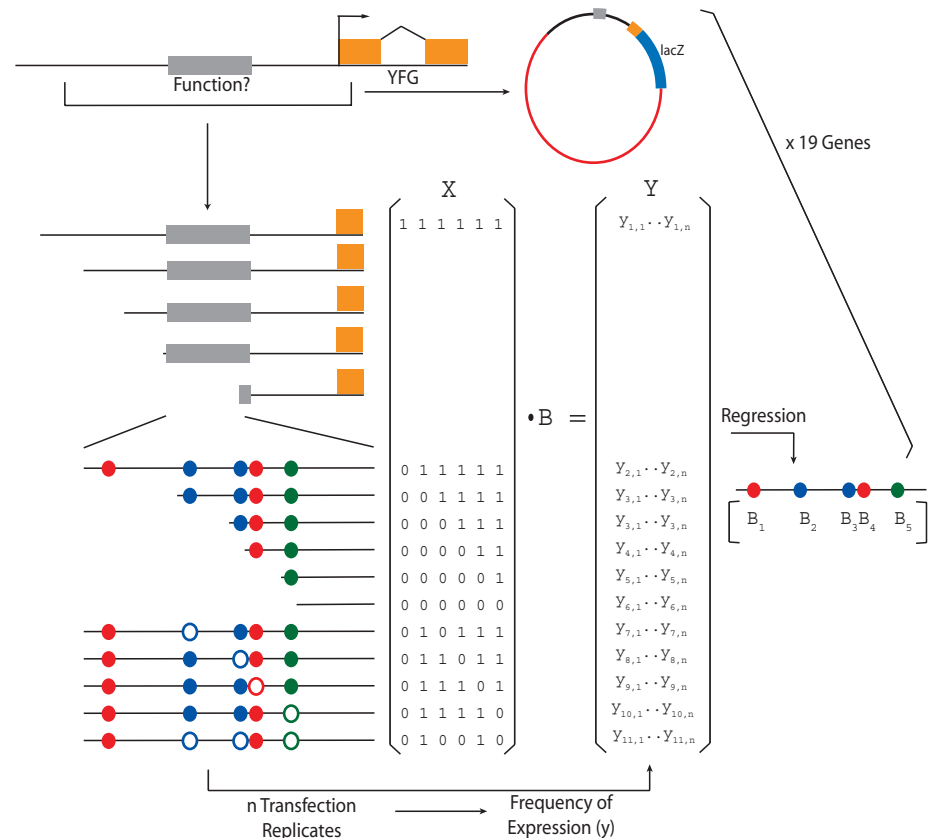
We investigated the functional architecture of the 19 cis-elements by a comprehensive mutagenesis effort coupled with a whole-embryo expression assay (13–15). Each reporter construct harboring specifically mutagenized sequences was transfected into hundreds of developing embryos. Activity of a mutagenized element was measured as the percentage of muscle cells expressing the reporter, which we show to correlate with average transcript levels [fig. S4 and supporting online material (SOM) text, section 3]. A first few hundred constructs, assayed in over 2000 transfections, defined the cis-elements responsible for the majority of function of each locus (Fig. 1 and table S1). We then dissected each cis-element using 220 constructs with small deletions [5 to 10 base pairs (bp)] or site-directed mutants that removed putative motifs in isolation or different combinations. Our quantitative results are based on 1237 transfections (five biological replicates per construct), which yielded a total of 85,506 transgenic embryos (table S2).

A quantitative and biologically meaningful representation of the functional architecture of each cis-element required an analysis framework to estimate the activity of each motif. To choose a framework, we needed to assess the relative importance of genetic interactions between motifs. Using the subset of the data that was appropriate for interaction analyses, we determined that most of the cis-elements examined functioned with little epistasis (SOM text, section 4, and fig. S5). Further evidence for motif independence was obtained by motif substitution experiments (SOM text, section 5, and fig. S6). This indicated that we could use multivariate regression (16) as the analytical framework to quantify motif activity.

Model predictions approximate the observed data well and are robust to several analytical scenarios (SOM text, section 6, fig. S7, and tables S1 and S3). After considering the advantages and drawbacks of these scenarios, we chose to continue our analyses with additive regression models (SOM text, section 6). These models explain 30 to 89% (mean 67%) of the variance of expression at each element (table S1), again underscoring that genetic independence explains most of the data well. Consistent with additivity, we express motif function in “expression frequency

units” (efu), meaning that a motif with an inferred activity of  $x$  efu increases by  $x$  the percentage of muscle cells in which expression is detected (SOM text, section 3). The mean per-element fraction of activity attributed to the motifs is 83%. The 77 motifs affect element activity by  $-0.14$  to  $0.45$  efu (Fig. 2A), with 39 motifs having significantly nonzero activating function (partial regression coefficient  $t$  statistic,  $P < 0.05$ ).

Having obtained quantitative estimates of motif activity, we examined each element's functional architecture. Apart from the obvious clustering of functional motifs, we were unable to discern any features (such as spacing, order, and relative orientation of motifs) that might explain the functions of individual motifs or of the elements as a whole, which would have shed light on organizational principles of regulatory elements. Indeed, there is notable heterogeneity among the loci: Elements are built from motifs of widely varying activity, from different combinations of motif types, and in diverse arrangements (Fig. 2B). For example, the cis-element at CK spans 31 bp and consists of one intermediate and one strong Tbx6 motif, whereas the AT1 cis-element consists of two weak CRE motifs, followed by two intermediate Tbx6 motifs and



**Fig. 1.** Experimental design. For each of 19 loci, initial deletion series (truncated lines) locate regions of concentrated function (gray bar). Fine-scale deletions and site-directed mutations (open circles) target putative motifs (solid circles). Set of constructs is represented as matrix X of categorical explanatory variables (ones and zeroes) whose replicated transfections yield expression measurements (matrix Y of  $y_{n,i}$ ). Functional contributions of individual motifs ( $B_1$  to  $B_5$ ) are estimated with regression models. YFG, your favorite gene.

<sup>1</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA.

\*To whom correspondence should be addressed. E-mail: arend@stanford.edu

a strong MyoD motif, across 35 bp. Although motif independence is prevalent, elements do somewhat differ in how much genetic interaction exists. At MBP and AT2, for example, the additive model explains the data very well with high correlations between the predictions and the actual data ( $r^2_{Cs-MBP} = 0.83$ ,  $r^2_{Cs-AT2} = 0.77$ ; tables S1 and S3) and with little function unexplained by the model. Function at MA1, by contrast, is not described as well by models without interactions (table S3).

Conceivably, the heterogeneous regulatory architectures specify subtle differences in expression pattern or timing during developmental stages or physiological conditions not assayed here. It is clear, however, that the genes' tight coexpression in the embryonic tail muscle is achieved by a common and restricted set of three transcription factors acting upon vastly different cis-element architectures, the diversity of which defies the expectation that commonalities in design underlie co-regulation.

In stark contrast to the apparent flexibility of regulatory architecture, we observed little change in motif activity, order, or composition between orthologous elements of *C. intestinalis* and *C. savignyi*. (Neutral sequence divergence between the two *Ciona* species is approximately equivalent to that between mammals and birds, ruling out the possibility that these sequences have not been afforded enough time to accumulate change.) At single-copy genes, 26 of the 27 motifs with statistically significant activity have a clearly orthologous counterpart. Orthologous motifs drive very similar, in many cases indistinguishable, amounts of activity (Fig. 3A). For example, both MBP orthologs are regulated by a strong MyoD, a weak Tbx6, and a weak CRE motif, with less than 0.039 efu average deviation in individual motif activity. In total, the activity of orthologous regulatory motif pairs is highly correlated between the two species (Spearman's  $\rho = 0.61$ ,  $P < 0.005$ ; Fig. 3B). Thus, co-regulated gene expression at these loci has been main-

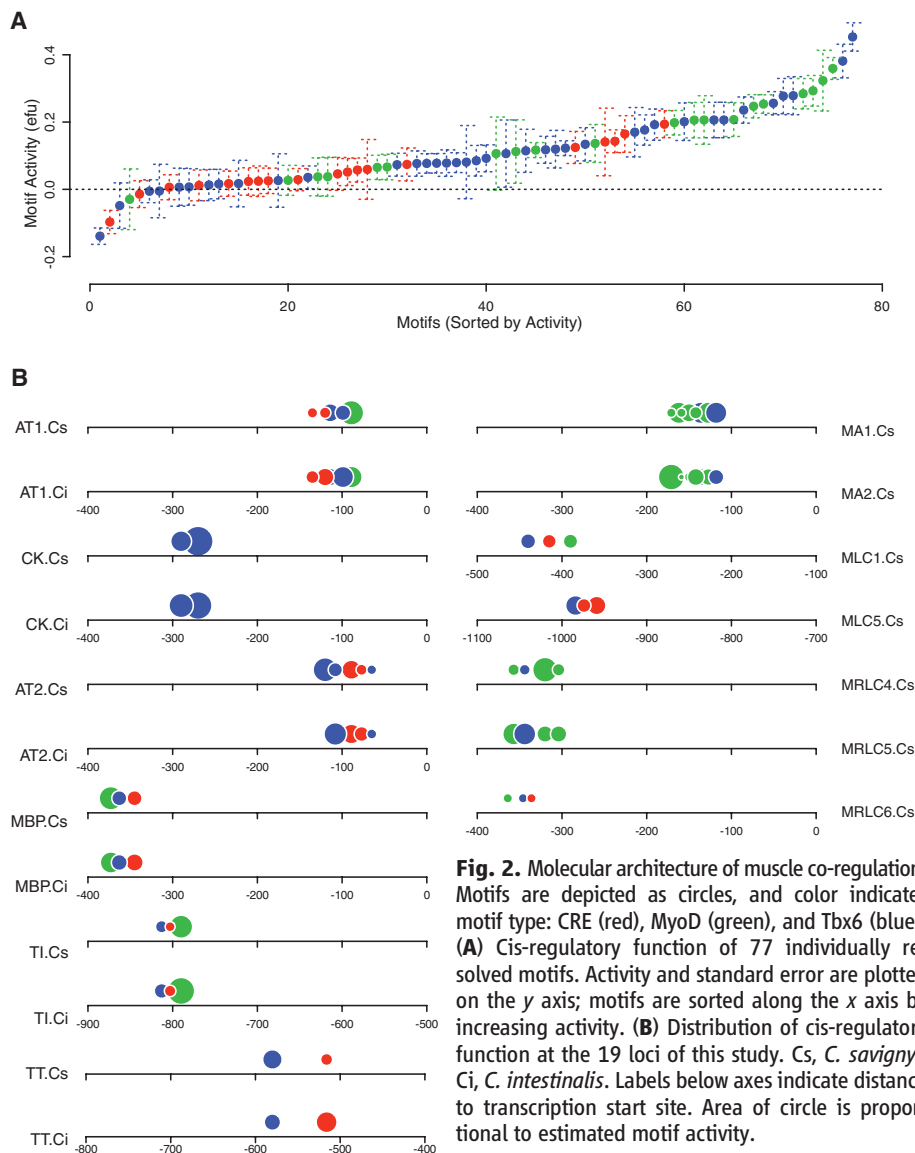
tained by conserving the locus-specific ancestral cis-regulatory architectures, with purifying selection tolerating little functional flexibility.

Strong constraint is evident at the sequence level as well. Functional regulatory motifs exhibit far fewer substitutions than the genome-wide average ( $P < 3.8 \times 10^{-10}$ ) (SOM text, section 7, and fig. S8A). The pairwise identity between orthologous functional motifs is 79%, whereas the genome-wide background identity is  $<20\%$  (including insertions and deletions). Sequence identity markedly drops off outside the boundaries of the functional motifs (Spearman's  $\rho = -0.57$ ,  $P < 0.05$ ), reaching genome-wide background levels within 12 bp (fig. S8B). Finally, there is less population genetic variation in the motifs than on average in the genome (17) (SOM text, section 7, and fig. S10). This demonstrates that the motif sequences are subject to far greater evolutionary constraint than flanking sequence and that the functional motifs themselves are the units maintained by purifying selection.

The sequence changes that have occurred are not distributed evenly among the orthologous motifs, with functionally strong motifs having accumulated fewer substitutions than weak motifs. Notably, motif activity is significantly correlated with percent identity (Spearman's  $\rho = 0.35$ ,  $P < 0.05$ ). This is likely due to strong regulatory motifs being responsible for a larger fraction of the total function of a regulatory element; substitutions in them will therefore result in greater phenotypic consequences and be subject to stronger levels of purifying selection, decreasing the evolutionary rate of the element.

Further emphasizing that the activity of an element is controlled tightly by evolutionary pressures is a case of compensatory evolution in AT2. Two Tbx6 motifs (Fig. 3, A and C) are functionally strong in one species and weak in the other, in a complementary pattern. These activity differences correlate with substitutions away from or toward the motif consensus (Fig. 3C). The functional differentiation of these cis-regulatory motifs did not result from motif gain or loss but from base substitutions that modified motif activity.

Comparison of the cis-regulatory architectures of the three groups of paralogs from *C. savignyi* presents a strong contrast to the highly constrained orthologous cis-elements. The paralogous architectures show a high degree of differentiation in the form of element and motif-level sequence turnover as well as functional divergence of well-aligned motifs (Figs. 2B and 3D). Thus, whereas purifying selection acting on orthologous motifs in single-copy genes is strong enough to maintain conservation of regulatory motif sequence and function over long evolutionary distances, the paralogous motifs of clusters of genes that encode the same protein exhibit far greater rates of turnover. We speculate that this greater flexibility in elements of clustered multicopy genes is tolerated because changes in



**Fig. 2.** Molecular architecture of muscle co-regulation. Motifs are depicted as circles, and color indicates motif type: CRE (red), MyoD (green), and Tbx6 (blue). **(A)** Cis-regulatory function of 77 individually resolved motifs. Activity and standard error are plotted on the y axis; motifs are sorted along the x axis by increasing activity. **(B)** Distribution of cis-regulatory function at the 19 loci of this study. Cs, *C. savignyi*; Ci, *C. intestinalis*. Labels below axes indicate distance to transcription start site. Area of circle is proportional to estimated motif activity.

the activity of one element have a small effect on the total function of the cluster.

The two most prominent developmental mechanisms that build a multicellular organism are pattern formation and cellular differentiation. Previous studies of regulatory architecture and evolution were conducted in pattern formation systems, either by leveraging sequence comparisons and broad functional genomic data (18–20) or by studying a single regulatory element in detail (21–24). By contrast, we dissected the regulation of coexpression during cellular differentiation and introduced a quantitative framework for targeted experimental analysis of motif function. Using the *Ciona* muscle system, we

demonstrated that coexpression is driven by regulatory motifs of broadly varying activity assembled into a diverse array of cis-elements. Despite this flexibility in cis-regulatory architecture, motif-level sequence and function are exquisitely maintained in distantly related orthologs. Thus, whereas a diversity of cis-regulatory architectures can generate nearly identical phenotypic outputs, the fitness landscapes separating them appear to be sufficiently rugged to constrain their evolution (25).

Our findings have implications for understanding genetic variation in such co-regulatory systems. Polymorphisms in cis-elements will range in phenotype, depending on the amount

of activity that the affected motif contributes to the function of its element; the most direct evidence for this view is the wide range of effects on cis-element function by the individual motif mutants we tested. Similarly, a polymorphism in a trans-acting factor will not affect expression of all targets equally but will instead have a target-specific effect whose magnitude is determined by the architecture of the target's cis-element. These conclusions highlight the challenges that lie ahead for interpretation of genetic variation in gene regulatory systems, including those of vertebrates, *Ciona*'s most advanced close relatives.

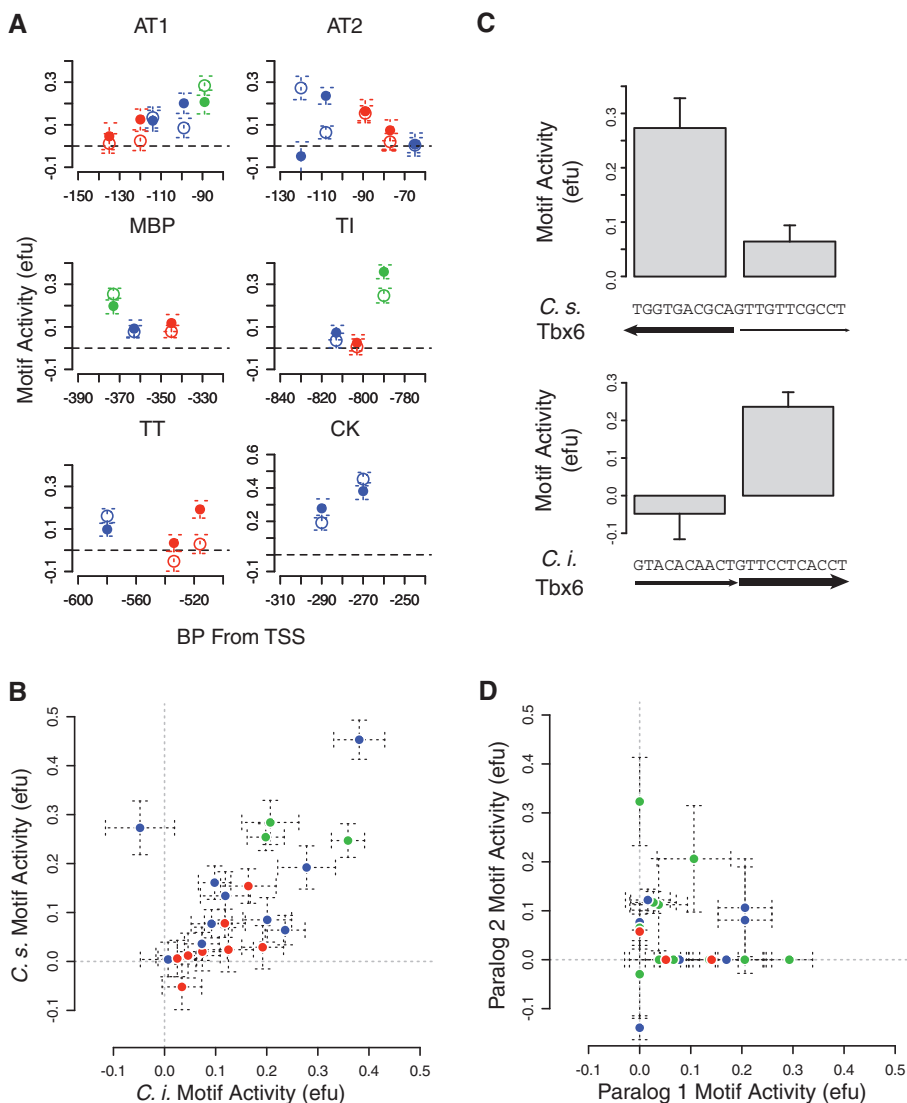
References and Notes

1. E. H. Davidson, *Genomic Regulatory Systems* (Academic Press, San Diego, CA, 2001).
2. Y. Pilpel, P. Sudarsanam, G. M. Church, *Nat. Genet.* **29**, 153 (2001).
3. B. P. Berman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 757 (2002).
4. Q. Zhou, W. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12114 (2004).
5. E. Segal *et al.*, *Nat. Genet.* **34**, 166 (2003).
6. D. S. Johnson *et al.*, *Genome Res.* **15**, 1315 (2005).
7. A. Tanay, A. Regev, R. Shamir, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7203 (2005).
8. T. Kusakabe, R. Yoshida, Y. Ikeda, M. Tsuda, *Dev. Biol.* **276**, 563 (2004).
9. T. K. Blackwell, H. Weintraub, *Science* **250**, 1104 (1990).
10. D. S. Johnson, B. Davidson, C. D. Brown, W. C. Smith, A. Sidow, *Genome Res.* **14**, 2448 (2004).
11. T. H. Meedel, P. Chang, H. Yasuo, *Dev. Biol.* **302**, 333 (2007).
12. K. Yagi, N. Takatori, Y. Satou, N. Satoh, *Dev. Biol.* **282**, 535 (2005).
13. Materials and methods are available as supporting material on Science Online.
14. R. W. Zeller, M. J. Virata, A. C. Cone, *Dev. Dyn.* **235**, 1921 (2006).
15. R. W. Zeller, *Methods Cell Biol.* **74**, 713 (2004).
16. R. R. Sokal, F. J. Rohlf, *Biometry* (Freeman and Co., New York, NY, 1995).
17. K. S. Small, M. Brudno, M. M. Hill, A. Sidow, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5698 (2007).
18. E. T. Dermitzakis, C. M. Bergman, A. G. Clark, *Mol. Biol. Evol.* **20**, 703 (2003).
19. A. M. Moses *et al.*, *PLoS Comput. Biol.* **2**, e130 (2006).
20. J. Zeitlinger *et al.*, *Genes Dev.* **21**, 385 (2007).
21. C.-H. Yuh, H. Bolouri, E. H. Davidson, *Science* **279**, 1896 (1998).
22. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, *Nature* **403**, 564 (2000).
23. V. Bertrand *et al.*, *Cell* **115**, 615 (2003).
24. I. Oda-Ishii *et al.*, *Development* **132**, 1663 (2005).
25. S. Wright, in *Proceedings of the Sixth Annual Congress of Genetics*, D. F. Jones, Ed. (Genetics Society of America, Austin, TX, 1932), vol. 1, pp. 356–366.
26. We thank G. Barsh, P. Collins, G. Cooper, J. Davis, J. Dean, M. Hosobuchi, R. Myers, C. Smith, and E. Stone for helpful insights, and K. Perez and E. Anton for expert technical assistance. This work was supported by an NSF Graduate Research Fellowship (C.D.B.), by grants from the NIH/National Human Genome Research Institute (NHGRI) and NIH/National Institute of General Medical Sciences (A.S.), and the Stanford Genome Training Program (NIH/NHGRI) (C.D.B. and D.S.J.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5844/1557/DC1  
 Materials and Methods  
 SOM Text  
 Figs. S1 to S11  
 Tables S1 to S3  
 Marker (Java image annotation tool)  
 References

31 May 2007; accepted 19 July 2007  
 10.1126/science.1145893



**Fig. 3.** Evolution of motif activity. (A) Motif-level distribution of regulatory activity at six orthologous gene pairs. Distance from transcription start site and motif activity are plotted along the x and y axes, respectively. Open and solid circles represent individually resolved *C. savignyi* and *C. intestinalis* motifs, respectively. Colors are same as in Fig. 2. (B) Conservation of orthologous motif activity. *C. intestinalis* and *C. savignyi* motif activity plotted against each other. (C) Compensatory evolution of AT2 regulatory elements for *C. savignyi* (top) and *C. intestinalis* (bottom). Arrow direction and thickness represent Tbx6 motif orientation and strength of match to its position-specific scoring matrix. Bar plots depict activity of each Tbx6 motif, as estimated from additive regression models. (D) Functional turnover in paralogous motifs. Plotted as in (B). Error bars in all panels depict standard error.