Title

Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development

Journal

Annual Review of Genomics and Human Genetics, 16(1)

ISSN

1527-8204

Authors

Rebeiz, Mark
Patel, Nipam H
Hinman, Veronica F

Publication Date

2015-08-24

DOI

10.1146/annurev-genom-091212-153423

Peer reviewed

Unraveling the tangled skein: the evolution of transcriptional regulatory networks in development

Mark Rebeiz[1*], Nipam Patel[2], Veronica Hinman[3]

Keywords: gene regulation, morphological evolution, gene regulatory network, anatomical novelty

[1] Department of Biological Sciences, University of Pittsburgh, PA, USA

[2] Department of Molecular & Cell Biology and Integrative Biology, University of California, Berkeley, CA, USA

[3] Department of Biological Sciences, University of Pittsburgh, PA, USA

[*]Correspondence: rebeiz@pitt.edu, nipam@uclink.berkeley.edu, veronica@cmu.edu

ABSTRACT (150 words max) (its at 131 now)

The molecular and genetic basis of the evolution of anatomical diversity is a major question that has inspired evolutionary and developmental biologists for decades. Because morphology takes form during development, a true comprehension of how anatomical structures evolve requires an understanding of the evolutionary events that alter developmental genetic programs. Vast gene regulatory networks (GRNs) that connect transcription factors to their target regulatory sequences control gene expression in time and space, and therefore determine the tissue specific genetic programs that shape morphological structures. In recent years, many new examples have greatly advanced our understanding of the genetic alterations that modify GRNs to generate newly evolved morphologies. Here, we will review several aspects of GRN evolution, including their deep preservation, mechanisms of alteration, and how GRNs originate to generate novel developmental programs.

TABLE OF CONTENTS

n. Adding genes to a network: the evolution of new expression patterns through the origination of enhancer sequences
o. Influences of enhancer origination mechanisms on network shape

I. What are gene regulatory networks, and how do they control development?

While trees (38) and landscapes (158) are widely used metaphors in evolutionary biology, networks serve as a driving analogy to conceptualize developmental genetics (39). Some of the first uses of the network metaphor were introduced to help visualize the context dependent regulation of enzyme expression in bacteria (59). In higher eukaryotic systems, perhaps the clearest articulation was a seminal review by Britten and Davidson (23). At their core, gene regulatory networks (GRNs) drive development through differential gene expression: despite identical DNA in every cell, only a subset of the genes in the genome is activated at any one time or place during development. This is exemplified by the specific expression of transcription factors in broad zones of the *Drosophila* embryo (140), and the corresponding absence or disruption of these regions in mutant embryos that lack these factors (107). Such transcription factors serve as leaders in the chain of command: as combinations of particular factors are expressed in restricted patterns, downstream genes are consequently activated, which eventually leads to the activation or repression of genes that are directly responsible for conferring cellular behavior, such as growth, migration, shape, adhesion, elasticity, etc. These subordinate genes of the network are generally thought to represent the network's periphery – they are usually unable to further influence transcriptional events in the network. Therefore, while nearly every transcription factor is an "**internal node**", they ultimately regulate the activity of "**terminal nodes**" (Figure 1).

Each gene in a network is connected through its cis-regulatory transcriptional control sequences

To participate in a particular GRN, a gene must contain a *cis*-regulatory apparatus that binds transcription factors of the network to activate expression in the zone of interest. In general, the recruitment of a single transcription factor is insufficient to ignite the downstream expression of target genes: **activator synergy** dictates a requirement for more than one activating transcription factor to bind. These factors bind to a special type of transcriptional regulatory sequence, variably referred to as **enhancers**, cis-regulatory modules (CRMs), or cis regulatory elements (CREs). Enhancers are generally 200-1,000 basepairs in length and contain multiple short stretches of DNA to which particular transcription factors specifically dock. Multiple protein:DNA complexes comprise a **combinatorial logic** of activators and repressors that sculpt a limited pattern of expression (83). The pattern of gene expression is dictated by the spatial distribution of the transcription factors that bind the enhancer. Hence, deciding as a committee, multiple transcription factors present in a particular cell type cooperate to activate target enhancer sequences, and drive expression of subordinate genes of the network.

One of the first examples of an enhancer that was demonstrated to integrate multiple positive and negative-acting transcription factor binding sites regulates the expression of the *even-skipped* gene in the early *Drosophila* embryo (Figure 2) (138). *even-skipped* encodes a transcription factor expressed in seven stripes that functions to set up segmented zones of the embryo (91). Several distinct modules of regulatory DNA in the non-coding sequences that surround the *even-skipped* coding unit each drive one or two of the seven stripes of *eve* expression (64). The “stripe 2 enhancer” contains binding sites for spatially restricted activators and repressors that combinatorially function to generate a discrete pattern of activation in the second stripe (Figure 2). Because

individual modules require the docking of multiple activating transcription factors in close proximity on the DNA, they are thought to act relatively independently, or exhibit **modularity**: the action of one enhancer is generally assumed to not interfere with the function of other non-overlapping modules.

Connecting nodes with genetic and biochemical tests of network hierarchy

There are several experimental methods that have been used to establish the hierarchical relationships between genes that form our picture of gene regulatory networks (Table 1). These range from simple tests of genetic epistasis (9) to the direct biochemical confirmation that a transcription factor directly binds to a given segment of regulatory DNA (138). A classical approach is to monitor the expression of a potential downstream gene in an animal that is mutant, or otherwise deficient for another gene in the network (30, 123). Such genetic tests can establish that one gene lies upstream of another in the regulatory network, but cannot distinguish whether the interaction is direct (i.e. the factor in question binds to the enhancer of the downstream gene) or indirect (e.g. the factor activates one or more downstream genes that ultimately regulate the gene in question). To determine whether a relationship in a network is direct, one must demonstrate the direct binding of that factor to the regulatory sequence responsible for the given pattern of gene expression. This can be done through a **gel-shift assay** *in vitro*, or using *in vivo* approaches such as **chromatin immunoprecipitation**. The demonstration of direct binding is usually complemented by a functional **reporter assay**, in which the identified binding site is mutated, and shown to affect the tissue-specific activation of the regulatory sequence. However, although these tests are the "gold

standard" for drawing direct connections in gene regulatory hierarchies of model organisms (e.g. fly, mouse, nematode, sea urchin) such tests are labor-intensive, and are not always possible, especially in an evolutionary context that involves non-model species.

Integration of multiple GRN tiers can generate complex and dynamic logical outputs

Gene expression, especially during early development, can be highly dynamic: transcripts can appear, recede, and reappear in different territories all within a few cell divisions. Enhancer modules must therefore interpret a **trans-regulatory landscape** that itself may be highly dynamic, to provide a complex temporal and spatial output of expression. One example of how such precise execution of dynamic expression is encoded in regulatory DNA derives from the sea urchin *foxa* gene locus (43). *foxa* is first expressed within the domain that is fated to become endomesoderm (i.e. both endoderm and non-skeletogenic mesoderm), but within a two hour window, expression is precisely extinguished from future non skeletogenic mesoderm cells, while expression persists exclusively in the outer tier of cells fated to become endoderm (115) (Figure 3A). Shortly thereafter, expression levels in this territory markedly increase (43). This change in gene expression is necessary for the segregation of endoderm and mesoderm in the sea urchin larva, and Foxa is a critical driver of the GRN for endoderm formation (115, 134). Genetic tests that perturbed endomesoderm formation using morpholinos had predicted the GRN that regulates *foxa* expression (114, 134) (Figure 3B). The Tcf-β-catenin complex and Notch effector Suppressor of Hairless [Su(H)], as well as the homeobox factor, Hox11/13b, initiate *foxa* expression throughout the endomesoderm lineage. β-

catenin is then cleared from the inner most cells through a Delta mediated signal, received from the adjacent micromeres (131). This results in the recruitment of the Groucho repressor to Tcf, which inactivates *foxa* in this inner-most tier of cells.

The cis regulatory apparatus that implements this dynamic pattern of expression is spread among multiple modules distributed across 20 kilobases surrounding the *foxa* locus (43). Employing a BAC vector system to untangle the regulatory control region of this gene, de-Leon and Davidson characterized four modules termed F, I, J, K in the upstream and downstream regions of *foxA* (Figure 3C). While each module executes a particular function, they control expression by interacting with the basal promoter in combination. Thus, Module F responds to Tcf, module J to Su(H) and module K to Hox11/13b. As Tcf associates with the Groucho co-repressor, module F then acts as a toggle switch, repressing *foxa.* In the presumptive endoderm, module K then also receives an input from Otx, while module I is activated by Brachyury. These new inputs function to ramp up *foxa* expression. This system illustrates how beyond encoding multiple tissue-specific domains of expression, modules can cooperate to control highly dynamic expression in a single territory.

Casting developmental evolution in terms of network biology

The intricately organized nature of gene regulatory networks naturally inspires the question of how their complexity evolves: *how do new phenotypes manifest through the alteration of networks? How do new networks arise?* Indeed, because development is proximately controlled at the upper-tiers by the spatial restriction of regulatory genes such as transcription factors and signaling pathway components, this framework lends

itself naturally to the study of evolutionary development. Every mutation occurs somewhere in a gene regulatory network: if a change occurs in a transcription factor that patterns the early embryo, this modifies the upper tiers of a network. The resulting change in the *trans*-regulatory landscape would be expected to alter many subordinate genes in the GRN, potentially causing drastic phenotypic differences. If the change resides in a gene that has no downstream regulatory connections (a terminal node) such as actin or myosin, one would expect that while there may be phenotypic consequences, the effect on the network would be minimal. Because the early phases of development depend so critically on the establishment of specific expression patterns, evolutionary change often occurs in the expression pattern generating machinery of *cis*-regulatory sequences. Such mutations change the expression of the gene in question, having relatively few **pleiotropic effects** (i.e. phenotypic consequences in other tissues). In the examples that follow, we discuss the deep evolutionary conservation of network components and circuits (section II), alterations at many levels of existing networks (section III), and the origins of networks that generate completely new developmental structures (section IV).

II. Deep conservation of network components and individual linkages

Ever since the first comparisons of gene coding sequences across animal phyla, it has been appreciated that many of the genes that pattern development are highly conserved. The small number of developmental signaling pathways used to generate differences in cell fate are for the most part, conserved across the animal kingdom. A core set of conserved transcription factors that sit at the termini of these signaling

pathways make signal-regulated decisions at enhancers (13). The major classes of transcription factor are highly conserved across animal phyla, appearing in even the most primitive of animals (139). Hence, the "developmental toolkit" (29) that builds the internal nodes of gene regulatory networks are deeply preserved in sequence. As we discuss below, these factors also often tend to maintain conserved roles during development.

Conserved roles of developmental patterning genes in evolutionarily distant homologous processes

One of the first, and perhaps still best characterized examples illustrating the unexpected conservation of developmental gene function comes from the Hox gene complex. The Hox transcription factors pattern body axial identity and are organized into gene clusters in a very similar way across bilaterians such that this gene complex was once described as the "Rosetta Stone" of animal pattern formation (137). However, many other examples of deeply conserved functional roles exist for the developmental toolkit. The transcription factor Distalless is expressed within the appendages of flies, mice, and everything in between (110). A conserved role for the Pax6/Eyeless transcription factor was characterized in vertebrate and invertebrate eye formation (122). The brain-patterning "proneural" transcription factors related to *Drosophila* Acaete-Scute are expressed in, and required for neuronal development across metazoa (71). These parallels have several implications for the evolution of GRNs: 1) they suggest that often, when a key role for a transcription factor evolves, it is preserved; 2) because transcription factors maintain conserved roles in gene regulatory networks, their DNA-binding specificities

tend to evolve slowly, or (as discussed in section III below maintain ancestral functionality; 3) the deep conservation of collections of transcription factors in certain developmental processes suggest that connections between regulatory factors often remain conserved within gene regulatory networks.

A large number of experiments in the 80's and 90's established that many of the genes of the developmental toolkit are **functionally equivalent**, and have been minimally altered during evolution. Tests of functional equivalence generally involve the mis-expression of a toolkit gene in a distantly related species, followed by a comparison to the phenotype generated by the endogenous species' gene. Surprisingly, many of these tests were successful. A key example of this was the expression of the mouse *Pax6* gene in *Drosophila*, which can generate ectopic eyes resembling those induced by its *Drosophila* homolog *eyeless* (62). While many examples of striking functional equivalence exist for developmental toolkit transcription factors (60, 61, 63, 92, 151), exceptions to this rule nevertheless exist (33, 53, 130). In Section III, we will discuss some established cases of transcription factor protein evolution.

The deep preservation of GRN subcircuits

Given that many of the transcription factors of the developmental toolkit have maintained similar DNA-binding specificities and functional roles in patterning organs or specifying particular cell types, one key question is to what degree this conservation extends to the underlying GRN. If protein-DNA linkages are conserved either at the DNA sequence or functional level, one could infer which portions of the GRN have been maintained, and what parts have changed (1). These studies must first determine if

orthologous genes are connected to each other by identical arrangements of regulatory linkages in distant taxa and if so, elucidate which developmental processes they control.

The GRN for specification of endomesoderm in the sea urchin is one of the most extensively studied networks, naturally forming a platform for comparison of GRN conservation and divergence (114, 115). Early studies of the evolutionary preservation of GRNs determined whether a similar topology could be detected in a distantly related echinoderm, the sea star (66, 96). Sea urchins and sea stars are representatives of two classes of the phylum Echinodermata that diverged around 400 million years ago (116). This comparison revealed that certain portions of the network (**subcircuits**) were indeed quite similar, in spite of an expansive span of time over which such linkages must have been maintained (66, 96). For instance, a subcircuit composed of four orthologous TFs (Gata4/5/6, Otx, Blimp1 and Foxa) were highly similar among these two taxa. Another subcircuit of four factors directing specification of mesoderm was also found to be highly similar (96). Strikingly, in both instances, the orthologous TFs were engaged in a **positive feedback loop**, e.g. Otx regulates the expression of *gata4/5/6* and in turn Gata4/5/6 regulates the expression of *otx*. The positive feedback loop is a commonly observed network motif present in many biological and non-biological systems that maintains activity at a node for an extended period (5). During development, we can envision that such positive regulatory feedback acts to reinforce gene expression states, particularly if the inputs are early and transient. Thus the conserved function may be to initiate, and then maintain the specification of cell types.

Studies of the GRN for the heart show an even deeper ancestry of conservation in network topology (37). The NK-2 transcription factor, named *tinman* in *Drosophila* due

to the lack of a heart in these mutants (19), shares a key position in the heart specification network of *Drosophila* and vertebrates (77, 84). In addition, the MEF2 transcription factor has a deeply ancestral role in specification of muscle cell identity, a role that was employed during the evolution of the heart (17). The vertebrate GATA4/5 and the homologous *Drosophila* Pannier factors also serve similar roles in these networks (6, 75, 104). These striking similarities in heart and endomesoderm GRNs, separated by 500 and 700 million years ago respectively, suggest that once these connections were established, their loss may have been absolutely disallowed. These subcircuits have been dubbed "kernels" by Davidson as an analogy to the part of a computer's operating system that is so crucial to the machine's function that it operates in a protected space of memory (40).

Deep conservation of individual transcription factor binding site sequences

In recent years, it has come to light that above and beyond a deeply preserved topology of certain GRNs, the connections between components of the network sometimes remain unchanged at the level of the DNA-sequences that bind individual transcription factors. **Phylogenetic footprinting** is a frequently employed tool to find important parts of regulatory elements (145). If a sequence serves an important transcription factor binding function, the logic follows that it may have withstood the erosion of evolutionary change, and thus its sequence would be conserved. This property of regulatory sequences has been successfully exploited to find many functional binding sites in regulatory DNA (124, 160). Over the past ten years, striking examples of sequence conservation that have delved progressively deeper into metazoan history have emerged.

Shortly after the generation of vertebrate genome sequences, a large number of studies documented the existence of :ultraconserved regions", stretches of DNA 100-200 basepairs in length that share an extraordinary level of sequence conservation between the genomes of rodents and humans, sometimes extending all the way to sequences in the genomes of fish (15, 113). *in vivo* reporter assays confirmed that many of these conserved elements represent transcriptional regulatory sequences (149), and yet paradoxically, mice engineered to lack some of these segments lacked obvious phenotypic consequences, possibly due to functional redundancy in the tested genomic regions (3). Utraconservation has proven to be a quite useful characteristic for inferring important evolutionary events (*see section on human evolution in Section III*). However, some studies have shown that regulatory sequence conservation between fish and humans may be just the tip of the iceberg.

For a small handful of cases, deep regulatory sequence conservation has been observed at the root of the bilaterian lineage (conserved between protostomes and deuterostomes). Two examples of this phenomenon have been documented for network connections in the nervous system GRN between the Notch signaling pathway, effectors of the Hairy/Enhancer of Split (HES) transcription factor family, and the proneural transcription factors that specify neuronal fates (125, 127). It was observed that a binding site matching a stringent 10 nucleotide definition for the Notch regulated HES factors was present upstream of both human and fly proneural genes of both the *atonal* and *achaete-scute* families (127). The invariant presence of this site close to the promoter, and conservation among alignable species allowed the authors to infer that this binding site most likely arose in the ancestral bilaterian proneural gene. Further, it was

discovered that the HES genes that target the proneural factors are themselves involved in a deeply conserved circuit consisting of a rigidly spaced pair of binding sites for the Notch regulated Su(H) transcription factor, combined with sites for the proneural factors present in both protostome and deuterostome lineages (125). Perhaps the most striking example of deep conservation of individual transcription factor binding sites is a conserved non-coding region present in the human SoxB2 gene locus, which is recognizable in the cnidarian, *Nematostella vectensis* (132). This remarkably conserved sequence was tested for regulatory activity in an impressive array of transgenic organisms (zebrafish, sea urchin, and *Drosophila*), and the common theme observed for each organism was that this DNA sequence drove reporter expression in the nervous system.

Hence, from the gross anatomy of network topology, to sequence conservation, and even individual binding sites encoded in the DNA, many aspects of GRN network architecture are deeply conserved. As we will see in the next section, the identification of these relationships provides a powerful platform for inferring major events during GRN evolution, particularly in the human lineage.

III. The evolutionary modification of GRNs

While the deep conservation of the developmental toolkit genes and their connections within metazoan networks highlights a remarkable degree of stability in some aspects of GRN architecture, these examples leave open the question of how GRNs change to generate diverse phenotypes. Here, we will illustrate examples in which different parts of a GRN's structure have been modified. These range from changes to the expression of terminal nodes to the alteration of high level factors in the network:

changing both patterns of expression as well as modifying DNA-binding specificity itself. We draw from examples that include both **microevolutionary** changes within populations or between closely related species, as well as **macroevolutionary** divergence that occurred in the distant past.

Modification of cis regulatory elements at the periphery of GRNs

One model system in which GRN evolution has been extensively studied is centered on the rapidly evolving pigment patterns that adorn the body parts of *Drosophila* species (154). With over 1,500 species in the genus *Drosophila,* there is ample variation in pigment patterning, including multiple examples of **intraspecific variation** (78, 106, 119, 146) and several cases of variation between closely related species that can be crossed for genetic mapping (28, 86, 156, 159). Many of the genes in the melanin synthesis pathway were discovered as some of the first genetic markers that occured as spontaneous mutants in the fly room of Thomas Hunt Morgan (21, 22, 105), over a century ago. Several of these enzymes are patterned in expression, correlating with the spatial patterns of pigments they are required to produce (69, 150, 155).

Perhaps the most well studied gene of the *Drosophila* pigmentation system is *yellow*, named for its coloration phenotype (105). *yellow* mutants exhibit a body-wide lightening of pigmentation on their wings, abdomen, and bristles (Figure 4A, B). Wing and abdomen expression of *yellow* is encoded by two separable enhancers in the gene's upstream regulatory region (57) (Figure 4C). This regulatory architecture also highlights where mutations underlying evolutionary shifts in *yellow* expression have occurred. Differences in gene expression may be caused by mutations to the regulatory apparatus of

the gene in question (a ***cis*** change), or could result from alterations in upstream regulators (a ***trans*** change) (153). A very simple way to distinguish between *cis* and *trans* mechanisms is to compare the activities of the given regulatory regions in a reporter assay within a common genetic background (128). If differences in expression are recapitulated by the enhancer sequences, then one can conclude that changes arose in *cis*. On the other hand, *trans* regulatory changes would result in an absent or poor recapitulation of the species-specific expression pattern.

Tests of the wing and abdominal enhancers of *yellow* have revealed a range of evolutionary shifts in these regulatory sequences (Fig. 4D). For example, mutations inactivating the abdominal enhancer were established in *D. kikkawai* (70), while its inactivity in *D. santomea* was attributed to changes upstream (69, 70). These changes are not only limited to instances of trait loss; the expansion of abdominal pigmentation was associated with *cis* changes in the abdominal enhancer of *yellow* (109). A stunning case of trait gain was observed in the wing spots of *D. biarmipes,* which have been attributed to a complex assembly of changes both in *cis* (58), combined with a *trans* change which altered the expression of the transcription factor Distalless (7). It is important to note that modification of *yellow* expression alone cannot solely account for any of these phenotypes (for example, in the absence of *yellow* gene function strong patterns of pigmentation still form). However, these cases have provided key examples illustrating *cis* regulatory mutations contributing to a morphological phenotype.

Altering the trans-regulatory landscape by changes in cis-regulation

As the operational definition of a terminal node would dictate, changes to their function (or expression) would result in a minimally modified GRN that differs only at the state of the gene that was changed. When higher tiers in the GRN are modified, ripples of change are predicted to emanate throughout the network. While this is likely to be quite pleiotropic, many examples of this phenomenon exist. Perhaps the simplest type of change is the modification to the expression pattern or timing of a transcription factor or signaling pathway ligand. Examples include the loss of armor plates and spines in sticklebacks (31, 35, 135), loss of trichomes in *Drosophila* larvae (51, 52, 98, 143), and the diversification of the wing spots mentioned in the above section (7).

One of the most striking cases illustrating changes in the spatial form of the trans landscape comes from work in butterfly mimicry. An enormous degree of variation exists in coloration patterns of closely related species of butterfly, exemplified by the elaborate differences between *Heliconious* butterflies (Figure 5). Early genetic studies in several *Heliconius* species suggested that a relatively small number of loci played a substantial role in the control of wing pattern variation (72, 148). In different regions, distinct species will converge on similar pigmentation patterns to share the burden of educating predators. Through a remarkable set of mapping analyses, genomics, and developmental work, we now know that much of this variation in phenotype is caused by differences in expression that have evolved by altering the extensive *cis*-regulatory apparatus of the same *trans*-regulators present in the butterfly wing. Thus far, of three regions that control the bulk of color-pattern variation, the *D* and *Sd* loci have been the most extensively studied.

Mapping of the *D* locus revealed a common genomic position for the gene causing drastic differences in red color patterns in subspecies of *H. melpomene* and *H. erato* (14, 111). An elegant microarray analysis in which pigmented portions of red-morph wings were compared to those of unpigmented subspecies revealed that within the *D* locus genomic region, the *optix* gene was uniquely upregulated in red-colored regions (129). Refined mapping data using naturally occurring hybrids also implicated *optix* as the causal locus for controlling red pigment patterns (129). An expansive comparison of *optix* expression among different *H. erato* morphs and other diverse red morphs uncovered a remarkable correlation between its mRNA pattern and phenotype. Notwithstanding ~25 million years of divergence, sequences of the *optix* coding region only differed at synonymous sites, eliminating the possibility that amino acid coding sequence variants might contribute to this phenotype. Taken together with association studies in hybrid zones that genetically pinpointed variation with *optix*, the picture that emerges is that *cis* regulatory changes at *optix* account for a large diversity of patterns of *optix* expression (129), downstream of which dozens of genes involved in synthesizing red pigment patterns have been altered in expression (65). A survey of Optix expression across the nymphalid clade that contains *Heliconius* revealed a complex and interesting history of its association with diverse types of modified wing scales, differing in color and morphology (94).

The *Sd* locus controls size, position, and shape of forewing color bands by changing the black regions of the wings, thus acting like a "shutter" to control these colored bands. Mapping of *Sd* locus implicated the *WntA* gene, which encodes a secreted signaling molecule (95). Elegant experiments using chemical treatment to modify Wnt

signaling further supported a role for *WntA* in the *Sd phenotype.* As with *optix*, the relevant variation appears to lie within the cis-regulatory DNA of *WntA*, and correlations are seen between the expression of *WntA* and the black domains in various *Heliconius* butterflies (95). Remarkably, this role of *WntA* is not restricted to *Heliconius* butterflies as it also appears to be responsible for similar variation in black patterns in distantly related butterflies (54). The origins of the diverse genetic programs that have evolved to be driven by *optix* and *WntA* remains an open and quite interesting question.

The Hox loci provide another important example of evolutionary changes to the expression of highly conserved genes. As mentioned previously, Hox genes control anterior-posterior patterning throughout the bilateria, and the overall structure and function of the Hox gene complex has been well conserved throughout more that 500 million years of evolution. Nevertheless, changes in Hox gene expression do underlie a number of examples of morphological evolution. Some of these involve later functions of Hox genes in fine tuning specific morphological features associated with particular body regions. For example, at the microevolutionary scale, changes in the levels of Ubx expression have been found to underlie differences in the pattern of hairs on the T2 legs of closely related *Drosophila* species (141). On the other hand, large shifts in the initial expression domains of Hox genes have been found to correlate with macroevolutionary changes in body plans. For example, in crustaceans, the anterior border of Ubx expression can reside at the first, second, third, or fourth thoracic segment, depending on the species (Figure 6). This expression correlates with the morphological transition from feeding to locomotory appendages in these animals (8). Further evidence for the functional significance of these shifts in expression come from gain and loss of function

studies in the crustacean, *Parhyale*, which show that the distinction between feeding and locomotory appendages is indeed regulated by *Ubx* expression (85, 112) (Figure 6). Knockdown of *Parhyale Ultrabithorax* recapitulates evolutionary changes in crustacean appendage morphology (85). While it is not yet known for certain that these changes are due to mutations in the cis-regulatory elements of crustacean *Ubx*, it points out how even the most highly conserved "toolkit" genes can evolve altered functions.

Diversification of transcription factor function at the protein-coding level

As discussed above, there is an abundance of direct evidence demonstrating that a significant source of evolutionary variation in GRNs resides within enhancers. However, a growing number of examples indicate how transcription factor coding regions can also acquire mutations that will modify their downstream GRNs. Theoretical support for the predominance of regulatory DNA mutations derives from the inherent modularity exhibited by enhancers, which is thought to limit pleiotropy (40, 157). Transcription factor proteins, however, can also be modular, and therefore, it has been argued, can also evolve in ways that reduce pleiotropic effects (90). The modulatory of protein function arises from the interaction domains that can independently execute their functions. For instance, transcription factors may have multiple DNA binding domains, as well as multiple protein-protein interaction domains. The DNA binding domains function to direct the protein to specific binding sites in enhancers to affect transcription of the target gene. The protein interaction domains bind other proteins to modify the mode of gene regulation. For instance, these domains can interact with obligate dominant repressors such as Groucho, cofactors that enhance binding and transcriptional control, or

proteins that modify the factor post-translationally to regulate its location, stability or activity.

There are several ways that proteins can evolve changes in the use of these modules to circumvent pleiotropy so that only a subset of their function is affected. The first source of variation comes from gene rearrangement processes that can shuffle the domains so that orthologs in different taxa have variations in the types and numbers of domains. Newly acquired domains can then direct new GRN functions while leaving the original functions intact. For example, a comprehensive analysis of chordate genomes revealed extensive domain shuffling (73). In particular, the authors found that some transcription factors acquired new transactivation domains, and that these appeared to be linked to the evolution of vertebrate specific characters (73).

Newly acquired domains may also execute their functions in only some spatiotemporal contexts if they facilitate binding of an interacting partner that is restricted in expression. Hence, not only can proteins evolve new domains that act independently of other domains, but these new functions may apply to only a subset of the protein's function in time and space. One particularly interesting example derives from studies of HoxA11 during the development and evolution of the placenta in mammals (20, 90). HoxA11 represses the expression of pregnancy related genes. During the evolution of placental mammals, HoxA11 acquired the ability to bind the FoxO1a transcription factor. In endometrial stromal cells, which express FoxO1a, HoxA11 interacts to activate, rather than repress pregnancy related genes.

Another mechanism through which proteins domains may act in only some fraction of their spatiotemporal environments is through context dependent splicing: i.e.

alternative splicing events that occur in particular times and places in the embryo. A recent study comparing genome wide splicing events in three primates showed that these events differed between lineages and sexes, thereby demonstrating that splicing regulation is context specific and changing rapidly (18). The functional consequences of context-specific splicing is highlighted by the *doublesex* gene in *Drosophila,* which is differently spliced in males and females, and these splice forms differently regulate downstream genes (56, 152) to direct sex specific GRNs.

More recently, another, entirely unexpected, source of modularity has been uncovered. Recent technological advances have permitted a sensitive and high throughput assessment of the binding preferences of transcription factors (16). In these experiments, proteins are allowed to bind to DNA microarrays of all possible 8, 10 or 12 mer-sequences. The motifs can be collated into a position weight matrix (PWM), which represents the preference for each base at each position within the motif. This analysis has revealed that many transcription factors have not only a primary preference PWM but also exhibit preferences for an additional, secondary PWM that cannot be explained by the primary PWM due to nucleotide interdependencies (11). Interestingly, when paralogous transcription factors are analyzed, their primary PWMs are quite similar, but they often differ in their preference for a secondary PWM (11). A critical finding from an evolutionary perspective is that orthologous transcription factors can also evolve altered preferences for secondary PWMs in different species. Cheatle-Jarvela et al (32), showed that orthologous Tbox transcription factors from mouse, sea urchin and sea star shared a highly similar primary PWM but had derived different preferences for a secondary, low affinity motif. This demonstrates that a single DNA binding domain can

evolve preferences for alternative DNA binding motifs without altering their preferences for their ancestral primary motif. The next challenge is to understand the types of GRNs that are controlled by enhancers using low affinity, secondary motifs. This will reveal the types of developmental processes that are under the regulation of these more pliable aspects of the protein's function.

Inferring evolutionary events within the human and primate GRNs

The question of how developmental networks evolve, naturally inspires the topic of how these studies apply to human evolution: what were the genetic changes accounting for our unparalleled mental capabilities, manual dexterity, and bipedalism? Indeed, early comparisons of human protein sequences to those of primates stimulated the idea that changes in gene regulation might underlie much of human uniqueness (74). This sentiment was only further bolstered by the realization that the human genome contained several times fewer genes than previously thought (12). Although it is technically challenging to experimentally validate changes within human developmental GRNs, several studies have mounted convincing cases for a number of changes on the human lineage. Because of experimental limitations, these studies have required the clever use of computational techniques and experimentation in non-primate models.

One line of inquiry depends upon the identification of highly conserved sequences among vertebrates that have been drastically altered in humans (117, 120). Because these "accelerated regions" show extreme alterations of highly conserved non-coding elements, it stands to reason that they would represent functional changes in the human lineage, possibly affecting gene regulation. The first documented example, *HAR1* (human

accelerated region 1), revealed a significantly altered non-coding RNA gene, whose expression in human and macaque brain were quite similar (118). Although this case represented a change in RNA secondary structure rather than gene expression, *HAR1* was found to be expressed in a region of the human neocortex that has been greatly elaborated along the human lineage. A second example, *HACNS1* (human accelerated conserved non-coding sequence 1) mapped to a non-coding region that drives expression in the vertebrate limb (121). Comparisons of regulatory sequence function in mouse reporter assays revealed that the human version of this regulatory sequence was much more active than any other primate reporter, extending into the thumb. Hence, one of the genes adjacent to this sequence may have evolved an expanded domain of expression in the thumb, possibly contributing to human adaptations of the hand.

While it is difficult to verify the phenotypic consequences of accelerated evolution of *HAR1* and *HACNS1*, studies examining the loss of conserved non-coding regions may provide more straightforward examples of changes underlying human uniqueness. McLean and colleagues systematically identified human-specific deletions that occurred within conserved non-coding elements, finding 510 such "hCONDELs" (99). From this set, the authors provided compelling evidence for two examples, showing that the chimp and mouse sequences deleted in the human lineage were active regulatory elements. The first hCONDEL occurred within the androgen receptor, deleting an enhancer active in penile spines. The androgen receptor confers responsiveness to circulating androgens, and this hormone signaling was previously known to be required for penile spine formation. The deletion of this spine enhancer correlates very well with

the absence of these androgen dependent structures in humans compared to many other mammals.

The second hCONDEL identified a change that may have contributed to increased brain size (99). The gene *GADD45G* is a tumor suppressor gene known to inhibit cell cycle progression and induce apoptosis. The loss of this enhancer, active in the subventricular zone may have led to decreased *GADD45G* expression, relaxing negative regulation on the growth of the human neocortex. Although such findings seem more clear-cut than the examples listed above for accelerated regions, there are many open questions concerning these case studies. For example, several conserved non-coding sequences have been deleted in the mouse with no apparent phenotypic consequence (3). Thus, without data concerning the expression status of these genes during human development, these changes may have been silent due to redundancy with other elements in these genes.

Several studies of human-specific changes have focused instead on tissues for which gene expression, activity, and even chromatin state can be measured, such as skin fibroblasts (136) or circulating blood (147). One prime example compared expression of glucose transporters in the human and chimp brain and muscle tissue, revealing a potential reduction in muscle expression of *SLC2A4* that was accompanied by an increase in brain activity of *SLC2A1* on the human lineage (49). These expression changes may have tipped the energy balance between muscle and brain tissues to support the more energetically demanding needs of the human brain. For both genes, an excess of substitutions was found within their 5' UTRs, a signature suggesting positive selection in a potential regulatory region.

To disentangle correlation from causation for any study of GRN evolution, the ultimate form of evidence should establish the sufficiency of implicated genetic changes to generate the evolved phenotype in question. This type of evidence is generally only available for a small number of model organisms that have tools for transgenesis or homologous recombination. This type of experimental support is especially problematic for studies of human GRN evolution, in that the closest model organism in which such studies are possible may be so different that results of placing a humanized allele will be meaningless. In a vanishingly small number of cases, such experiments have been attempted. Of these, perhaps the greatest amount of information is available for the *FOXP2* gene.

FOXP2 is a forkhead domain transcription factor for which mutations were identified in individuals with speech disorders that stem from a lack of coordinated face and mouth movements (80). Stimulated by the finding that variation in *FOXP2* could affect one's ability to speak, a comparative sequencing analysis between humans and several primates revealed an excess of changes to its amino acid sequence on the human lineage, a sign of positive selection (48). Intriguingly, these changes predated the divergence of Neanderthals from modern humans (79). To the extent that differences can be distinguished, *foxp2* expression in mouse and human brains are quite similar (81).

To test the potential functional implications of the amino-acid changes in *FOXP2*, a humanized mouse was engineered, in which the two human specific amino acids were introduced into the endogenous mouse gene. These changes resulted in increases in dendrite length, depressions in dopamine levels, and an alteration in ultrasonic vocalization (47) in humanized mice compared to controls. Recently, behavioral assays

have shown that the humanized *FOXP2* mice have a slight increase in learning how to use spatial cues to obtain rewards, and are more dynamic in transitioning between reward systems (133) compared to wildtype individuals. The GRN-level effects of the amino acid changes at *FOXP2* have been more recalcitrant to functional characterization. Transcriptome profiling approaches revealed very few major changes in gene expression. This could be due to these amino acid changes affecting FOXP2 function in just a few cell types (133). While the amino acid changes at *FOXP2* may be our most detailed example of GRN level change in the human lineage, the applicability of the mouse model to human evolution must be interpreted with caution. Some evidence suggests that *FOXP2* is subject to long distance regulation with elements perhaps 3 Mb away from the promoter (2). Further, cis-regulatory mutations along the human lineage have been implicated to alter *FOXP2* expression as well (93). Currently this is the best example we have, but it very likely represents one of a vast number of changes involved in human brain evolution. Thousands of the genes exhibit differential brain expression among primates (10), no small number of which may contribute to our exquisite mental capabilities. As more information and increasingly inventive ways to probe the phenotypic consequences of mutations specific to the human lineage arise, the role of *FOXP2* relative to these other changes will become apparent.

IV. Network origination

Perhaps the biggest challenge in understanding how networks evolve is the problem of understanding novelty: how do completely new structures that bear no obvious precursor originate? In particular, this is a complex problem because of the

difficulty in elucidating evolutionary events in the distant past, where the most obvious of novelties such as appendages, eyes, or the turtle's shell arose.

Co-option of the appendage-patterning network during beetle horn evolution

One system that has gained much traction is the origination of beetle horns. The beetle horn has long been appreciated as a novelty of particular adaptive significance. As one of the most successful collections of species, thousands of sorts of beetles bear horns that project from various parts of the body (46). Used by males as weaponry in the struggle to find mates, this sexually selected structure can increase the chances of a male's success in obtaining and preserving mates (100). At the molecular level, this structure is a perfect exemplar of GRN co-option, the re-deployment of a pre-existing network to evolve a new structure.

Across the arthropods, the genes that participate in appendage formation are highly conserved (76, 110). The ectopic deployment of top-level regulators of this network is sufficient to induce additional appendage outgrowths (26), suggesting that when deployed at the right place and time, this program could be co-opted to generate novel structures. Analysis of the appendage battery in developing horns has demonstrated the expression of several genes in this network during development (101, 103). Indeed, the *distalless* gene, associated broadly with appendages in many species of protostomes and deuterostomes (110), is expressed both in pronotal and head horns (101). The deployment of these genes in the novel context of the horn suggests that a conserved downstream program for proximo-distal axis formation is activated in the developing

horn, causing proximo-distal axis responsive enhancers to be re-deployed in the novel setting.

An emerging theme of this research program is the realization that not all genes of the network appear to contribute to the morphology of the horn. While proximo-distal - axis genes *distalless* and *homothorax* exhibit altered horn morphology upon RNAi knockdown, the *dachshund* gene failed show a horn phenotype, even though RNAi conditions generated phenotypes in other tissues (102). This finding suggests that perhaps not all co-opted circuits make developmental contributions to the morphological novelty.

Co-option of the Echinoderm larval skeleton

While many examples of GRN co-option have been uncovered among groups of arthropods, this phenomenon has also been implicated in other developmental systems, supporting the concept that network co-option may be a universal evolutionary phenomenon. One such example is of the echinoderm larval skeleton. Sea urchin larvae develop long biomineralized skeletal rods that provide the characteristic pyramidal shape to their plutei larvae. Representatives of outgroup species of echinoderm (e.g. crinoids and sea stars) do not make larval skeletons. All echinoderms however make a skeleton during their adult stages, as it is a defining feature of the phylum. Gao and colleagues (55) investigated whether the GRN for the larval skeleton could have been co-opted from an ancestral GRN that directed the formation of adult skeletogenesis. They showed that many transcription factors, signaling molecules, as well as terminally acting genes involved in the formation of the biomineral were indeed co-expressed during both the development of the larval skeleton and during the formation of the adult skeleton.

Because the GRN for the development of the larval skeleton has been characterized in extraordinary detail (108), the authors were able to show that genes with overlapping expression patterns mapped to regions of the GRN acting downstream of initial specification processes. Importantly, some orthologs of these genes were also co-expressed in the developing skeleton of the adult sea star, suggesting that this adult skeletogenic GRN was basal. Recently, it has also been shown that the sea cucumber, which is a sister taxa to the sea urchin, also expresses some, but not all of these genes during the formation of its much reduced larval skeleton (97). This suggests that upon co-option there has been further remodeling of the GRN to modify skeletal morphology.

Where do nodes in the network come from?

Two major themes unify the examples of novelties described above: (i) rather than evolving new genes to generate a new morphological structure, novelties are often associated with new and unique expression patterns of ancestral genes. (ii) The ancestral genes associated with novelties are often transcription factors and signaling pathway ligands that are predicted to alter many subordinate genes in their respective networks. In some of these cases, the network that has been co-opted can be reasonably inferred (e.g. beetle horn co-option of the proximo-distal axis specification network), but in many cases the downstream network's identity remains unknown. Hence, these two properties also stimulate two emerging questions in understanding how GRNs underlying these novel morphological structures evolve: How do genes evolve novel expression patterns that place them into new networks? How does network co-option proceed at the level of

individual genes and the genome? Recent advances provide a rich picture of the molecular mechanisms by which new expression patterns evolve.

The question of how expression patterns arise can be rephrased into terms of how enhancers originate or evolve new activities at the molecular level. Several mechanisms have been hypothesized to underlie the evolution of new enhancer activities. The simplest way is through *trans* regulatory modification, in which a pre-existing enhancer is activated in a new location because its upstream regulators have been deployed in this new tissue. The beetle horn and sea urchin larval skeleton co-option studies imply that dozens if not hundreds of pre-existing enhancers are activated in the novel site of GRN deployment.

In addition to *trans*-regulatory co-options, multiple mechanisms by which novel enhancer activities have arisen through changes in *cis* have been documented (Figure 7). These mechanisms typically differ by the type of ancestral information that was present preceding the evolution of the novelty. For example, many studies have implicated the role of transposable elements (TEs) in the origination of enhancer sequences (50). The human genome is composed of as much as two-thirds repetitive sequences derived from transposons (41). Indeed, many characterized enhancers overlap TEs, and display striking patterns of conservation within the TE sequence, suggesting purifying selection has acted on these regions (15, 87). During the evolution of pregnancy, hundreds of genes expressed in the endometrial stromal cells have become associated with nearby transposon insertions that alter stromal cell gene activity (89). However, of the many cases of TE **exaption**, very few are subject to functional analysis in vivo (42). In one detailed example, the neuronal enhancers of the vertebrate *Pomc* gene are TE derived, but

clearly replaced an ancestral enhancer, which was likely lost after the TE insertion gained enhancer activity (44). Furthermore, current examples of TE exaption that have been linked to phenotypes are associated with increased abundance, rather than the evolution of expression that was previously absent. In the case of *teosinte-branched1*, a combination of two transposons increased expression in domesticated maize (142). Similarly, pesticide resistance in *Drosophila* is associated with increased expression of *Cyp6G* in the gut, caused by a transposon insertion that is sufficient to drive gut expression (34). Thus, while it is entirely possible that many genes have gained new expression patterns via TE exaption, examples that show clearly novel expression domains are currently needed.

Another potential mechanism driving the evolution of new expression patterns is to simply alter enhancer-promoter specificity (Figure 7). Enhancers are subject to tight control of their associations with promoters, often exclusively interacting with only a single promoter. Insulator elements restrict enhancers from interacting with the wrong promoter (24), while promoter tethering elements facilitate interactions with specific target promoters (4, 25). A famous example that illustrates the exquisite control of promoter specificity is the limb enhancer of Sonic hedgehog, which is located ~1 megabase upstream of its promoter, in the intron of a gene two genes away (82). Evolution of enhancer-promoter control sequences may cause an adjacent gene to adopt the activity of the nearby enhancer. In practice, chromosomal rearrangement has been documented to drive promoter-switching events. In the *tinman* gene complex of the beetle *T. castaneum*, a chromosomal inversion caused a change in an enhancer's position relative to an insulator sequence, which allowed the *C15* gene to adopt an enhancer from

the *ladybird* gene (27). The *Rose-comb* locus of the domesticated chicken offers a phenotypically relevant example of chromosome rearrangement leading to novel expression (68). In this case, the *MNR2* gene, encoding a Homoedomain transcription factor, is located 3 kilobases away from the edge of a large-scale (7 megabase) chromosomal inversion in animals bearing the *Rose-comb* phenotype. Placement of this transcription factor in this new regulatory environment is associated with the novel embryonic expression of *MNR2* in the developing comb mesenchyme, suggesting that the inversion placed *MNR2* in close proximity to a comb mesenchyme enhancer (68). One potential pitfall of the promoter-switching mechanism is pleiotropy – by interacting with a new promoter, the ancestral pattern of the original target gene of the enhancer may be altered. However, the wide prevalence of **shadow enhancers**, in which a single expression pattern is encoded by multiple separable enhancer elements (67), may alleviate the potentially pleiotropic effects of such promoter switching events. In this case, the loss of one enhancer due to alteration of enhancer-promoter specificity or chromosomal rearrangement may have little effect on the overall expression of the ancestral gene.

A potentially powerful mechanism to generate novel expression domains is the reuse, or co-option of individual ancestral enhancers (Figure 7, "co-option"). In this case, the pre-existing enhancer already contains a dozen or more binding sites for upstream transcription factors. As most transcription factors are expressed in multiple tissues during development, any instance of a transcription factor binding site could lead to ectopic activation in a new setting that contains the upstream factor. Hence, most enhancers may be merely a few mutations away from generating a novel activity in a new

tissue. A study of a novel expression pattern of the *Nep1* gene in the *Drosophila santomea* optical lobe revealed a co-option event that arose over the 400,000 years separating *D. santomea* from its sister species *D. yakuba* (126). In this case, short stretches of sequence within this enhancer were shown to affect function in both the novel and ancestral expression activities. This demonstrates how individual transcription factor binding sites can be re-used to generate new patterns of expression. In a genome-wide survey of an enhancer-associated chromatin marks in human, macaque, and mouse limbs, it was found that a large fraction (18%) of enhancers gained in the human lineage showed evidence of ancestral enhancer-associated marks in other tissues (36). Hence, co-option may be a driving mechanism for the generation of enhancers in the human genome.

Finally, enhancers may simply evolve from non-functional sequences that contain chance binding sites. A study of pseudogenized exons in zebrafish established that sequences previously used as exons exhibit enhancer activity, suggesting that upon pseudogenization, a regulatory role evolved to generate novel patterns of gene expression (45). While this mechanism may require more steps than we generally think is likely, it may nevertheless be more prevalent than we anticipate. For example, several studies have indicated that some enhancers rapidly turn over binding sites, evolving new sites to replace old ones to maintain a conserved output (88, 144). This indicates that the evolution of a complex enhancer containing many new binding sites is perhaps not so implausible. However, more examples are needed, and the identification of clear instances of *de novo* generation is difficult, as it requires evidence that the ancestral sequence contained no activity in any tissue, anywhere in the developing organism or adult.

Influences of enhancer origination mechanisms on network shape

It may be that the mechanisms at work during the evolution of enhancer activities dictate the shape and hierarchical relationships within a GRN. For example, if a network evolved through the widespread exaption of a transposable element, as proposed for mammalian pregnancy (89), the resulting network may be flat, and wide, with very little stratification. On the other hand, models of wholesale network co-option posit that the hundreds of genes deployed in a new cell type result from a small number of changes in the upper tiers of the GRN that cause many subordinate genes to be expressed. A variety of mechanisms may lie at the heart of these changes that initiate a network co-option: promoter switching, *de novo* enhancer evolution, transposition, or co-option of pre-existing binding sites within an enhancer. Given the pre-existing structure of a co-opted network, it may be more stratified than a network rewired by TE exaption.

However, the case of GRN co-option via the expansion of a pre-existing enhancer's activity to new location poses an interesting conundrum. If a pre-existing enhancer evolves a new expression pattern in a new territory due to its pre-existing inputs, this would, in a sense, lead to the novel setting becoming more like that enhancer's ancestral setting. If the ancestral setting matches that of the GRN that is being re-deployed, this raises the question of how the novelty can differ from the structure encoded by the ancestral network, and how such co-opted networks can become specialized while mitigating pleiotropic effects on other tissues. Perhaps one underappreciated solution to this problem is that the derived networks always exist in the context of a new cell type or tissue that implements its own pre-existing network.

Currently, we know little of how new networks interact with the pre-existing networks already in place.

CONCLUDING REMARKS

The last decade or so has seen great strides in our understanding of GRN structure and function. This has paved the way for a clearer definition of how GRNs evolve to generate morphological phenotypes. Below, we briefly discuss some of the challenges for the future in achieving a deeper understanding of GRN evolution, origination, and conservation.

The deep conservation of regulatory factors in developmental GRNs was one of the observations that motivated the hypothesis that regulatory sequence evolution may drive morphological diversification. However, such findings also led to the discovery that many subcircuits within GRNs may have a deep ancestry that has been preserved since their origination in early metazoans. The systematic identification of these ancient linkages will reveal the core circuits of GRNs that are recalcitrant to change, and broadly apply to a large number of species (including humans). A key challenge in this endeavor is the relatively small number of GRNs that have been mapped in any species, which compounds the effort required to map such GRNs in multiple species. A second major task is to explore how deeply conserved regulatory sequences, such as those found in SoxB2 are employed in different organisms. Why are they preserved for so long? And what principles dictate their re-deployment to other developmental settings?

Alongside deep conservation, some of the most interesting aspects of GRN biology are the parts that are changing. We now have several initial examples of GRN

evolution, and there are many conceptual advances to be made in this arena. The ubiquity of cis-regulatory mutations that alter morphological phenotypes is becoming increasingly accepted, and a major challenge now is to obtain a clearer picture of how these mutations affect transcription factor binding, and the logical operations that determine transcriptional output. Only a small number of enhancer sequences are well studied in any one species, and a mutation to such an element may create or destroy the binding of a small handful of the hundreds of transcription factors present in the metazoan genome. It is fair to say that any case study of enhancer evolution is likely to be missing some important aspects of what factors are bound in vivo. This is an area that will improve with the state-of-the-art of regulatory biology in model organisms.

For alterations of transcription factor coding sequences, we are now at the point where likely changes have been implied, but very few detailed examples of their direct phenotypic consequences exist. Progress in this respect has been hampered to some degree by the focus of these studies on macroevolutionary changes. It may well be easier to detect a difference in coding sequence function over a longer evolutionary distance, but the trade-off is that it then becomes difficult to form a "before and after" picture of what phenotypic consequences these changes engendered. We believe that this is an area of GRN biology that is ripe for intense experimental treatment.

Perhaps one of the most exciting aspects of GRN evolution is an area that will require much experimental attention: the origination of GRNs and novel morphological phenotypes. While all of the current examples rest upon the tried and true inferences based on comparative analyses of gene expression, examples that probe the circuits themselves are needed. We need examples in which we can point to the causative

changes, and downstream enhancers that were activated during the incipient stages of network co-option. Systems in which the network co-option event can be recapitulated through gene mis-expression studies would allow us to probe which connections were required to generate the novel phenotype, versus connections lacked phenotypic consequence (e.g. *dachshund* in the beetle horn). How sloppy is the origination of a novel structure? And after its establishment, how were connections refined, altered, or pruned? Answers to these questions will illuminate the origins of morphological complexity, one of our biggest challenges in the understanding of evolution.

Summary Points (8)

- Gene Regulatory Networks (GRNs) lie at the heart of animal development, controlling which subset of the genome is activated in time and space to generate a tissue's physical properties.
- Most transcription factor families and developmental signaling pathways are conserved across the animal kingdom.
- The GRNs for many organ systems have a deep metazoan ancestry, with some individual transcription factor binding sites remaining conserved for hundreds of millions of years.
- The alteration of transcriptional regulatory sequences (cis-regulatory evolution) is a commonly observed mechanism for morphological evolution that modifies the pattern and level of a gene's expression.
- A multitude of mechanisms exist to modify transcription factor function at the coding-sequence level.
- Many human adaptations are thought to have resulted from cis-regulatory evolution, and mounting evidence for this mechanism has accumulated in several case studies.
- The co-option, or re-use of pre-existing GRNs is thought to underlie their origination, leading to the generation of novel structures.
- Several mechanisms for the deployment of genes into new tissues exist, which may dictate a GRN's resulting shape and topology.

Future Directions (8)

- Systematic identification of ancient subcircuits that form the kernels of GRNs.
- Deeper investigation into the function and re-usage of ancient circuit connections.
- Elucidation of how mutations are translated into differences in transcription factor binding, and transcriptional output is a crucial, and yet understudied facet of connecting non-coding sequence variation to phenotypic consequences.
- The phenotypic consequences of coding-sequence evolution within GRNs is a pressing issue that is difficult to disentangle, given the long evolutionary distances that many of these changes have been inferred to occur over.
- Increasingly clever usage of model systems is required to advance our understanding of human evolution.
- Investigations of morphological novelties in systems amenable to the evaluation of regulatory sequence function is crucial to obtain a clearer picture of GRN origination.

ACKNOWLEDGMENTS

We thank members of the Rebeiz, Patel, and Hinman Labs, as well as members of the Pittsburgh area Molecular Evolution Laboratory Discussion group for helpful discussions. M.R. was supported by grants IOS 1145947 (NSF) and GM107387 (NIH), as well as a fellowship from the Alfred P. Sloan Foundation. NHP was supported by grant IOS 1257379 (NSF). VH was supported by grants IOS 0844948 and IOS 1024811. (NSF) and a Charles E. Kaufman award of the Pittsburgh Foundation

DEFINITIONS

internal node: A gene that functions within the GRN. These genes typically control the expression of other genes, and are therefore, most often, transcription factors, and members of signaling pathways.

terminal node: A gene that is activated at the terminus of a GRN. Typically these genes have no role in regulating gene expression and instead influence the final differentiation of the cell types.

pleiotropy: Having multiple, independent functions. Regulatory genes frequently affect the expression of multiple other genes, at multiple times and places in the developing embryos, and are therefore especially pleiotropic.

subcircuit: A set of several genes, or nodes, within a GRN, which have regulatory interconnections. Such sets of genes may execute some particular, discrete function.

network topology: The shape, or structure of the GRN defined by the regulatory connections, or edges, between nodes. No kinetic information is provided by the topology, which is more simply a static representation of the GRN.

enhancer: Also termed a cis regulatory module, or a cis regulatory element. A regulatory DNA sequence, usually 500 to 1000 bases in length, to which a set of transcription factors bind to regulate the expression of a target gene. Enhancers are often within several kilobases, up- or downstream-, or within introns of the coding sequence of the regulated gene.

combinatorial logic: The control of gene expression by the binding of multiple different transcription factors that each provide an input, which in combination determines the precise dynamics of the final gene expression.

modularity: Property of regulatory sequences that allows them to be altered independently, resulting in few pleiotropic consequences for other activities of a given gene. This is due to the requirement for binding sites to reside close to one another within a typical enhancer, rendering mutations that alter binding sites in one enhancer to have little effect on other enhancers at a locus.

trans-regulatory landscape. The content of regulatory genes present in a cell, cell type or territory of the embryo. The genes of this landscape are the subset of genes that are available to bind enhancers and hence regulate the next set of genes to be expressed.

EMSA: Electrophoretic Mobility Shift Assay, also known as a gel-shift assay. The mobility of a labeled DNA probe is measured by gel electrophoresis in the presence and absence of a potential binding transcription factor. If the probe's mobility depends on the factor's presence, the factor is inferred to bind the sequence in question.

chromatin immunoprecipitation: in vivo assay to determine transcription factor occupancy on regulatory DNA. An antibody raised to a transcription factor of interest is used to "pull down" isolated chromatin from a tissue in which the factor is active. Segments of DNA pulled down can be measured by quantitative PCR, or high throughput sequencing ("ChIP-seq").

reporter assay. The in vivo analysis of an enhancer's function. Typically, the predicted enhancer is cloned into a region of DNA that contains a basal promoter upstream of a transgene that encodes a readily detectible protein such as GFP or luciferase. The expression of the transgene serves to reveal the pattern of gene expression normally driven by the enhancer.

positive feedback loop. A subcircuit in which the genes positively and mutually regulate the expression of other genes in the subcircuit. In its simplest form, “gene A" regulates a “gene B", and in turn “gene B" regulates the expression of “gene A".

phylogenetic footprinting: The use of sequence conservation to identify important binding sites within regulatory DNA.

microevolutionary: evolutionary change occurring within a population, or between very recently diverged species.

macroevolutionary: evolutionary change occurring at or beyond the species level.

intraspecific variation: variation that exists within a population of individuals of the same species

shadow enhancer: A second enhancer that encodes a specificity that is identical to another enhancer of the same gene. The enhancer referred to as the ‘shadow’ enhancer is generally further away from the promoter.

PWM: position weight matrix. A matrix of a transcription factor’s binding specificity based on multiple sequences that have been shown to bind the factor. PWMs can be used to score and rank potential target sequences within a query regulatory sequence of interest.

TE: transposable element. A self-replicating element that can insert itself into the genome.

exaption: The reuse or co-option of an ancestral gene, feature, or trait for a new usage. For example, a transposable element that evolves a function as an enhancer.

LITERATURE CITED

1. Abouheif E. 1999. Establishing homology criteria for regulatory gene networks: prospects and challenges. *Novartis Found. Symp.* 222:207–21; discussion 222–25

2. Adegbola AA, Cox GF, Bradshaw EM, Hafler DA, Gimelbrant A, Chess A. 2014. Monoallelic expression of the human foxp2 speech gene. *Proc. Natl. Acad. Sci. U. S. A.*

3. Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, et al. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 5(9):e234

4. Akbari OS, Bae E, Johnsen H, Villaluz A, Wong D, Drewell RA. 2008. A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the drosophila embryo. *Development*. 135(1):123–31

5. Alon U. 2007. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8(6):450–61

6. Alvarez AD, Shi W, Wilson BA, Skeath JB. 2003. Pannier and pointedp2 act sequentially to regulate drosophila heart development. *Development*. 130(13):3015–26

7. Arnoult L, Su KFY, Manoel D, Minervino C, Magrina J, et al. 2013. Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science (80-.).* 339(6126):1423–26

8. Averof M, Patel NH. 1997. Crustacean appendage evolution associated with changes in hox gene expression. *Nature*. 388(6643):682–86

9. Avery L, Wasserman S. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* 8(9):312–16

10. Babbitt CC, Fedrigo O, Pfefferle AD, Boyle AP, Horvath JE, et al. 2010. Both noncoding and protein-coding rnas contribute to gene expression evolution in the primate brain. *Genome Biol. Evol.* 2:67–79

11. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. 2009. Diversity and complexity in dna recognition by transcription factors. *Science*. 324(5935):1720–23

12. Baltimore D. 2001. Our genome unveiled. *Nature*. 409(6822):814–16

13. Barolo S, Posakony JW. 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* 16(10):1167–81

14. Baxter SW, Papa R, Chamberlain N, Humphray SJ, Joron M, et al. 2008. Convergent evolution in the genetic basis of müllerian mimicry in heliconius butterflies. *Genetics*. 180(3):1567–77

15. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, et al. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*. 441(7089):87–90

16. Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW 3rd, Bulyk ML. 2006. Compact, universal dna microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* 24(11):1429–35

17. Black BL, Olson EN. 1998. Transcriptional control of muscle development by myocyte enhancer factor-2 (mef2) proteins. *Annu. Rev. Cell Dev. Biol.* 14:167–96

18. Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. 2010. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* 20(2):180–89

19. Bodmer R. 1993. The gene tinman is required for specification of the heart and visceral muscles in drosophila. *Development*. 118(3):719–29

20. Brayer KJ, Lynch VJ, Wagner GP. 2011. Evolution of a derived protein-protein interaction between hoxa11 and foxo1a in mammals caused by changes in intramolecular regulation. *Proc. Natl. Acad. Sci. U. S. A.* 108(32):E414–20

21. Bridges CB, Morgan TH. 1919. Contributions to the genetics of drosophila melanogaster. *Carnegie Inst. Washingt. Publ.* 278:123–304

22. Bridges CB, Morgan TH. 1923. The third-chromosome group of mutant characters of drosophila melanogaster. *Publ. Carnegie Inst.* 327:50

23. Britten RJ, Davidson EH. 1969. Gene regulation for higher cells: a theory. *Science (80-.).* 165(891):349–57

24. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, et al. 2002. The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A*. 99 Suppl 4:16433–37

25. Calhoun VC, Stathopoulos A, Levine M. 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the drosophila antennapedia complex. *Proc Natl Acad Sci U S A*. 99(14):9243–47

26. Campbell G, Weaver T, Tomlinson A. 1993. Axis specification in the developing drosophila appendage: the role of wingless, decapentaplegic, and the homeobox gene aristaless. *Cell*. 74(6):1113–23

27. Cande JD, Chopra VS, Levine M. 2009. Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, tribolium castaneum. *Development*. 136(18):3153–60

28. Carbone MA, Llopart A, deAngelis M, Coyne JA, Mackay TF. 2005. Quantitative trait loci affecting the difference in pigmentation between drosophila yakuba and d. santomea. *Genetics*. 171(1):211–25

29. Carroll SB, Grenier JK, Weatherbee SD. 2005. *From DNA to diversity : molecular genetics and the evolution of animal design*. Malden, MA: Blackwell Pub. 2nd ed.

30. Carroll SB, Scott MP. 1986. Zygotically active genes that affect the spatial expression of the fushi tarazu segmentation gene during early drosophila embryogenesis. *Cell*. 45(1):113–26

31. Chan YF, Marks ME, Jones FC, Villarreal Jr. G, Shapiro MD, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitx1 enhancer. *Science (80-.).* 327(5963):302–5

32. Cheatle Jarvela AM, Brubaker L, Vedenko A, Gupta A, Armitage BA, et al. 2014. Modular evolution of dna binding preference of a tbrain transcription factor provides a mechanism for modifying gene regulatory networks. *Mol. Biol. Evol.*

33. Chen L, Zhao P, Wells L, Amemiya CT, Condie BG, Manley NR. 2010. Mouse and zebrafish hoxa3 orthologues have nonequivalent in vivo protein function. *Proc. Natl. Acad. Sci. U. S. A.* 107(23):10555–60

34. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, et al. 2007. Cis-regulatory elements in the accord retrotransposon result in tissue-specific expression of the drosophila melanogaster insecticide resistance gene cyp6g1. *Genetics*. 175(3):1071–77

35. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal Jr. G, Dickson M, et al. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science (80-.).* 307(5717):1928–33

36. Cotney J, Leng J, Yin J, Reilly SK, DeMare LE, et al. 2013. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell*. 154(1):185–96

37. Cripps RM, Olson EN. 2002. Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev. Biol.* 246(1):14–28

38. Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, Vol. 138. John Murray

39. Davidson EH. 2001. *Genomic regulatory systems : development and evolution*. San Diego: Academic Press

40. Davidson EH. 2006. *The regulatory genome : gene regulatory networks in development and evolution*. Burlington, MA : Academic

41. De Koning a PJ, Gu W, Castoe T a, Batzer M a, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7(12):e1002384

42. De Souza FSJ, Franchini LF, Rubinstein M. 2013. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol. Biol. Evol.* 30(6):1239–51

43. de-Leon SB-T, Davidson EH. 2010. Information processing at the foxa node of the sea urchin endomesoderm specification network. *Proc. Natl. Acad. Sci. U. S. A.* 107(22):10103–8

44. Domené S, Bumaschny VF, de Souza FSJ, Franchini LF, Nasif S, et al. 2013. Enhancer turnover and conserved regulatory function in vertebrate evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368(1632):20130027

45. Eichenlaub MP, Ettwiller L. 2011. De novo genesis of enhancers in vertebrates. *PLoS Biol.* 9(11):e1001188

46. Emlen DJ, Szafran Q, Corley LS, Dworkin I. 2006. Insulin signaling and limb-patterning: candidate pathways for the origin and evolutionary diversification of beetle "horns". *Heredity (Edinb).* 97(3):179–91

47. Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, et al. 2009. A humanized version of foxp2 affects cortico-basal ganglia circuits in mice. *Cell*. 137(5):961–71

48. Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, et al. 2002. Molecular evolution of foxp2, a gene involved in speech and language. *Nature*. 418(6900):869–72

49. Fedrigo O, Pfefferle AD, Babbitt CC, Haygood R, Wall CE, Wray GA. 2011. A potential role for glucose transporters in the evolution of human brain size. *Brain. Behav. Evol.* 78(4):315–26

50. Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 9(5):397–405

51. Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory dna. *Nature*. 474(7353):598–603

52. Frankel N, Wang S, Stern DL. 2012. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc. Natl. Acad. Sci.*

53. Galant R, Carroll SB. 2002. Evolution of a transcriptional repression domain in an insect hox protein. *Nature*. 415(6874):910–13

54. Gallant JR, Imhoff VE, Martin A, Savage WK, Chamberlain NL, et al. 2014. Ancient homology underlies adaptive mimetic diversity across butterflies. *Nat. Commun.* 5:4817

55. Gao F, Davidson EH. 2008. Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105(16):6091–96

56. Garrett-Engele CM, Siegal ML, Manoli DS, Williams BC, Li H, Baker BS. 2002. Intersex, a gene required for female sexual development in drosophila, is expressed in both sexes and functions together with doublesex to regulate terminal differentiation. *Development*. 129(20):4661–75

57. Geyer PK, Corces VG. 1987. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in drosophila melanogaster. *Genes Dev.* 1(9):996–1004

58. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in drosophila. *Nature*. 433(7025):481–87

59. Gottesman S. 1984. Bacterial regulation: global regulatory networks. *Annu. Rev. Genet.* 18:415–41

60. Greer JM, Puetz J, Thomas KR, Capecchi MR. 2000. Maintenance of functional equivalence during paralogous hox gene evolution. *Nature*. 403(6770):661–65

61. Grens A, Mason E, Marsh JL, Bode HR. 1995. Evolutionary conservation of a cell fate specification gene: the hydra achaete-scute homolog has proneural activity in drosophila. *Dev. Cambridge Engl.* 121(12):4027–35

62. Halder G, Callaerts P, Gehring WJ. 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in drosophila. *Science (80-.).* 267(5205):1788–92

63. Hanks MC, Loomis CA, Harris E, Tong CX, Anson-Cartwright L, et al. 1998. Drosophila engrailed can substitute for mouse engrailed1 function in mid-hindbrain, but not limb development. *Development*. 125(22):4521–30

64. Harding K, Hoey T, Warrior R, Levine M. 1989. Autoregulatory and gap gene response elements of the even-skipped promoter of drosophila. *EMBO J.* 8(4):1205–12

65. Hines HM, Papa R, Ruiz M, Papanicolaou A, Wang C, et al. 2012. Transcriptome analysis reveals novel patterning and pigmentation genes underlying heliconius butterfly wing pattern variation. *BMC Genomics*. 13(1):288

66. Hinman VF, Nguyen AT, Cameron RA, Davidson EH. 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100(23):13356–61

67. Hong J-W, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science (80-.).* 321(5894):1314

68. Imsland F, Feng C, Boije H, Bed'hom B, Fillon V, et al. 2012. The rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet.* 8(6):e1002775

69. Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The evolution of gene regulation underlies a morphological difference between two drosophila sister species. *Cell*. 132(5):783–93

70. Jeong S, Rokas A, Carroll SB. 2006. Regulation of body pigmentation by the abdominal-b hox protein and its gain and loss in drosophila evolution. *Cell*. 125(7):1387–99

71. Johnson JE, Birren SJ, Anderson DJ. 1990. Two rat homologues of drosophila achaete-scute specifically expressed in neuronal precursors. *Nature*. 346(6287):858–61

72. Kapan DD, Flanagan NS, Tobler A, Papa R, Reed RD, et al. 2006. Localization of müllerian mimicry genes on a dense linkage map of heliconius erato. *Genetics*. 173(2):735–57

73. Kawashima T, Kawashima S, Tanaka C, Murai M, Yoneda M, et al. 2009. Domain shuffling and the evolution of vertebrates. *Genome Res.* 19(8):1393–1403

74. King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science (80-.).* 188(4184):107–16

75. Klinedinst SL, Bodmer R. 2003. Gata factor pannier is required to establish competence for heart progenitor formation. *Development*. 130(13):3027–38

76. Kojima T. 2004. The mechanism of drosophila leg development along the proximodistal axis. *Dev. Growth Differ.* 46(2):115–29

77. Komuro I, Izumo S. 1993. Csx: a murine homeobox-containing gene specifically expressed in the developing heart. *Proc. Natl. Acad. Sci. U. S. A.* 90(17):8145–49

78. Kopp A, Graze RM, Xu S, Carroll SB, Nuzhdin S V. 2003. Quantitative trait loci responsible for variation in sexually dimorphic traits in drosophila melanogaster. *Genetics*. 163(2):771–87

79. Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, et al. 2007. The derived foxp2 variant of modern humans was shared with neandertals. *Curr. Biol.* 17(21):1908–12

80. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 413(6855):519–23

81. Lai CSL, Gerrelli D, Monaco AP, Fisher SE, Copp AJ. 2003. Foxp2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain*. 126(Pt 11):2455–62

82. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, et al. 2002. Disruption of a long-range cis-acting regulator for shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A*. 99(11):7548–53

83. Levine M. 2010. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20(17):R754–63

84. Lints TJ, Parsons LM, Hartley L, Lyons I, Harvey RP. 1993. Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development*. 119(2):419–31

85. Liubicich DM, Serano JM, Pavlopoulos A, Kontarakis Z, Protas ME, et al. 2009. Knockdown of parhyale ultrabithorax recapitulates evolutionary changes in crustacean appendage morphology. *Proc. Natl. Acad. Sci. U. S. A.* 106(33):13892–96

86. Llopart A, Elwyn S, Lachaise D, Coyne JA. 2002. Genetics of a difference in pigmentation between drosophila yakuba and drosophila santomea. *Evolution (N. Y).* 56(11):2262–77

87. Lowe CB, Bejerano G, Haussler D. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A*. 104(19):8005–10

88. Ludwig MZ, Bergman C, Patel NH, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*. 403(6769):564–67

89. Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* 43(11):1154–59

90. Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. *Evolution*. 62(9):2131–54

91. Macdonald PM, Ingham P, Struhl G. 1986. Isolation, structure, and expression of even-skipped: a second pair-rule gene of drosophila containing a homeo box. *Cell*. 47(5):721–34

92. Malicki J, Schughart K, McGinnis W. 1990. Mouse hox-2.2 specifies thoracic segmental identity in drosophila embryos and larvae. *Cell*. 63(5):961–67

93. Maricic T, Günther V, Georgiev O, Gehre S, Curlin M, et al. 2013. A recent evolutionary change affects a regulatory element in the human foxp2 gene. *Mol. Biol. Evol.* 30(4):844–52

94. Martin A, McCulloch KJ, Patel NH, Briscoe AD, Gilbert LE, Reed RD. 2014. Multiple recent co-options of optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo*. 5(1):7

95. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, et al. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a wnt ligand

96. McCauley BS, Weideman EP, Hinman VF. 2010. A conserved gene regulatory network subcircuit drives different developmental fates in the vegetal pole of highly divergent echinoderm embryos. *Dev. Biol.* 340(2):200–208

97. McCauley BS, Wright EP, Exner C, Kitazawa C, Hinman VF. 2012. Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. *Evodevo*. 3(1):17

98. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, et al. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*. 448(7153):587–90

99. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, et al. 2011. Human-specific loss of regulatory dna and the evolution of human-specific traits. *Nature*. 471(7337):216–19

100. Moczek A, Emlen D. 2000. Male horn dimorphism in the scarab beetle, onthophagus taurus: do alternative reproductive tactics favour alternative phenotypes? *Anim. Behav.* 59(2):459–66

101. Moczek AP, Nagy LM. 2005. Diverse developmental mechanisms contribute to different levels of diversity in horned beetles. *Evol. Dev.* 7(3):175–85

102. Moczek AP, Rose DJ. 2009. Differential recruitment of limb patterning genes during development and diversification of beetle horns. *Proc. Natl. Acad. Sci. U. S. A.* 106(22):8992–97

103. Moczek AP, Rose D, Sewell W, Kesselring BR. 2006. Conservation, innovation, and the evolution of horned beetle diversity. *Dev Genes Evol*. 216(11):655–65

104. Molkentin JD, Lin Q, Duncan SA, Olson EN. 1997. Requirement of the transcription factor gata4 for heart tube formation and ventral morphogenesis. *Genes Dev.* 11(8):1061–72

105. Morgan TH, Bridges CB. 1916. Sex-linked inheritance in drosophila. *Carnegie Inst. Washingt. Publ.* 237:1–88

106. Ng CS, Hamilton AM, Frank A, Barmina O, Kopp A. 2008. Genetic basis of sex-specific color pattern variation in drosophila malerkotliana. *Genetics*. 180(1):421–29

107. Nusslein-Volhard C, Wieschaus E, Kluding H. 1984. Mutations affecting the pattern of the larval cuticle in drosophila melanogaster. *Roux's Arch. Dev. Biol.* 193:267–82

108. Oliveri P, Tu Q, Davidson EH. 2008. Global regulatory logic for specification of an embryonic cell lineage. *Proc. Natl. Acad. Sci. U. S. A.* 105(16):5955–62

109. Ordway AJ, Hancuch KN, Johnson W, Wiliams TM, Rebeiz M. 2014. The expansion of body coloration involves coordinated evolution in cis and trans within the pigmentation regulatory network of drosophila prostipennis. *Dev. Biol.*, pp. 1–10

110. Panganiban G, Irvine SM, Lowe C, Roehl H, Corley LS, et al. 1997. The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. U. S. A.* 94(10):5162–66

111. Papa R, Morrison CM, Walters JR, Counterman BA, Chen R, et al. 2008. Highly conserved gene order and numerous novel repetitive elements in genomic regions linked to wing pattern variation in heliconius butterflies. *BMC Genomics*. 9:345

112. Pavlopoulos A, Kontarakis Z, Liubicich DM, Serano JM, Akam M, et al. 2009. Probing the evolution of appendage specialization by hox gene misexpression in an emerging model crustacean. *Proc. Natl. Acad. Sci. U. S. A.* 106(33):13897–902

113. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*. 444(7118):499–502

114. Peter IS, Davidson EH. 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* 340(2):188–99

115. Peter IS, Davidson EH. 2011. A gene regulatory network controlling the embryonic specification of endoderm. *Nature*. 474(7353):635–39

116. Pisani D, Feuda R, Peterson KJ, Smith AB. 2012. Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Mol. Phylogenet. Evol.* 62(1):27–34

117. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, et al. 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2(10):e168

118. Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, et al. 2006. An rna gene expressed during cortical development evolved rapidly in humans. *Nature*. 443(7108):167–72

119. Pool JE, Aquadro CF. 2007. The genetic basis of adaptive pigmentation variation in drosophila melanogaster. *Mol Ecol*. 16(14):2844–51

120. Prabhakar S, Noonan JP, Pääbo S, Rubin EM. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science*. 314(5800):786

121. Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science (80-.).* 321(5894):1346–50

122. Quiring R, Walldorf U, Kloter U, Gehring WJ. 1994. Homology of the eyeless gene of drosophila to the small eye gene in mice and aniridia in humans. *Science*. 265(5173):785–89

123. Rafiq K, Cheers MS, Ettensohn CA. 2012. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development*. 139(3):579–90

124. Range R, Lapraz F, Quirin M, Marro S, Besnardeau L, Lepage T. 2007. Cis-regulatory analysis of nodal and maternal control of dorsal-ventral axis formation by univin, a tgf-beta related to vg1. *Development*. 134(20):3649–64

125. Rebeiz M, Castro B, Liu F, Yue F, Posakony JW. 2012. Ancestral and conserved cis-regulatory architectures in developmental control genes. *Dev. Biol.* 362(2):282–94

126. Rebeiz M, Jikomes N, Kassner VA, Carroll SB. 2011. The evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A*. 108(25):10036

127. Rebeiz M, Stone T, Posakony JW. 2005. An ancient transcriptional regulatory linkage. *Dev Biol*. 281(2):299–308

128. Rebeiz M, Williams TM. 2011. Experimental approaches to evaluate the contributions of candidate cis-regulatory mutations to phenotypic evolution. *Methods Mol. Biol.* 772:351–75

129. Reed RD, Papa R, Martin A, Hines HM, Counterman BA, et al. 2011. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*. 333(6046):1137–41

130. Ronshaugen M, McGinnis N, McGinnis W. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature*. 415(6874):914–17

131. Röttinger E, Croce J, Lhomond G, Besnardeau L, Gache C, Lepage T. 2006. Nemo-like kinase (nlk) acts downstream of notch/delta signalling to downregulate tcf during mesoderm induction in the sea urchin embryo. *Development*. 133(21):4341–53

132. Royo JL, Maeso I, Irimia M, Gao F, Peter IS, et al. 2011. Transphyletic conservation of developmental regulatory state in animal evolution. *Proc. Natl. Acad. Sci. U. S. A.* 108(34):14186–91

133. Schreiweis C, Bornschein U, Burguiere E, Kerimoglu C, Schreiter S, et al. 2014. Humanized foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl. Acad. Sci.* 111(39):14253–58

134. Sethi AJ, Wikramanayake RM, Angerer RC, Range RC, Angerer LM. 2012. Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. *Science*. 335(6068):590–93

135. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, et al. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*. 428(6984):717–23

136. Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, et al. 2012. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 8(6):e1002789

137. Slack J. 1984. A rosetta stone for pattern formation in animals? *Nature*. 310(5976):364–65

138. Small S, Blair A, Levine M. 1992. Regulation of even-skipped stripe 2 in the drosophila embryo. *EMBO J.* 11(11):4047–57

139. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, et al. 2010. The amphimedon queenslandica genome and the evolution of animal complexity. *Nature*. 466(7307):720–26

140. St Johnston D, Driever W, Berleth T, Richstein S, Nüsslein-Volhard C. 1989. Multiple steps in the localization of bicoid rna to the anterior pole of the drosophila oocyte. *Dev. Cambridge Engl.* 107 Suppl:13–19

141. Stern DL. 1998. A role of ultrabithorax in morphological differences between drosophila species. *Nature*. 396(6710):463–66

142. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43(11):1160–63

143. Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*. 424(6951):935–38

144. Swanson CI, Schwimmer DB, Barolo S. 2011. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* 21(14):1186–96

145. Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (galago crassicaudatus). nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol.* 203(2):439–55

146. Takahashi A, Takahashi K, Ueda R, Takano-Shimizu T. 2007. Natural variation of ebony gene controlling thoracic pigmentation in drosophila melanogaster. *Genetics*. 177(2):1233–37

147. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. 2007. Convergent adaptation of human lactase persistence in africa and europe. *Nat Genet*. 39(1):31–40

148. Tobler A, Kapan D, Flanagan NS, Gonzalez C, Peterson E, et al. 2005. First-generation linkage map of the warningly colored butterfly heliconius erato. *Heredity (Edinb)*. 94(4):408–17

149. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, et al. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40(2):158–60

150. Walter MF, Black BC, Afshar G, Kermabon AY, Wright TR, Biessmann H. 1991. Temporal and spatial expression of the yellow gene in correlation with cuticle formation and dopa decarboxylase activity in drosophila development. *Dev. Biol.* 147(1):32–45

151. Wang VY, Hassan BA, Bellen HJ, Zoghbi HY. 2002. Drosophila atonal fully rescues the phenotype of math1 null mice: new functions evolve in new cellular contexts. *Curr. Biol.* 12(18):1611–16

152. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in drosophila. *Cell*. 134(4):610–23

153. Wittkopp PJ. 2005. Genomic sources of regulatory variation in cis and in trans. *Cell. Mol. Life Sci.* 62(16):1779–83

154. Wittkopp PJ, Carroll SB, Kopp A. 2003. Evolution in black and white: genetic control of pigment patterns in drosophila. *Trends Genet*. 19(9):495–504

155. Wittkopp PJ, Vaccaro K, Carroll SB. 2002. Evolution of yellow gene regulation and pigmentation in drosophila. *Curr Biol*. 12(18):1547–56

156. Wittkopp PJ, Williams BL, Selegue JE, Carroll SB. 2003. Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc Natl Acad Sci U S A*. 100(4):1808–13

157. Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 8(3):206–16

158. Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. sixth Int. Congr. Genet.* 1(6):356–66

159. Yeh S-D, Liou S-R, True JR. 2006. Genetics of divergence in male wing pigmentation and courtship behavior between drosophila elegans and d. gunungcola. *Heredity (Edinb).* 96(5):383–95

160. Yuh C-H, Brown CT, Livi CB, Rowen L, Clarke PJC, Davidson EH. 2002. Patchy interspecific sequence similarities efficiently identify positive cis-regulatory elements in the sea urchin. *Dev. Biol.* 246(1):148–61

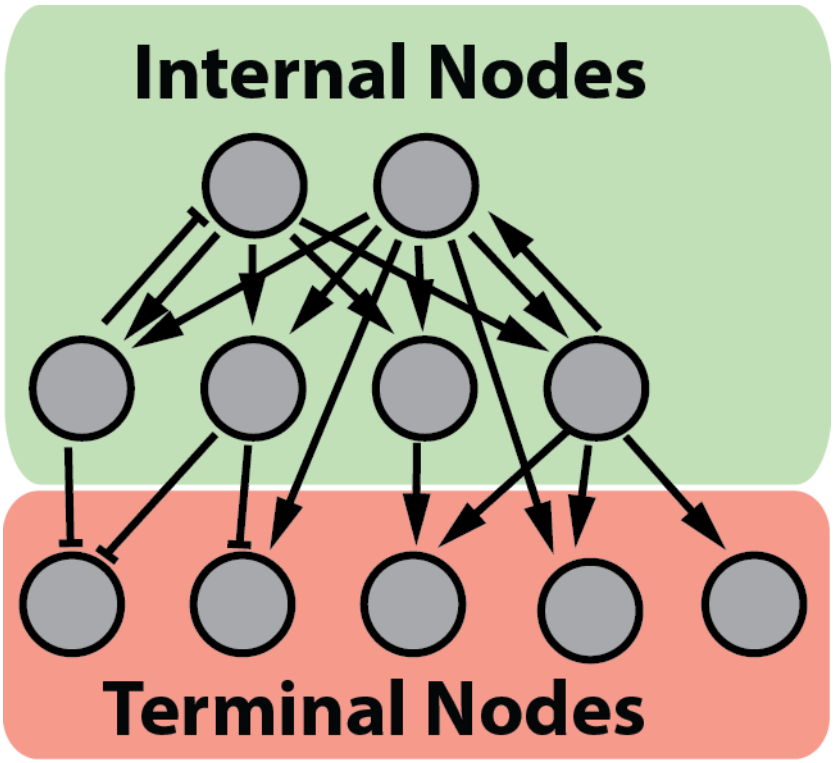

Figure 1. Abstract Gene Regulatory network depicting internal and terminal nodes. Each gray circle is a gene that either activates (arrow) or represses (horizontal bar) its targets. Nodes that affect the expression of other nodes are internal (green shading), while terminal nodes do not impinge upon other nodes (pink).

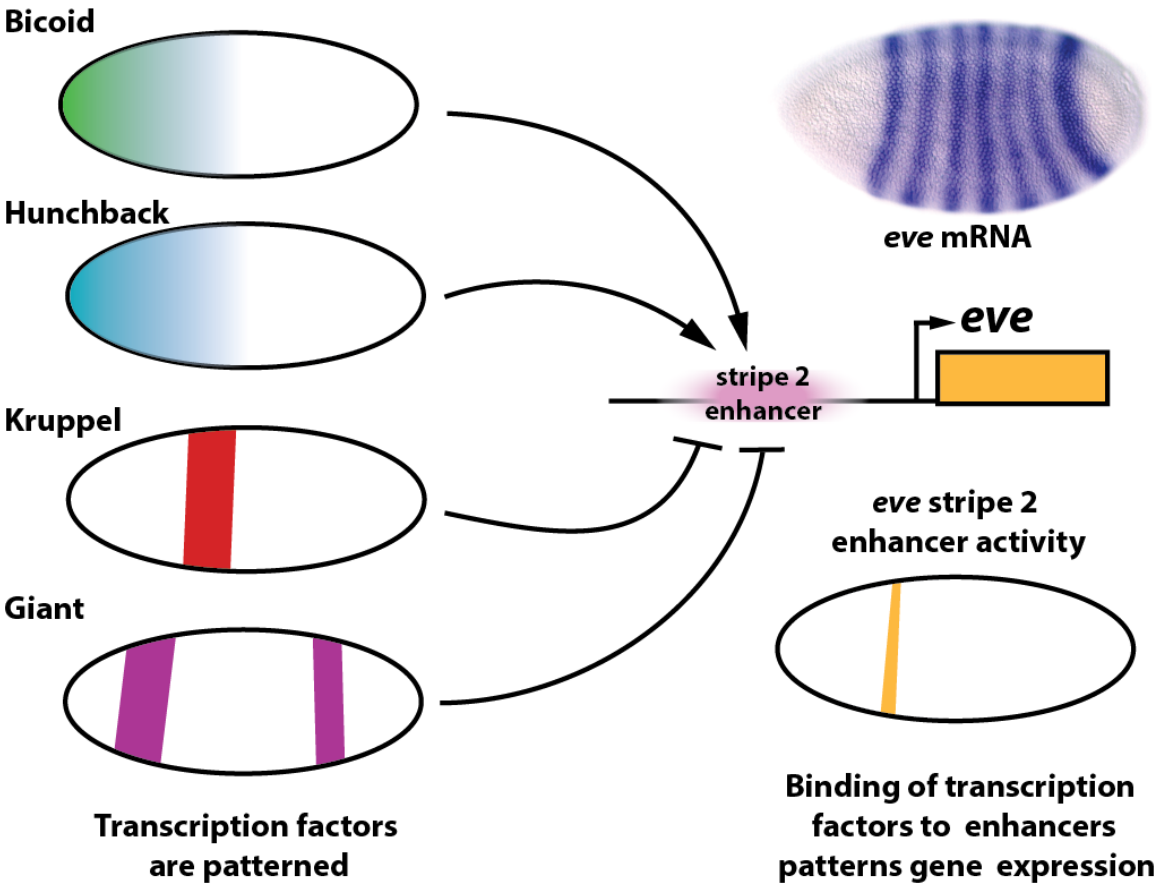

Figure 2. Patterned transcription factors act at enhancers to sculpt downstream gene expression. (left) The gap genes are transcription factors that define broad zones of the early *Drosophila* embryo, in part through their regulation of the *even-skipped* (*eve*) gene regulatory region. (right) *eve* is expressed in seven stripes in the early *Drosophila* embryo. The stripe 2 enhancer of *eve* is regulated by the gap gene transcription factors, which activate and repress the enhancer to generate a thin stripe of expression at a precise register along the embryo.

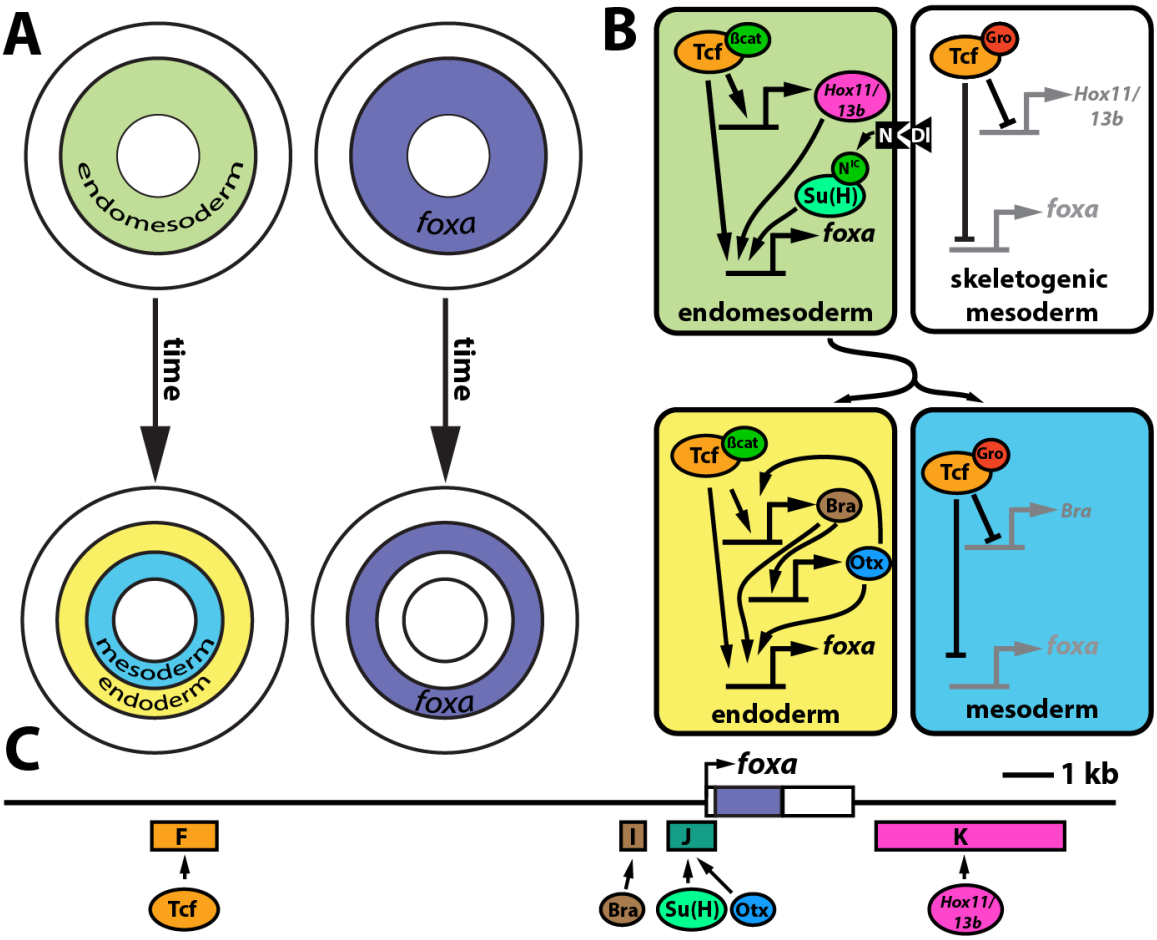

Figure 3. Dynamic regulation of *foxa* by multiple enhancer modules. (A) (left) Vegetal view of a sea urchin embryo, showing the endomesoderm, which segregates into endoderm and mesoderm during development. (right) *foxa* mRNA is first expressed throughout the endomesoderm, and is subsequently maintained in the endoderm. (B) GRN architecture that controls *foxa* expression during endoderm development. A Notch-Delta signal from the skeletogenic mesoderm activates the Su(H) transcription factor. (C) Schematic of the *foxa* locus, detailing the relative positions of four modules that integrate diverse inputs in the endomesoderm network to control dynamic *foxa* expression. Abbreviations: N, Notch; N^{IC}, Notch intracellular domain; Dl, Delta; Su(H), Suppressor of Hairless; βcat, β-catenin; Bra, Brachyury; Gro, Groucho.

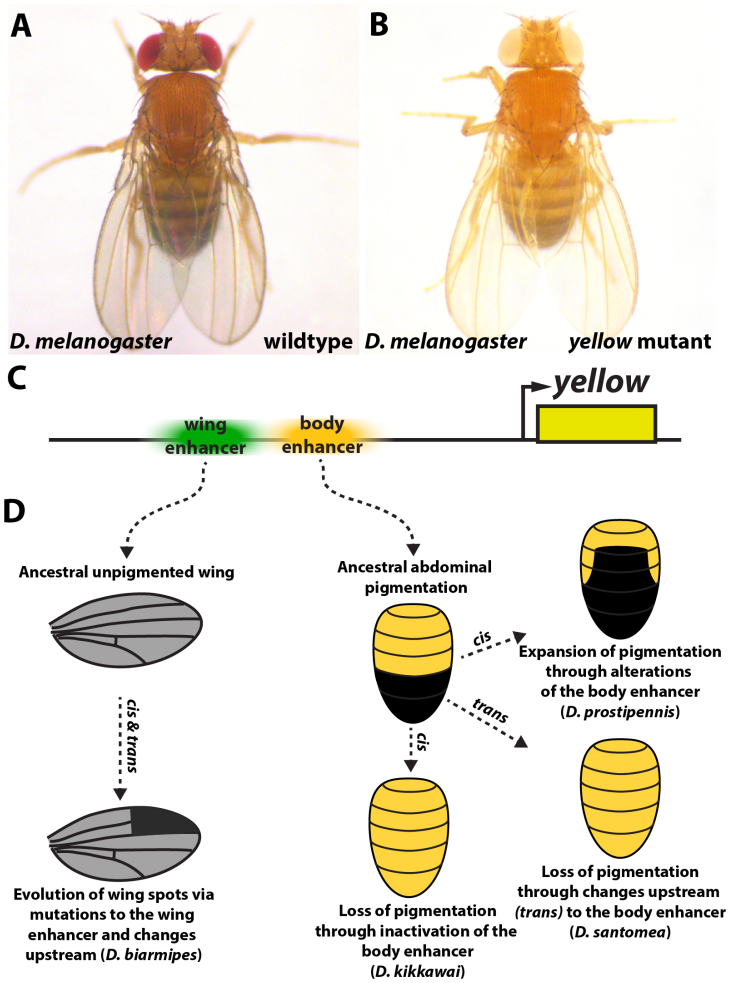

Figure 4. Examples of alterations to a terminal GRN node, the pigment-promoting enzyme gene *yellow.* (A) Pigmentation phenotype of a wildtype *D. melanogaster* male shows dark pigmentation of the posterior abdomen, bristles, and wing. (B) *D. melanogaster* male mutant for the *yellow* gene, causing a lightening of the body, bristles, and wings. The imaged individual is also mutant for the *white* gene, causing a loss of red eye color. (C) Schematic of the *yellow* gene, highlighting the relative position of wing and body enhancers, elements that have been altered in *cis* and *trans* during the evolution of pigment phenotypes. (D) Modifications to the wing and body elements of the *yellow* gene. (left) A combination of mutations to the wing enhancer, as well as upstream factors that regulate its activity underlie the transition from an ancestrally uniformly gray wing (top) to the dark spot of pigmentation present in *D. biarmipes* (bottom). (right) Three independent alterations of the body enhancer accompanied the expansion and loss of ancestral male-specific pigmentation. In *D. kikawai,* the enhancer was inactivated by *cis* changes, while a similar loss in *D. santomea* occurred entirely through changes in upstream factors. The expansion of pigmentation in *D. prostipennis* was traced to changes within the body element.

Figure 5. Mullerian mimicry among *Heliconius* subspecies. Images of subspecies of *H. erato* (top), and *H. melpomene* (bottom) that have evolved co-mimic phenotypes to share the burden of educating predators of their unpalatability.

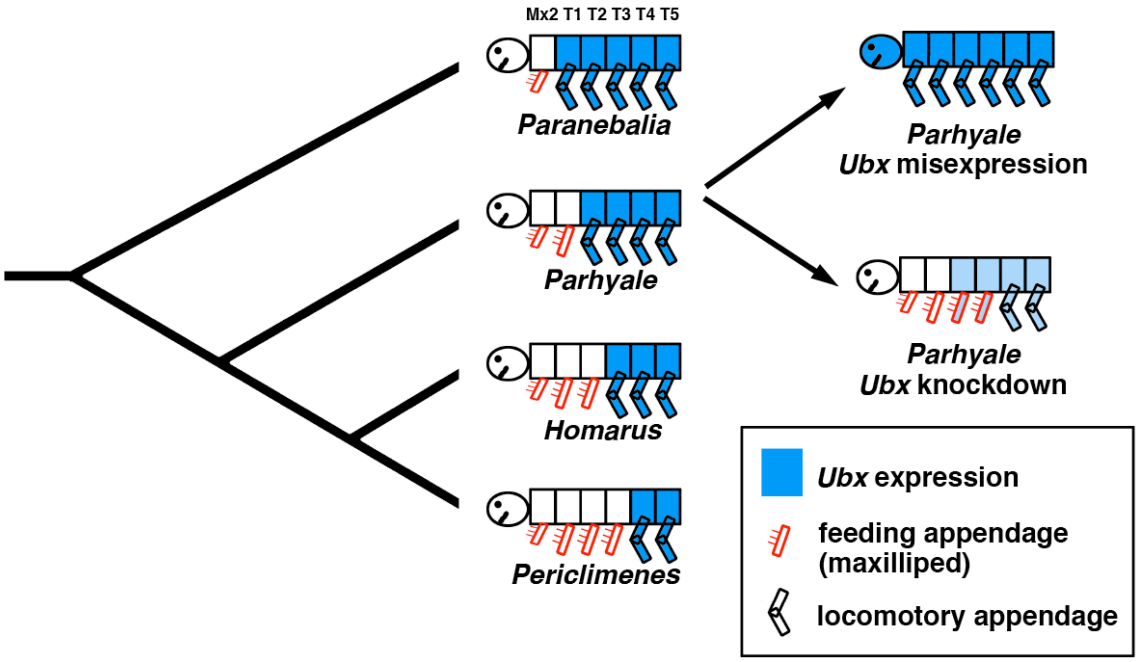

Figure 6. Diversification of crustacean limb patterns by changes in Ubx expression. The anterior boundary of *Ubx* expression (blue) correlates with the morphological transition between locomotory appendages (black), and feeding appendages (maxillipeds, red) in several species of crustaceans. (right) RNAi experiments in *Parhyale* that reduce *Ubx* function transform locomotory appendages into maxillipeds. In contrast, *Ubx* misexpression results in the conversion of maxillipeds into locomotory appendages.

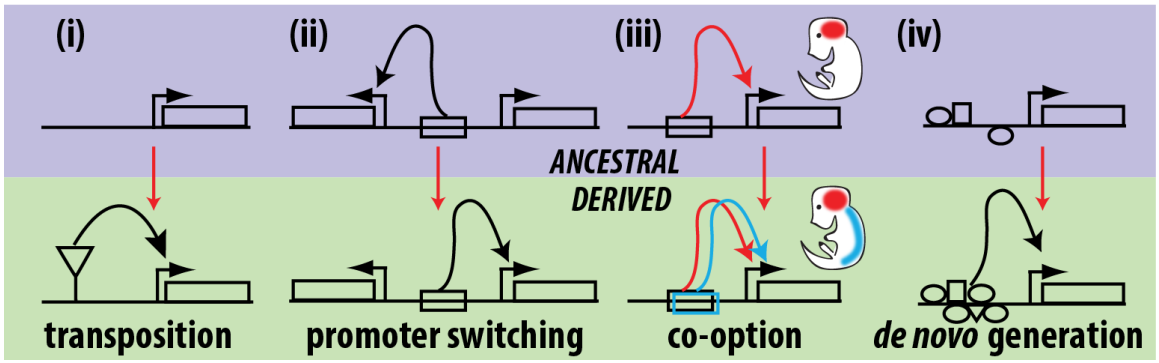

Figure 7. Models for enhancer origination (i) The insertion of a transposable element, which contains functional enhancer sequences may activate the adjacent gene. (ii) A pre-existing enhancer may evolve changes in the control of enhancer-promoter specificity that cause it to interact with an alternate promoter. (iii) The alteration of a pre-existing enhancer may co-opt it to drive a new expression pattern in an additional tissue. (iv) The evolution of new binding sites in otherwise non-functional DNA may result in the *de novo* generation of an enhancer sequence.

Assay	Description of method	Direct or Indirect Linkage
Genetic test	Measurement of a target gene's expression in a background for which a regulating gene's function has been manipulated (genetically or by RNAi or Morpholino knockdown)	Indirect Linkage
DNA Binding Assay	A small portion of the regulatory region is detected to bind the factor in an electrophoretic mobility shift assay	Direct Linkage
Chromatin Immunoprecipitation	The region of interest binds the factor *in vivo* as measured by the recovery of the bound region upon formaldehyde fixation and antibody pull down for a particular DNA-binding factor	Direct Linkage
Reporter Assay Mutation	Introduction of a mutation that will disrupt the binding of a factor in a reporter assay shows the predicted effect on reporter gene expression (e.g. increase, decrease, expansion or contraction)	Direct Linkage

Table 1. Methods for inferring regulatory connections within GRNs