Genetic Networks

Stefan C Materna, *School of Natural Sciences, University of California Merced, Merced, California, USA*

Stephanie Woo, *School of Natural Sciences, University of California Merced, Merced, California, USA*

Based in part on the previous versions of this eLS article 'Genetic Networks' (2001, 2007).

Introductory article

Article Contents

Online posting date: 14th June 2018

Genetic networks control the execution of the genetic program stored in an organism's deoxyribonucleic acid (DNA) by orchestrating gene expression. Any biological function, in physiology or development, is dependent on the combined action of many genes, requiring their precise control. Expression of individual genes is determined by associated regions of regulatory DNA. These are bound by sequence-specific regulatory proteins, called transcription factors, leading to activation or repression of transcription. The interactions among regulatory genes display the features of a network where the linkages are determined by the binding sites in the regulatory region of downstream genes. The architecture of genetic networks is intrinsically hierarchical, and discrete subcircuits that accomplish particular tasks can be identified. Certain linkage patterns are recurrent in subcircuits with similar biological function, although the regulatory genes involved differ. This suggests that the topology of the network, and not the identity of its constituents, determines the function.

Introduction

Cellular function is determined by proteins, which carry out the countless tasks that underlie the structure and activity of every cell. The phenotype of any given cell is thus largely determined by the proteins that it contains, which is in turn a function of gene expression. Implicit in the term 'gene expression' is the fact that any given cell uses or 'expresses' only a fraction of the genetic information contained in its deoxyribonucleic acid (DNA) to produce protein. Cells are complex systems, and function thus requires the coordinated expression of many genes. This control is exerted by genetic networks, which are sets of regulatory genes that are functionally linked through sequence-specific interactions. Genetic networks thus are the source of regulatory programming in the genome.

It has long been known that heritable information determines the phenotypic traits of an organism and that this information is transmitted from generation to generation via the organism's genes. Gregor Mendel established the fact of 'particulate', or genetic, inheritance between 1853 and 1863, by showing experimentally in pea plants that phenotypic traits of a sexually reproducing organism behave as though they are determined by the action of discrete factors (genes) that follow statistical laws of assortment. The term 'gene' was initially defined in terms of function, which was revealed through mutant phenotypes. It thus included what we think of as 'genes' today, that is transcribed regions that are processed into message and code for a particular protein, but it also included those regions that exert transcriptional control of associated genes (Lewis, 1978). Loss of either functionality may result in aberrant phenotypes. It is now clear that the fraction of the genome with regulatory function matches in size the fraction coding for proteins. **See also: Mendel, Gregor Johann**; **Genes: Definition and Structure**

Genetic networks are made up of regulatory genes, that is transcription factors and signalling molecules. Transcription factors bind DNA in a sequence-specific manner to modulate the transcription of genes nearby. Signalling factors are secreted ligands that enable intercellular communication and cause transcriptional changes in the receiving cell. The expression of any given gene is typically influenced by the activities of several different 'upstream' regulatory genes. If the gene is itself a regulatory gene, it will affect the expression of numerous other 'downstream' genes. In consequence, the interactions among multiple regulatory genes display the features of a network, where regulatory genes form the nodes and the edges represent their linkages to downstream genes. As we will discuss, individual regulatory modules are the information-processing units that interpret the regulatory state of a cell, given by the transcription factors currently expressed, and determine the transcriptional

eLS subject area: Molecular Biology

How to cite:
Materna, Stefan C and Woo, Stephanie (June 2018) Genetic Networks. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0001170.pub3

state of individual genes. **See also:** **Transcription Factors**; **Transmembrane Signalling**

Biological function emerges through the combinatorial use of regulatory genes that control particular physiological or developmental events. The fact that some linkage patterns are recurrent and appear in subcircuits with similar biological function indicates that it is the topology of the network that determines function and not the identity of individual genes. Maps of well-studied genetic networks can be complicated, and it is easy to treat them as abstract ideas. However, it is important to remember that they represent a physical reality, that is the molecular interactions that underlie the execution of genomic information.

The examples in this article primarily focus on genetic networks in organismal development. Development is characterised by an increase in spatial complexity, and, as a consequence, many functions executed by a particular piece of genetic network circuitry have a spatial outcome. This is most obvious in the institution of divergent gene expression patterns that underlies the patterning of the embryo. Nonetheless, genetic networks also control the response to physiological stimuli wherever this requires the execution of genomic programming. Unlike in development, the resulting changes in gene expression are often temporal rather than spatial. A good example is the control of the cell cycle that, despite its many physiological check points, is at its core a genetic network.

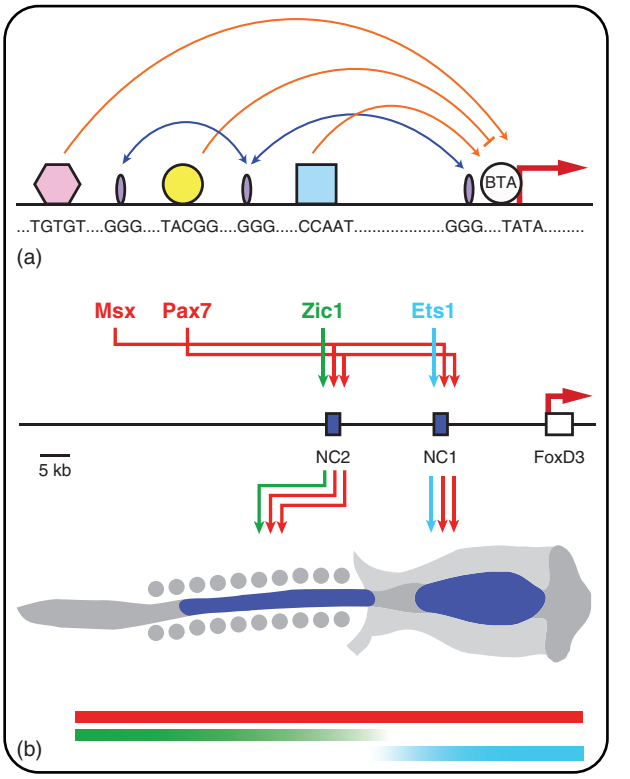

Figure 1 Gene regulatory domains are modular. (a) In a hypothetical example, the gene is represented by a horizontal line and the transcription start site by the bent red arrow. Coloured shapes represent the proteins bound to the DNA at their specific binding sites. The activity of the basal transcription apparatus (BTA) is positively affected by the proteins bound at sites TGTGT and CCAAT and negatively by the protein bound at site TACGG (orange lines). The proteins bound at sites GGGG interact with each other, causing the DNA to loop and thereby facilitating the interactions between the other regulatory proteins and the BTA (blue lines). (b) Modular gene regulation in neural crest cells. The transcription of the *foxD3* gene in the neural crest of chickens (shaded blue area of schematised chick embryo, anterior to the right) is controlled by two regulatory modules, NC1 and NC2. Both receive the transcription factors Msx and Pax7, which are expressed throughout the entire neural crest (red bar below), as activating inputs. NC1 also requires binding of Ets1, but its expression is restricted to the head region (blue bar), thus restricting activity of this module to the cranial neural crest. NC2 is bound by the Zic1 protein that is present in the trunk neural crest (green bar), driving FoxD3 expression there.

Principles of Gene Regulation

Structure, function and logic of genetic regulatory systems

The regulation of gene expression occurs primarily at the level of transcription – that is, through the process of turning genomic sequence into translatable messenger ribonucleic acid (RNA). The molecular mechanisms of transcription are best understood in bacteria where activator or repressor proteins bind regulatory sequence directly upstream of the site of transcription initiation in response to physiological stimuli to initiate transcription. The situation is vastly more complex in eukaryotic cells that contain genomes magnitudes larger than those of bacteria. In eukaryotes, genes for proteins that act in the same pathway are often dispersed throughout the genome and must be controlled individually, whereas in bacteria they are packaged in single transcription units (operons). In eukaryotes, DNA is stored in the nucleus and bound up in a protein scaffold composed of histone proteins, which can be chemically modified to alter the accessibility of particular sites, or entire regions. Such epigenetic modifications, that is modifications that do not alter the DNA sequence, contribute to long-term maintenance of gene silencing or gene activity. **See also:** **Transcription Activation at Bacterial Promoters**; **Histone Acetylation: Long-range Patterns in the Genome**; **Chromatin Structure and Domains**; **Transcription Activation in Eukaryotic Cells**; **Chromatin Remodelling and Histone Modification in Transcription Regulation**

The rate of transcription is largely a function of the frequency of initiation by the enzyme RNA polymerase. In eukaryotes, three different ribonucleic acid polymerases (RNAPs) are responsible for transcription. Coding genes, such as those that make up genetic networks, are transcribed by RNAP II (RNAP I transcribes ribosomal RNAs, and RNAP III transcribes small RNAs, e.g. tRNAs). RNAP II is part of a large multiprotein complex containing numerous proteins, called basal transcription factors, which usually bind the sequence 'TATA' near the site of transcription initiation (**Figure 1a**). This multiprotein complex is known as the basal transcription apparatus (BTA) and stabilises the interaction of the polymerase enzyme with DNA in preparation for transcription initiation. However, through its size and complexity, it also provides a multifaceted regulatory interface that allows cells to differentially transcribe genes in response to specific signals.

Such signals are ultimately transmitted to the gene by transcriptional regulatory proteins, that is transcription factors that interact with specific DNA sequences in the vicinity of the gene, and are thereby brought into relative proximity to the transcription initiation site and the BTA. This proximity facilitates interaction between the regulatory proteins and the BTA, which in turn influences the activity of the RNAP. The fact that transcriptional regulatory proteins bind specific DNA sequences means that the regulation of a gene's expression is genetically programmed. **See also: RNA Polymerase II Holoenzyme and Transcription Factors**; **RNA Polymerases: Subunits and Functional Domains**

The region of DNA within a gene that contains target sequences for transcriptional regulatory proteins is the 'cis-regulatory domain' of the gene. Here, cis refers to positioning of the regulatory region on the same strand of DNA, usually in the vicinity to, or even within, the transcribed region. The complexity of genetic cis-regulatory domains varies. Genes that are expressed ubiquitously and require little regulation typically contain relatively simple cis-regulatory domains. In contrast, genes that are expressed in specific cells, at specific times, or in response to specific physiological signals tend to have more complex regulatory domains. The most complex cis-regulatory domains are found in genes that are expressed in multiple cell types, but at different times in development. The regulation of such genes is usually modular – that is, carried out by discrete regions or 'modules' (sometimes referred to as 'enhancers') of cis-regulatory DNA.

Each cis-regulatory module contributes a particular function to the overall control of the gene's expression and may determine timing, location or amplitude. Each module typically contains target sequences for several different transcriptional regulatory proteins, which include both activators and repressors. Many regulatory modules for genes with complex expression patterns have been studied in detail. One example is depicted in **Figure 1b** and shows the regulation of *foxD3* transcription in the neural crest of chick embryos (Simões-Costa *et al.*, 2012). Two regulatory modules, NC1 and NC2, have been identified whose combined activity reproduces the expression pattern of the endogenous gene. Both regulatory modules are bound by Msx and Pax7, two transcription factors that are expressed throughout the entire neural crest. However, NC1 requires an additional activating input from Ets1, and all the three proteins must be bound for NC1 to activate transcription. Ets1 expression is restricted to the head region, and, as a consequence, NC1 controls *foxD3* specifically in the cranial neural crest. Zic1 is a required input into NC2 but expressed only in trunk neural crest, thus limiting NC2 activity to this region.

As in prokaryotes, regulatory modules may be located directly adjacent to the transcriptional start site of the gene they regulate. However, they may also be located at distances of tens of thousands of base pairs away, as is the case in the example above. Because transcription factors interact with the BTA directly or through adapter proteins, these distal regulatory modules must be moved to the vicinity of the BTA through looping of the DNA (**Figure 1a**). This looping is accomplished by the interaction of proteins that bind to the active regulatory module with proteins that bind in the vicinity of, or directly to, the BTA. The sequence directly upstream of the site of transcription initiation thus often functions as a landing pad for distant regulatory modules. **See also: DNA Looping and Transcription Regulation**

In essence, the cis-regulatory domain of a gene constitutes an information-processing system that is encoded in the DNA. The 'inputs' processed by the system are the activities of transcription factors for which the system has target sites, and which are present in the nucleus at a particular moment. The logic of the processing is specified by the target sites, packaged in discrete modules, which anchor regulatory proteins and thus determine what interactions these proteins engage in. The final 'output' of the cis-regulatory system's information processing is simply the rate of transcription initiation.

Genetic network architecture

The number of regulatory genes in the genome is finite and, in animal genomes, accounts for a rather small fraction of protein coding genes of typically less than 5%. Thus, any particular biological process usually requires the activity of several different regulators that are used in varying combinations to achieve specific outcomes. While the cis-regulatory domain of each gene provides the fundamental information-processing capacity that is hardwired in the genome, the linkages between regulatory genes determine the flow of information from which specific functions emerge.

The total set of linkages together with the particular response of each gene at the nodes constitutes the architecture of the genetic network. Mapped out as a whole, it is an abstract representation of genomic regulatory function integrated over time and space. The current state of the network of a given cell at a given time, also referred to as its regulatory state, is defined by the transcriptional status of each node in the network. The information carried by sequence-specific transcription factors present in a given nucleus are themselves products of numerous different regulatory genes expressed in the history of the cell lineage that gave rise to that particular nucleus. Thus, the activity of each regulatory gene is linked via cis-regulatory target sequences to the activities of numerous upstream genes, as well as the activities of numerous downstream genes. If one had a complete map of an organism's genetic network architecture and knowledge of the state of the network in each cell at time t, one could in principle predict the subsequent state of the network in each cell at subsequent time $t+1$ and so on. Each regulatory interaction is contingent on the state of other interactions in the network.

Despite their complexity, genetic networks are made more manageable by their modular organisation (**Figure 2**). Within any given genetic network (which in totality occupies the entire genome) there exist numerous functionally linked, subsidiary networks that constitute integrated subcircuits. Subcircuits are minimal assemblages of usually a handful of genes that together execute a particular task. Among others, this may be the interpretation of an intercellular signal, lockdown of a particular cell fate, the exclusion of alternative fates or the sharpening of a boundary between adjacent groups of cells. While the combination of genes working together in a given subcircuit is usually unique and specific to that particular task, more general architectural features can be extracted. Such network motifs are recurrent in different subcircuits where similar tasks are executed. For example, positive

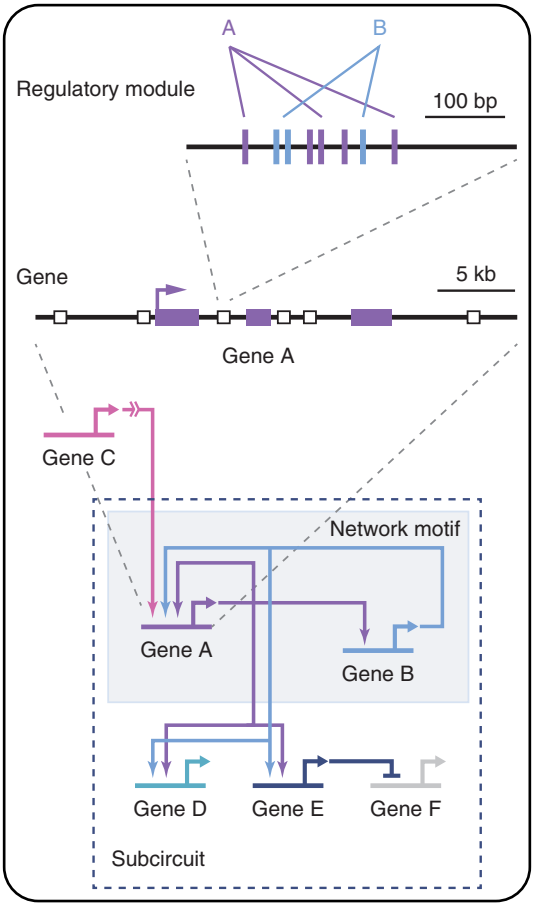

Figure 2 Organisational level of genetic networks. Regulatory modules are short stretches of sequence that contain binding sites for transcription factors that determine whether transcription of the gene they control is initiated. Several regulatory modules (white boxes), each controlling a discrete feature of the overall expression pattern, may be associated with a gene. Genes with more complex expression patterns usually contain more regulatory modules. Genes A and B are connected in a positive feedback loop, and both are direct activating inputs into the regulatory module of gene A, as confirmed by the presence of their binding sites. Simple network motifs as feedback loops are the building blocks of genetic networks and are used in combination to implement particular functions. In this simple example, their mutual activation ensures that the input from gene C is required only transiently. A unique combination of genes is connected in a network subcircuit to achieve a particular biological function. The subcircuit depicted here locks down the regulatory state and leads to expression of genes D and E. Expression of gene E prevents activation of gene F, excluding alternative fates.

feedback loops as between gene A and gene B in **Figure 2** are usually employed downstream of transient activating signals to lock down the new regulatory state making it independent of the initial signal. This suggests that the primary determinant of function is the topology of the subcircuit and not the identity of its constituents.

The structure of genetic networks is inherently hierarchical. Development proceeds from a comparatively simple structure, the single cell of the zygote, to the vastly more complex, fully formed organism. The regulatory processes at each stage determine what will happen next, thereby causing the progressive partitioning of the embryo in time and space. The fact that the state of the network at a preceding stage gives rise to new embryonic regions that are established at the next stage imparts a strict hierarchical organisation. The fate of cells in different regions now depends on the regulatory functions encoded in the genome that are specific to that cell type, which are called upon by the particular set of regulatory genes present in their nuclei. As a consequence, developmental gene expression patterns are often transient and represent the activities of regulatory genes that are internal nodes deep within the genetic network.

Development of multicellular organisms is, at least early on, characterised by spatial decision-making. The immediate outcome is the institution of disparate regulatory states in cells that are descendent from the same progenitor, thus putting them on divergent developmental trajectories. However, development is a highly regulated process that requires coordination across cell types and tissues. Therefore, genes that encode signalling functions are particularly important, as they are the mechanism that links the expression of regulatory genes from one cell to another (**Figure 2**). In other words, the regulatory nodes in network subcircuits are linked together by genes that encode proteins required to convey signals between cells.

The flow or regulatory information in developmental genetic networks is unidirectional. This is a consequence of the nature of gene regulation that depends on the presence of specific binding sites. Thus, a transcription factor will bind to the cis-regulatory domain of a downstream target and activate, or repress, its transcription. The upstream gene controls the downstream gene. Directionality is also ensured by particular subcircuit topologies. In the example in **Figure 2**, the feedback loop between A and B locks down the regulatory state and makes its maintenance independent of the initial activating signal. Although the individual biochemical reactions, such as the binding of a transcription factor to its target sites, are reversible reactions, the developmental process as a whole is irreversible.

In summary, the increase in embryonic complexity is determined by the regulatory functions that emerge from the architecture of the genetic network. Modular subcircuits show recurrent linkage patterns that fulfil particular functions. These are embedded in the larger network hierarchy and drive development forward.

Examples of Developmental Genetic Network Subcircuits

Studies of genetic networks controlling diverse developmental and cellular functions have been reported in the literature. They are the result of sometimes years of experimentation and, in a few cases, have reached astounding complexity. The experimental approach to unravelling genetic networks is twofold: regulatory function can be revealed by interfering with the function of a candidate gene while closely monitoring its effect on potential downstream targets. Alternatively, and starting at the opposite end, cis-regulatory modules that control expression of genes of interest can be isolated and characterised. By identifying the proteins that bind to these pieces of regulatory DNA, a link to the

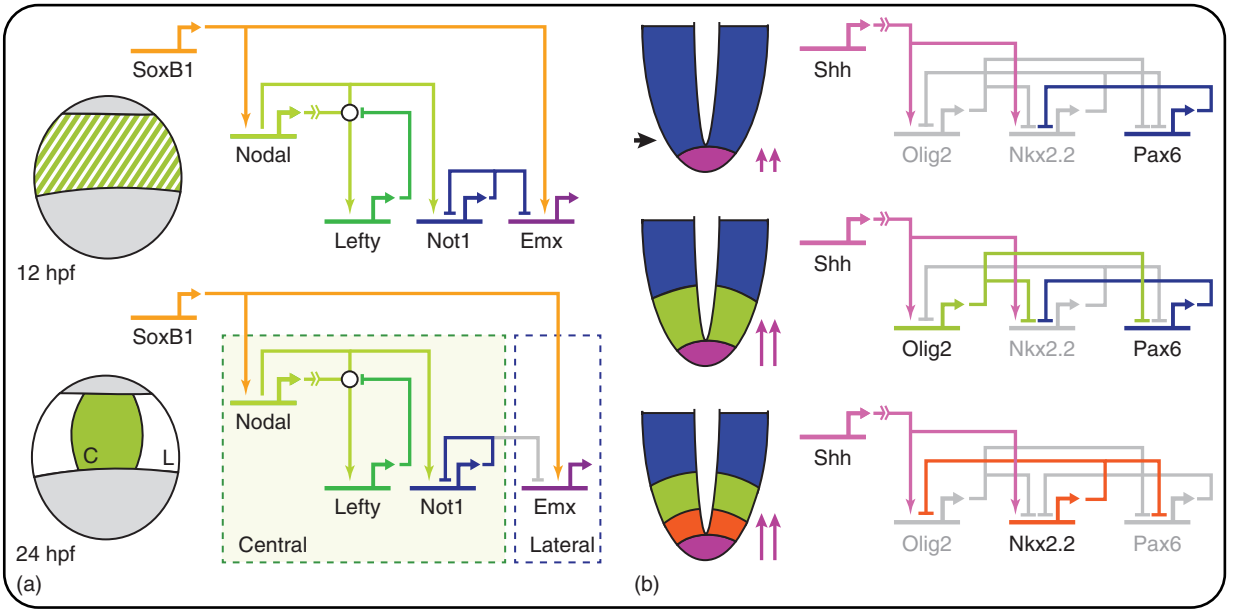

Figure 3 Examples of genetic network subcircuits that pattern embryonic tissues. (a) In the ectoderm of sea urchin embryos, the SoxB1 transcription factor is an activating input into the *nodal* and *emx* genes. Initially, both genes are expressed in overlapping patterns (cross-hatched area in 12 h postfertilisation embryo (hpf)). Nodal is a signalling ligand that activates the *lefty* and *not1* genes. Lefty antagonises Nodal signalling and limits *nodal* and *not1* expression to the central domain (C) of the ventral ectoderm, where later the mouth will form. Not1 represses transcription of the *emx* gene, thus excluding its expression from the central domain. In consequence, *emx* expression is relegated to the lateral domain (L) by 24 hpf and two distinct domains have been established. (b) In the neural tube of vertebrates, the signalling ligand Sonic Hedgehog (Shh) is expressed in the most ventral compartment. It diffuses dorsally into the neural tube (indicated by double arrows) that initially uniformly expresses Pax6. Shh activates expression of several transcription factors that regulate each other as summarised in the wiring diagram on the right. These interactions determine the progression of regulatory states (top to bottom) depicted here as observed in the compartment directly adjacent to the floor plate (black arrow in top panel), where *nkx2.2* expression comes to dominate.

upstream regulators can be established. It is important to emphasise the causal nature of these experiments. Correlation of gene expression patterns alone does not allow inference of causation, although they are a good starting point to query function. Only perturbation of the system, at the gene or DNA level, can reveal the topology of the genetic network under study.

A discussion of entire genetic networks is beyond the scope of this article. To demonstrate some of the salient features of genetic networks, we will review a few examples of network subcircuits.

Subcircuits executing spatial patterning

One of the best-studied genetic networks is the gene regulatory network underlying early sea urchin development (Davidson *et al.*, 2002). In its latest iteration, it covers the progressive establishment of embryonic territories in sea urchin embryos up to the end of gastrulation. Maternally deposited transcripts of regulatory genes establish anisotropies within the early embryo resulting in differential gene activation at the earliest stages. Over time, this gives rise to the embryonic territories that presage the appearance of morphological features. In total, the number of regulatory genes, transcription factors and signalling molecules, which has been incorporated into this network, is approaching triple digits. **See also**: **Sea Urchin Embryo: Specification of Cell Fates**

The dorsal/ventral axis of the sea urchin embryo is determined by Nodal signalling, a member of the Tgf-β family of secreted ligands. Its expression is activated on the future ventral side downstream of a redox gradient and also requires the maternally inherited SoxB1 transcription factor, which at this time is present throughout the apical half of the embryo (**Figure 3a**). Nodal signalling activates transcription of *nodal* in receiving cells, and thus perpetuates its expression through a positive feedback loop. However, one of its direct downstream targets is the *lefty* gene, which encodes a secreted Nodal antagonist that is thought to spread slightly faster than the Nodal protein. In effect, Lefty limits the range of Nodal signalling, and, by disrupting the Nodal positive feedback loop, causes the area of *nodal* expression to shrink. A second downstream target of Nodal is the *not1* gene. It is a repressor that prevents expression of dorsal genes on the ventral side of the embryo. It also represses itself, and, as a consequence clears from cells that no longer receive the Nodal signal, thus mimicking the expression of *nodal* (Li *et al.*, 2013). Eventually these two genes become confined to the central region of the ventral ectoderm, where later the mouth will form. The clearance of Not1 from the lateral ventral domain relieves repression of *emx*, which is also activated by SoxB1. Thus, SoxB1, Nodal/Not1 and Emx form an incoherent feedforward loop (see section titled 'Recurrent Subcircuit Linkage Patterns') that results in the partitioning of the ventral ectoderm into lateral and central domains (Li *et al.*, 2014).

The neural tube of vertebrates is the embryonic structure that will give rise to the central nervous system, that is the brain in

the head and the spinal cord throughout the body. Initially, the neural tube is a uniform cylinder that is then patterned along the dorsoventral axis throughout the body (**Figure 3b**). This establishes different compartments that each expresses a particular combination of regulatory genes. Eventually these compartments will give rise to neurons whose identity is determined by the compartment they originated from. Patterning is controlled by the expression of the signalling factor *Sonic Hedgehog* (Shh) in the ventral most compartment, called the floor plate. Shh diffuses into the neural tube, forming a ventral to dorsal gradient along which the boundaries between the different compartments are defined. For a long time it was assumed that the distance of a particular cell from the source of Shh would determine the concentration of signal received, and that precise concentrations would trigger specific cellular reactions. However, as is now widely appreciated, the resulting pattern is a function of both the activating signal *and* the mutual regulation of the genes thus activated (Balaskas *et al*. 2012). **See also: Regulation of Neuronal Subtype Identity in the Vertebrate Neural Tube (Neuronal Subtype Identity Regulation)**

The subcircuit that determines ventral compartment formation in the neural tube is shown in **Figure 3b**. It is characterised by mutual repression between the transcription factors Olig2, Nkx2.2 and Pax6. The network diagrams in **Figure 3b** show the progression of regulatory states at the level of the compartment adjacent to the floor plate: before the onset of Shh signalling, the transcriptional repressor Pax6 is expressed throughout the neural tube and actively represses *nkx2.2*. Shh activates both *olig2* and *nkx2.2*, but because *nkx2.2* is actively repressed by Pax6, only Olig2 expression is first initiated. However, Olig2 transcriptionally represses *pax6*, thereby relieving repression of *nkx2.2*. Nkx2.2 is a potent repressor of *olig2* expression and is thus able to displace Olig2 adjacent to the floor plate. Further dorsally, where less Shh is received, Olig2 never reaches high enough levels to fully repress *pax6*. Thus *nkx2.2* expression never gathers enough steam to overcome repression by Pax6. Other mechanisms will subsequently take over to lock down these expression patterns. It should be noted that the particular subcircuit discussed here covers only the establishment of the ventral-most compartments. Additional genes are required for the formation of more dorsally located compartments.

Subcircuits controlling morphogenetic events

The example shown in **Figure 4a** depicts a patterning event with a morphogenetic outcome (Christiaen *et al*., 2008). Sea squirts are basic chordate animals that form a simple heart tube although they do not have a closed vascular system. The origin of the heart can be traced back to two clusters of four cells located at the anterior end of the trunk. These cells express the transcription factor Mesp, a highly conserved gene that is similarly involved in heart formation in vertebrates. Mesp activates the transcription factor Ets.b, and together they activate another transcription factor,

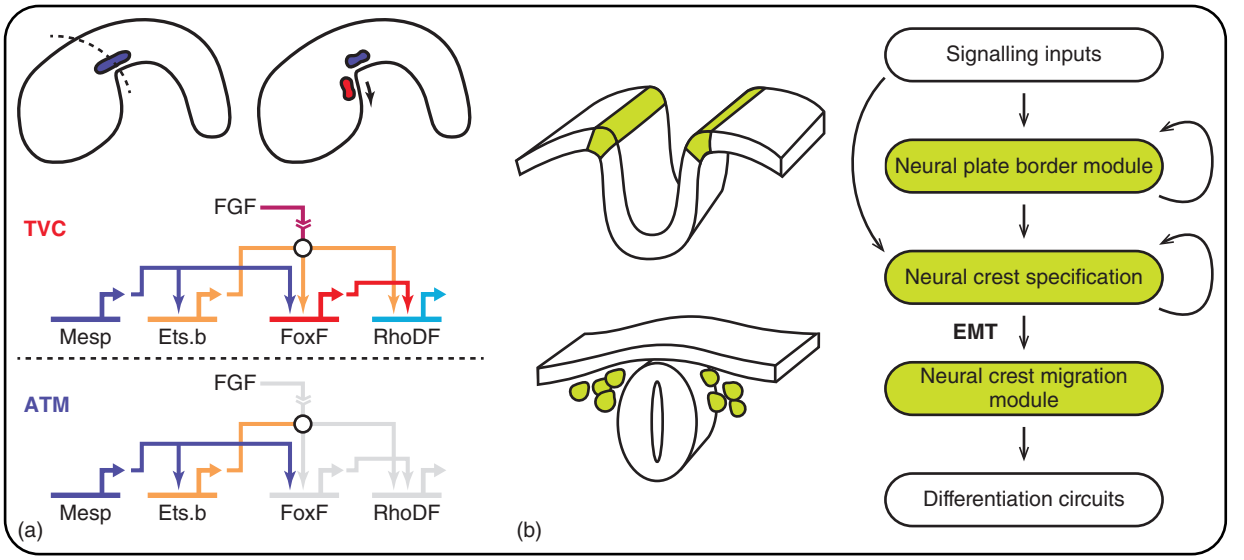

Figure 4 Genetic network control of morphogenetic events. (a) Subcircuit controlling migration of heart progenitors in the sea squirt *Ciona intestinalis*. The mesodermal progenitor cells (purple patches) are located in two lateral patches of four cells at the anterior end of the trunk. They express Mesp, which controls transcription of Ets.b. FGF signalling in the trunk is received in the two anterior progenitor cells where it activates Ets.b through phosphorylation, leading in turn to initiation of FoxF transcription. In consequence, the Rho GTPase RhoDF is activated in the two anterior cells, which affects actin cytoskeleton dynamics and thus promotes migration of trunk ventral cells (TVC, red patch). The cells left behind differentiate into anterior tail muscles (ATM). (b) Process diagram for neural crest development in vertebrates. Neural crest specification is set in motion by first demarcating the neural plate border region downstream of various signalling inputs. As the neural tube forms, neural crest-specific genes are activated along the ectodermal ridges (green, top panel) that in turn lock down the neural crest-specific regulatory state. The neural crest specification genes cause cells to become migratory and undergo epithelial-to-mesenchyme transition. Migrating neural crest cells express a new set of regulatory genes that determine their behaviour according to their position along the body axis and presage what they will eventually differentiate into.

FoxF. But Ets.b function is dependent on chemical modifications, namely phosphorylation of particular amino acids, which are placed as the result of Fgf signalling. Fgf is active throughout the trunk of the *Ciona* embryo but only reaches far enough into the tail to be received in the two anteriormost cells. As a result, FoxF expression is limited to these two cells where it activates the effector gene *rhoDF*. The RhoDF protein is an enzyme involved in remodelling the cytoskeleton and thus a part of the program that installs a migratory phenotype. Although RhoDF is not a regulatory gene as defined before, it nevertheless holds sway over the fate of the trunk ventral cells by enabling their migration to a new environment that allows their differentiation into heart cells. The two remaining cells differentiate into tail muscle. **See also: Cell Motility**; **Cell Migration during Development**

Development is characterised by cell movements that dramatically alter the morphology of the embryo. These movements must be coordinated and are thus one of the immediate results of genetic network subcircuits. This is of particular importance in vertebrate embryos that, unlike many protostomes, do not have fixed lineages and where cell fate is determined almost exclusively in response to inductive signalling. Thus, such rearrangements propel cells into new regulatory environments that determine their eventual fate. One striking example is the specification and migration of neural crest cells in vertebrates. These cells are specified at the neural plate border around the time of neurulation (**Figure 4b**). They adopt a migratory phenotype and relocate to a number of different regions and contribute to craniofacial tissues, sensory organs, the heart and the central nervous system, among others. The genetic network underlying neural crest specification has been mapped in great detail (Simões-Costa and Bronner 2015). It is summarised as a process diagram in **Figure 4b**, where each of the boxes contains at least a dozen interconnected regulatory genes. FoxD3 and its upstream regulators Msx, Pax7, Zic1 and Ets1 are part of this network that in the end controls events not unlike those discussed in the migration of heart progenitors in *Ciona,* but on a larger scale requiring more complex molecular decision-making. **See also: Neural Crest: Origin, Migration and Differentiation**; **Genetics of Non-alcoholic Fatty Liver Disease**

Recurrent Subcircuit Linkage Patterns

Detailed studies of actual genetic networks such as those described earlier have revealed a number of recurring architectural modules. The fact that no two subcircuits that accomplish a similar task (e.g. lockdown of cell fate and exclusion of alternative fates) are put together from the same individual regulatory genes suggests that function does not depend on the particular biochemical properties. Instead, function is entirely due to the topological design of the subcircuit. In this section, we will briefly discuss the properties of a few of such recurrent linkage patterns (**Figure 5**). As more genetic networks are studied in detail, this list is likely to expand.

Positive feedback loops are one of the most ubiquitous patterns in genetic networks. They usually involve at least two genes

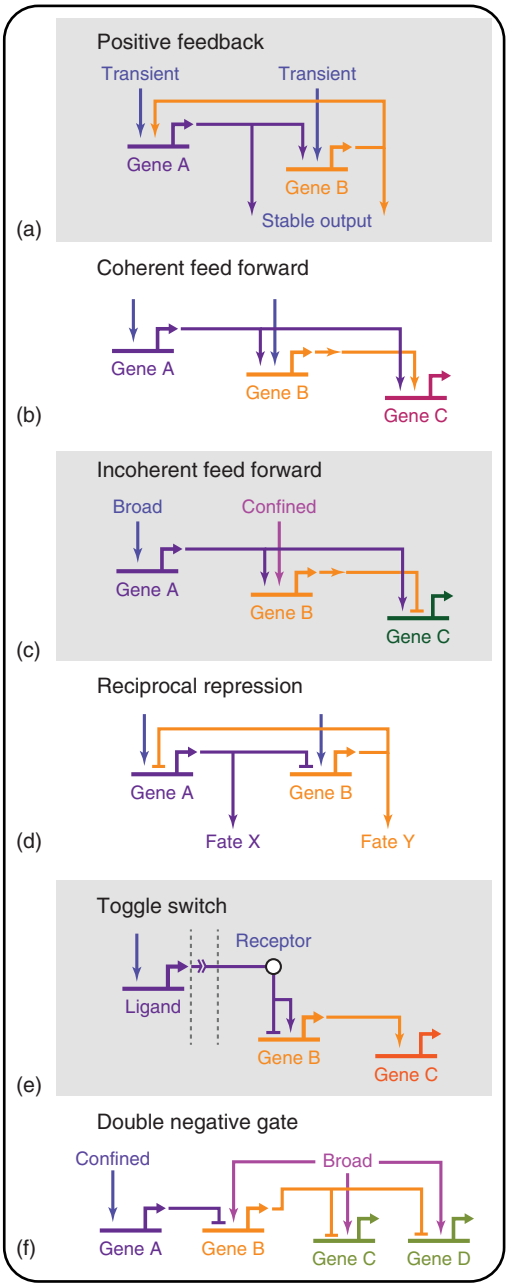

Figure 5 Recurring subcircuit architectures in developmental genetic networks. (a) Positive feedback loops stabilise regulatory states by transforming transient inputs into stable outcomes. (b) Coherent feed forward subcircuits integrate inputs. Activation of C is contingent on A and B but may be delayed. (c) Incoherent feed forward subcircuits are used for spatial subdivision as in the examples in **Figure 2**. This circuitry can also be used to cause a burst of expression of gene C, which is quickly shut off. (d) Reciprocal repression circuits prevent establishment of alternative regulatory states. These are often employed in cells descendent from a common progenitor to extinguish alternative regulatory states. (e) Toggle switch wiring ensures binary outcomes of signalling events. The immediate effectors of many signalling systems, for example Notch or Hedgehog (**Figure 3**), function as repressors in the absence of signal leading to a switch-like behaviour. (f) Double negative gate architecture is a device ensuring that genes are activated in only a particular domain but off elsewhere.

that positively regulate each other's expression (**Figure 5a**). The advantage of using more than one genes is that it allows the reuse of transcription factors in different combinations, thereby increasing the regulatory capacity. Positive feedback loops are often activated by a transient inductive signal. Once set in motion, the feedback loop locks down the new regulatory state. The amplitude of expression of the genes in the feedback loop determines the level of downstream genes and can thus serve as a buffer for fluctuations of the initial signal. A special case of positive feedback circuitry is the community effect phenomenon that is characteristic for many signalling systems as is the case for Nodal signalling as discussed earlier (**Figure 3a**).

Coherent feed forward subcircuits are circuits involving linkages between three genes, where, in its simplest form, gene A activates genes B and C, but B also activates gene C. Such topologies are often used in physiology to delay the onset of C, which would require intermittent activation of B. This pattern can also be used for spatial control (e.g. in the subcircuit from *Ciona* heart development, **Figure 4a**), where gene A defines a broad territory. If gene B receives additional inputs that activate it in a subset of A, the result is that gene C, being dependent on both, turns on specifically in the area of overlap. As C is also dependent on A, B may be used elsewhere without also activating C.

The most common form of ***incoherent food forward circuits*** are circuits where gene A activates both genes B and C, but gene B represses C. In genetic networks controlling physiological responses, this is used to cause a spike of expression in C before enough B accumulates to repress C, thereby sharply limiting its expression in time. In development, this topology is frequently used for spatial patterning to carve out a smaller domain out of a bigger one. If gene B is activated in a subset of region A through auxiliary inputs and B represses C, then gene C is expressed in A wherever B is not, for example as is the case of Emx expression in sea urchin ectoderm, which is excluded from the central domain by Nodal/Not (**Figure 3a**).

Reciprocal repression circuitry is commonly employed to exclude alternative fates. In this arrangement, two genes A and B are repressors of each other, thus ensuring they never run together in the same cell. More generally, A and B could disrupt subcircuits that are instituted as part of alternative programs and would not necessarily have to repress each other directly. This is a common feature of developmental genetic networks and beautifully demonstrated by the subcircuit controlling neural tube patterning (**Figure 3b**).

Toggle switch topologies are found in many signalling systems. Inductive signals often behave in an all or none manner. This is ensured by downstream effectors that change their behaviour in response to the signal from an obligate repressor to an activator. Shh is one such system that turns its effector protein Gli from a repressor into an activator through proteolytic cleavage.

Double negative gates are subcircuits that include two repressors that are employed together. That is, one broadly expressed repressor prohibits activation of genes that are themselves under the control of widely available activators. Wherever the repression is relieved by local activation of the second repressor, the target genes will become activated. This circuitry allows for a tighter control of the target genes than simply putting them under activating control of A. One now famous example is the control of specification of skeletogenic cells in sea urchin embryos. Here, the ubiquitous repressor HesC prevents the activation of skeletogenic genes, which are under the control of broadly expressed activators. Pmar, a second repressor, becomes expressed specifically in the skeletogenic lineage to relieve HesC repression, thereby setting in motion the genetic network underlying specification of skeletogenic cells.

Conclusion: Genetic Networks Constitute the 'Hard-wiring' of Development and Physiology in Organisms

The flow of genetic information underlies every biological phenomenon. Both cell physiology and organismic development require that cells respond to their environment by regulating their internal state and external interactions through the coordinated expression of gene products that perform specific functions. A key to understanding how genotype determines phenotype lies in the fact that this coordinated regulation is itself genetically encoded in the DNA, in the form of modular arrays of transcription factor target sites within the cis-regulatory domains of genes. The genetic network topology determined by these sites and the factors that bind them precisely constrains the response of cells to whatever environmental conditions they normally encounter. It therefore constitutes the 'hard-wiring' of cell physiology and organismal development.

Evolutionary changes must be made at the level of DNA to be passed on to the offspring. Because genetic networks determine the outcome of the developmental process, differences in phenotype are caused by changes in network architecture. Particular parts of the network may be species specific, while others are fundamental and held in common between species, or even across large phylogenetic distances. Changes in network architecture can be enacted by changing the amino acid sequence of regulatory genes, which may alter their DNA-binding or transcriptional activation capacity. But changes can also be made to the sequence of individual cis-regulatory modules causing changes in the expression of the gene they are associated with. Transcription factors usually have many targets and are used in multiple subcircuits. Thus, changes in their amino acid sequence will affect many processes, whereas changes in cis-regulatory modules are of more limited consequence. It is thus not surprising that, on an evolutionary timescale, regulatory DNA changes more rapidly than protein sequence. **See also:** **Evolution of Development**; **Evolutionary Developmental Biology: Homologous Regulatory Genes and Processes**

Beyond genetic networks, the representation of biological systems as networks of specific interactions can be extended to other levels of regulation, such as RNA–RNA interactions (e.g. involving microRNAs that block translation of the messenger ribonucleic acid (mRNA)), protein–RNA interactions and protein–protein interactions. All of these interactions depend directly or indirectly upon the primary sequence of the interacting partners, which is encoded in DNA. Knowing the architecture

of the interaction network is key to understanding the mechanism by which the genotype determines the phenotype. The quite remarkable fact of life is that the complex network of molecular interactions that generate a living organism is precisely specified in the linear DNA sequence residing in each nucleus. **See also: 5'-UTRs and Regulation**; **Interaction Networks of Proteins**

Summary

The genome sequence determines the phenotype of an organism by both specifying the amino acid sequence of proteins and regulating when and where each protein is expressed. The regulation of gene expression depends directly on the genetically specified interactions, that is those that occur between DNA-binding transcriptional regulatory proteins (plus their cofactors) and their binding sites within the cis-regulatory domains of genes. The interactions among many regulatory genes form a network that constitutes the 'hard-wiring' of physiology and development in any organism. The hierarchical organisation of genetic networks in development reflects the increase in organismal complexity and ensures that development is irreversible. The recurrent use of particular linkage patterns suggests that the specific topology and not the individual constituent genes determines the biological function. Because phenotype is the result of the execution of the genetic program, modifications of genetic network circuitry are the underlying cause of changes in organismal traits in evolution.

References

Balaskas N, Ribeiro A, Panovska J, *et al.* (2012) Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* **148**: 273–284. DOI: 10.1016/j.cell.2011.10.047.

Christiaen L, Davidson B, Kawashima T, *et al.* (2008) The transcription/migration interface in heart precursors of *Ciona intestinalis*. *Science* **320**: 1349–1352. DOI: 10.1126/science.1158170.

Davidson EH, Rast JP, Oliveri P, *et al.* (2002) A genomic regulatory network for development. *Science* **295**: 1669–1678. DOI: 10.1126/science.1069883.

Lewis EB (1978) A gene complex controlling segmentation in Drosophila. *Nature* **276**: 565–570.

Li E, Materna SC and Davidson EH (2013) New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN. *Developmental Biology* **382**: 268–279. DOI: 10.1016/j.ydbio.2013.07.027.

Li E, Cui M, Peter IS and Davidson EH (2014) Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E906–E913. DOI: 10.1073/pnas.1323105111.

Simões-Costa MS, McKeown SJ, Tan-Cabugao J, *et al.* (2012) Dynamic and differential regulation of stem cell factor FoxD3 in the neural crest is encrypted in the genome. *PLoS Genetics* **8**: e1003142. DOI: 10.1371/journal.pgen.1003142.s008.

Simões-Costa M and Bronner ME (2015) Establishing neural crest identity: a gene regulatory recipe. *Development* **142**: 242–257. DOI: 10.1242/dev.105445.

Further Reading

Alon U (2007) Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**: 450–461. DOI: 10.1038/nrg2102.

Peter I and Davidson EH (2015) *Genomic Control Process*. Oxford: Academic Press.

Ptashne M (2004) *A Genetic Switch*. New York: CSHL Press.