

Gene regulatory networks governing haematopoietic stem cell development and identity

JOHN E. PIMANDA^{*,1} and BERTHOLD GÖTTGENS^{*,2}

¹Lowy Cancer Research Centre and the Prince of Wales Clinical School, University of New South Wales, Sydney, Australia and ²Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

ABSTRACT Development can be viewed as a dynamic progression through regulatory states which characterise the various cell types within a given differentiation cascade. To understand the progression of regulatory states that define the origin and subsequent development of haematopoietic stem cells, the first imperative is to understand the ontogeny of haematopoiesis. We are fortunate that the ontogeny of blood development is one of the best characterized mammalian developmental systems. However, the field is still in its infancy with regard to the reconstruction of gene regulatory networks and their interactions with cell signalling cascades that drive a mesodermal progenitor to adopt the identity of a haematopoietic stem cell and beyond. Nevertheless, a framework to dissect these networks and comprehend the logic of its circuitry does exist and although they may not as yet be available, a sense for the tools that will be required to achieve this aim is also emerging. In this review we cover the fundamentals of network architecture, methods used to reconstruct networks, current knowledge of haematopoietic and related transcriptional networks, current challenges and future outlook.

KEY WORDS: *haematopoietic stem cell, gene transcription, regulatory network*

Introduction

Transcription factor proteins respond to biological signals and regulate the transcription rates of genes. Gene regulatory networks (GRNs) describe the interactions between transcription factor proteins and the *cis*-regulatory modules (CRMs) in their target genes, often with a particular emphasis on regulatory genes including those that express spatially important signalling proteins (Alon, 2007a, Davidson, 2006). These intra- and inter-cellular networks control the temporal and spatial expression of proteins and govern the development and function of cells and the organism they constitute.

The gene regulatory state of a cell continually defines its identity throughout development. The prior state controls the current regulatory state of a cell which in turn determines its fate. Therefore, it is not surprising that the information processing power required to maintain a cell's regulatory state is enormous. Approximate estimates of the number of regulatory functions operating in the human body at any given moment based on the number of genes that need to be regulated in a cell and the total

number of cells in the body range in the order of 10^{16} - 10^{20} (Istrail *et al.*, 2007). These interactions however are not random and large GRNs contain a small set of recurring patterns called network motifs control basic logic operations and information flow (Alon, 2007b, Milo *et al.*, 2002; Shen-Orr *et al.*, 2002). These

Abbreviations used in this paper: AGM, Aorta-Gonad-Mesonephros; AP1, Adaptor protein complex 1; ARACNe, Algorithm for the Reconstruction of Accurate Cellular Networks; ATF3, Activating Transcription Factor 3; ChIP, Chromatin Immuno Precipitation; CRMs, Cis-Regulatory Modules; CSF1R, Macrophage colony-stimulating factor 1 receptor Precursor; EST, Expressed Sequence Tags; Fli1, Friend leukemia integration 1; GATA2, GATA Binding Protein 2; GRN, Gene Regulatory Networks; Hhex, Hematopoietically-expressed homeobox protein; H3K4M3, Tri-methylation of the fourth Lysine residue of Histone H3 subunits; HSC, Haematopoietic Stem Cell; Myc, Myc Oncoprotein; Nanog, Homeobox Transcription Factor Nanog; NF-kB, Nuclear factor NF-kappa-B; Oct4, Octamer-Binding Protein 4; SCL, Stem Cell Leukaemia; Sox4, Transcription Factor Sox- 4; Mothers against decapentaplegic homolog 6, Smad6; Tcf3, Transcription Factor 3; Tcf3, Transcription Factor 3; VEGF, Vascular Endothelial Growth Factor.

*Address correspondence to: John Pimanda, Prince of Wales Clinical School, Lowy Cancer Research Centre, University of New South Wales, Sydney NSW 2052, Australia. Fax: +612-9385-1389. e-mail: jpimanda@unsw.edu.au or Bertie Göttgens, Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom. Fax: +44(0)1223-762-670. e-mail: bg200@cam.ac.uk

motifs are assembled to form larger subcircuits which determine which proteins are expressed, when, where and at what concentration to specify various body parts. The integrity of the developing organism is maintained by linking these subcircuits into a coherent GRN (Levine and Davidson, 2005; Levine and Tjian, 2003).

Development is a modular process that starts from a common template representing the undifferentiated state. This template is progressively refined as development proceeds to yield the various body parts with their distinct anatomical and functional properties (Davidson, 2006; Davidson and Erwin, 2006). GRNs that govern development are also inherently modular and hierarchical. The major structural domains and the transcriptional subcircuits that specify them are evolutionarily highly conserved. Indeed it has been proposed that conservation of the basic body plan in extant phyla in the animal kingdom is predicated by underlying conservation of GRN subcircuits or regulatory 'kernels' (Davidson and Erwin, 2006). Following domain specification and patterning, functional properties of a body part are detailed by recruiting gene differentiation batteries that appropriately vary considerably between species.

The formation of blood, like the formation of other organ systems requires prior specification of the structural domain in the developing embryo from which it develops. As detailed elsewhere in this supplement, these spatial domains are broadly conserved through evolution and the underlying regulatory kernels within large GRNs are themselves likely to be evolutionarily highly conserved. Therefore, although these networks may at first seem overwhelmingly complex, their analysis is facilitated by the modular architecture of both the organ systems and the regulatory circuitry as well as the use of a finite set of logic operations to control gene expression. In this review we cover the fundamentals of network architecture, methods used to reconstruct networks, current knowledge of haematopoietic and related transcriptional networks, current challenges and future outlook.

Fundamentals

Overview

At the cellular level, development represents a journey through successive regulatory states starting from a pluripotent stem cell and ending with fully differentiated mature cells (Waddington, 1957) (Fig. 1). Transcription factor concentrations in a cell are determined by events immediately preceding the current regulatory state and their activities determine the state that follows. The combination of transcription factors that are expressed in a cell will depend on the spatial context within the animal. Whether these factors are activated or not will in turn depend on regional cues supplied by signal transduction pathways. Transcription factors serve as vehicles to assemble activator and repressor protein complexes on DNA of target genes. If the regulatory proteins of a cell as a whole are considered an input, the output is determined by the DNA sequences of the CRMs and their accessibility to transcription factor complex binding. Therefore in terms of a

hardwired regulatory circuit, the logic of the circuit ultimately resides within these physically disparate CRMs that can be strung out over tens or even hundreds of kilobases of DNA, up or downstream of a gene promoter or within an intron (Davidson, 2006).

Logic circuits are not unique to biology and the switches, gates and quantity controllers that control genetic information flow also underlie electronic circuits (Istrail *et al.*, 2007). However, a fundamental difference between them is that although biological circuits are also hardwired in the sense that particular inputs result in predetermined outputs; biological circuits generate their own templates for successive regulatory events. Biological circuits can broadly be categorized as (i) developmental networks that underlie events during development that produce an animal from a fertilized egg and (ii) sensory networks that help maintain homeostasis within the fully formed animal.

Sensory transcription networks respond rapidly and make reversible decisions in response to environmental stimuli. Unlike signalling cascades, transcriptional networks can take up to one cell generation time to pass on a signal one step down a cascade (Alon, 2007b). As they operate on a short time-scale, long regulatory cascades are rare in signalling networks and most genes are regulated just one step away from their activator. Developmental transcription networks have different requirements and operate over longer time scales. Decisions often are largely irreversible and last long after the initial activator signal has disappeared as such developmental regulatory cascades are strung out over several cell generations (Alon, 2007b).

Cis-regulatory modules are information processing units and network anchors

A fundamental principle of developmental biology is that spatial patterning of gene expression is directly determined by the heritable *cis*-regulatory DNA sequence code (Davidson, 2006). When fragments of DNA representing *cis*-regulatory modules of genes are transferred into the genome, although they integrate randomly they still generate correct and predicted reporter ex-

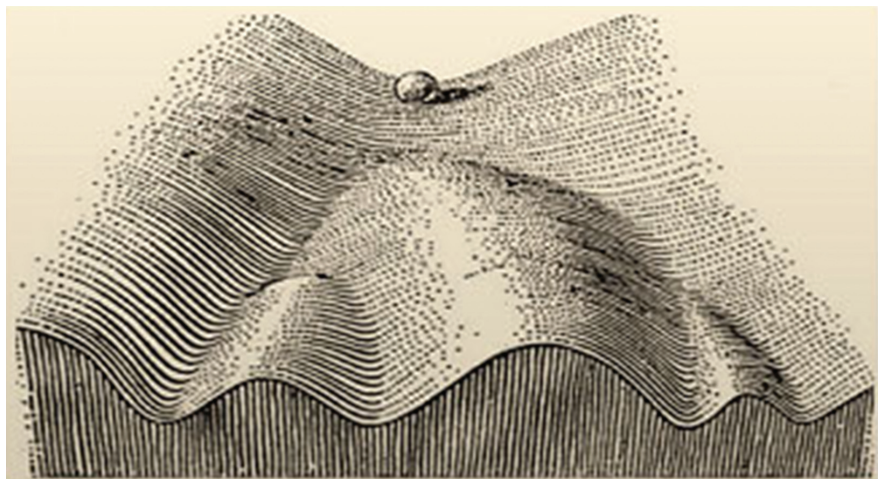


Fig. 1. Conrad Waddington's epigenetic landscape. The ball represents a cell and the bifurcating system of valleys represents trajectories of cell state. This diagram by C.H. Waddington neatly encapsulates the developmental pathways and progressive divergence of cells as they differentiate in the embryo. Reproduced from Waddington, CH © (1957) George Allen and Unwin (London).

pression that mirrors the developmental and spatial expression patterns of the endogenous genes they regulate (Delabesse *et al.*, 2005; Gottgens *et al.*, 2000; Gottgens *et al.*, 2004; Sanchez *et al.*, 1999; Silberstein *et al.*, 2005). Genes are equipped with multiple, disparate CRMs, each several hundred base pairs long. As the transcription factor binding properties of each module varies, different modules are co-opted at specific stages of development to regulate the transcriptional output of a gene. Although, CRMs function as independent information processing units their outputs are communicated elsewhere. Protein complexes assembled on distant modules are brought into contact with each other and the basal transcription apparatus on the promoter by alternative looping of DNA to regulate transcriptional activity (Davidson, 2006). How accessibility at individual modules is coordinated at

different stages of development and the precise determinants of this dynamic DNA conformation are not well understood (Fig. 2).

During specification of a body part, it may be required that a particular set or battery of genes needs to be expressed. Although the CRMs of these genes may indeed share target sites for critical transcription factors, no two CRMs of genes in a battery are exactly the same. The specific target sites in these modules vary in composition, spacing and order and respond to varying combinations of transcription factors that result in distinct gene outputs (Arnone and Davidson, 1997). These genomic regions anchor transcriptional networks by acting as information processors, which control basic logic operations and determine transcriptional output.

Information processing at CRMs can be reduced to suitable combinations of universal logic operations AND, OR and NOT (Istrail *et al.*, 2007). If a CRM has sites for two transcription factors and its output is conditional on both sites being bound, then this module operates as an AND gate. Alternatively if the output is conditional on at least one of the inputs, then it acts as an OR gate. Although one input results in some activity many OR gates are additive in that the output is the sum of both inputs. The NOT gate acts as a switch; when the input is on, the output is off and vice versa.

Network Motifs

The wiring of transcriptional networks is not random. Analyses of GRNs in bacteria and yeast have shown that a small set of regulation patterns or network motifs are used repeatedly to regulate transcriptional output (Eichenberger *et al.*, 2004; Lee *et al.*, 2002; Milo *et al.*, 2002). Where the CRMs act as information processing units that determine output of a particular gene, network motifs describe the patterns in which these individual outputs interact to regulate transcription over time.

The motifs found in biological networks are reviewed in detail elsewhere (Alon, 2007a, Alon, 2007b) and are summarized

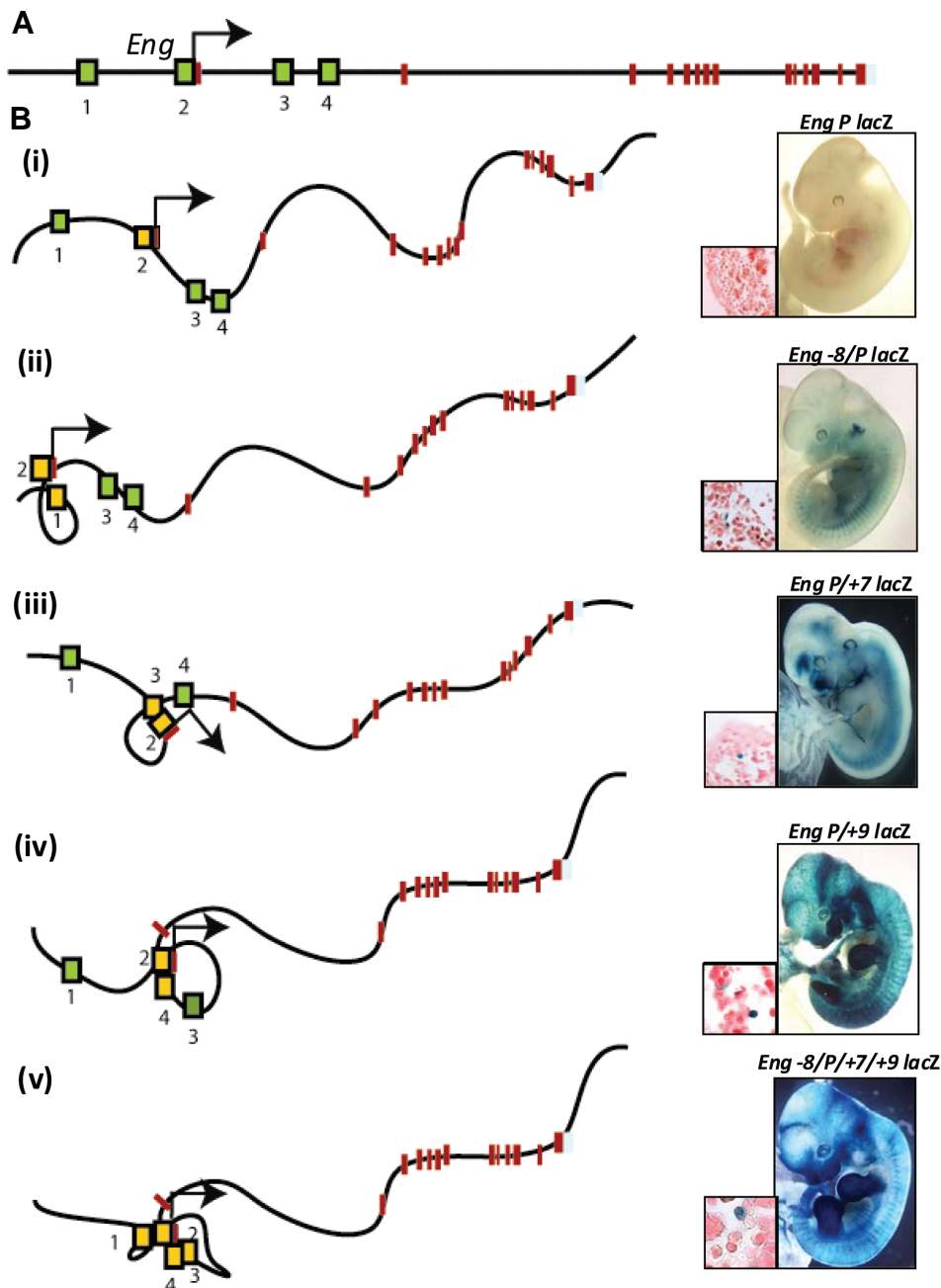


Fig. 2. Distinct enhancers are used to regulate the expression of the same gene in different tissues. (A) The endoglin gene drawn to scale with the promoter (2) and -8kb (1), +7kb (3) and +9kb (4) enhancers marked in green and exons in brown. **(B)** Transgenic mice generated with Endoglin promoter constructs (*Eng P lacZ*) show little endothelial and no blood activity. Transgenic mice generated with the -8 enhancer and endoglin promoter (*Eng -8/P lacZ*) shows robust endothelial but no blood activity. Various combinations of the *Eng P* and *Eng -8/P* constructs together with the *Eng +7* and *Eng +9* enhancers target blood and endothelium.

here. The most basic interactions (simple regulation) are those between two genes, a transcription factor and its target gene. Following modification of a transcription factor by an activator, target gene concentrations could rise until it reaches a steady state level or feed-back and regulate itself either negatively (Lee *et al.*, 2002; Rosenfeld *et al.*, 2002) or positively (Kalir *et al.*, 2005; Maeda and Sano, 2006). The kinetics of each of these responses with respect to the rate of expression of a target gene varies. Another commonly found network motif is the feed forward loop of which a number of types are described. This ‘three gene’ motif consists of a regulator X, that controls a second regulator Y, and gene Z that is controlled by both X and Y. The on-off kinetics of Z expression varies depending on the types of logic gates on the *CRMs* of both Y and Z. In a common coherent feed forward loop (Mangan and Alon, 2003), X positively regulates both Y and Z; an AND gate on Z results in slow activation and rapid inactivation of expression whereas an OR gate has the reverse response. In a common incoherent feed forward loop, X activates Z, but also

represses Z by activating repressor Y (Mangan *et al.*, 2003). This network motif utilizes both transcription factors that assemble repressor complexes on Z as well as microRNAs as repressor Y (O'Donnell *et al.*, 2005). Double positive or double negative feedback loops are commonly found in developmental networks and describe the activation of both Y and Z by X following which Y and Z continue to positively or negatively regulate each other even after X is extinguished (Gardner *et al.*, 2000; Johnston *et al.*, 2005; Milo *et al.*, 2004). This motif therefore ensures memory of X activity. It is possible to establish a series of differentiation steps by combining a series of these network motifs. Complex responses to signals can be mediated by single input modules or densely overlapping regulons which describe network motifs where either a single factor or groups of factors in varying combinations regulate sets of genes that must be activated or silenced for a particular functional response (Amit *et al.*, 2007) (Fig. 3).

Experimental validation of these network motifs has mostly been performed in bacteria but data are being gathered in eukaryotic organisms. The ultimate challenge however, is to experimentally validate the complex networks that must operate in higher order metazoans. The methodology used to discover and experimentally validate one such motif, a variant of the fully connected triad that operates during blood stem cell development in the mouse embryo is detailed later in this review. In a fully connected triad, transcription factors X, Y and Z all positively regulate each other upon activation of one or more of its constituents a mechanism that maybe used more widely during development (Milo *et al.*, 2002; Pimanda *et al.*, 2007b).

Cellular Memory

As cells share the same genome and the gene regulatory state of a cell determines its identity and function, installing restricted chromatin states that play a role in expression or silencing of genes is essential for cells to adopt a more restricted phenotype. Gene expression patterns that are initially mandated by *CRMs* are retained long after the initial stimulus has been extinguished by locking in specific chromatin states in a cell. This is achieved by a variety of chromatin modifications including DNA methylation, histone acetylation and deacetylation, differential methylation of lysine residues of histones which install polycomb or trithorax group complexes and control accessibility of the transcription apparatus to promoters and distant *CRMs* (Allis *et al.*, 2006; Davidson, 2006). Gene expression states can also be maintained by continuously functional sub-circuits that direct expression of down-stream genes long after the initial trigger has ceased. Alternatively, transcriptionally driven memory subcircuits can be maintained by the community effect where all cells of a territory signal to one another and these inter-cellular feedback circuits maintain the regulatory state of cells within that territory (Davidson, 2006; Gurdon, 1988).

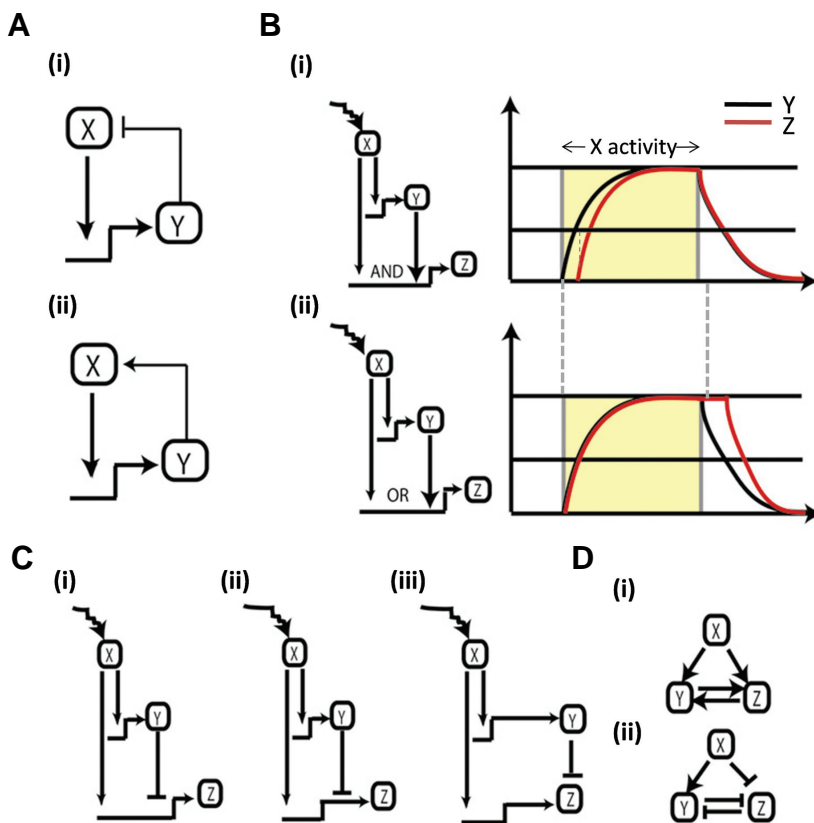


Fig. 3. Different network motifs can be used to regulate the expression kinetics of developmentally important genes. (A) (i) Simple negative regulation; the activation of X, up-regulates Y, which down regulates the activator (ii) Simple positive regulation; the activation of X, up-regulates Y, which further increases X. (B) Coherent feed-forward loops with different logic gates determine rates of target gene expression (i) shows an AND gate with delayed onset and rapid loss (ii) shows an OR gate with rapid onset and delayed loss of target gene expression. (C) Incoherent feed-forward loops, where the activator X up-regulates Z expression but also inhibits Z by up-regulating Y. These incoherent feed-forward loops act at different levels (i) transcription (ii) post-transcription (iii) post-translational. (D) Double positive and negative motifs. (i) Y and Z can continue to up-regulate each other after the activator X has been extinguished. (ii) The inhibition of Z can be initiated by the activation of X and maintained by Y.

Chromatin remodelling and activation of *CRMs* can be temporally linked to the expression of transcription factors that subsequently bind these modules (Bonifer and Hume, 2008). Occupancy of CRMs by pioneer transcription factors also helps maintain their ongoing accessibility and the adoption of inactive chromatin marks can follow the loss of expression of these factors (Tagoh *et al.*, 2002; Xu *et al.*, 2007). Chromatin condensation and inactivation of gene expression can also be achieved by tissue specific transcription factors recruiting histone bound co-repressors to gene promoters (Xu *et al.*, 2007).

The *Csf1r* surface protein is required for macrophage development and the *csf1r* gene is silenced in the lymphoid lineage. However, whilst the gene is truly epigenetically silenced by inaccessibility of the transcription start site and promoter DNA methylation in T- cells, the promoter and intronic regulatory elements are maintained in a partially active chromatin conformation in B- cells. This has been shown to be due to the continuous expression in B- cells of Pu.1, an *Ets* transcription factor that binds to these regions of the *Csf1r* locus. *Csf1r* is not however expressed in B- cells owing to the active suppression of the promoter by Pax5, which is required to maintain B- cell identity (Tagoh *et al.*, 2006). Therefore chromatin remodelling is a dynamic process and intermediate states can exist with a limited capacity for cells to transit between types.

Robustness

Establishing one developmental fate usually prohibits an alternate fate. This is important to maintain structural integrity in a given spatial domain. This is achieved by establishing multiple inter-cellular and intra-cellular regulatory feedback circuits that actively maintain a given state and repress alternate possibilities. A robust design would also safeguard against failure. Regulatory sub-circuits that specify various body parts have been selected in evolution to provide maximum insurance against the possibility of developmental failure (Cripps and Olson, 2002; Hinman *et al.*, 2003; Pimanda *et al.*, 2007b). Furthermore, the subcircuit elements of any developmental system may be of different ages and origins in evolution, selected and assembled piece-meal to best suit circumstance (Davidson and Erwin, 2006; Tokuoka *et al.*, 2004). Robustness is also achieved by a degree of built in redundancy. *CRMs*, which constitute the nodes of a network often have multiple binding sites for transcription factors and utilize multiple related factors to drive gene expression.

Network reconstruction

Cell type specific transcription factor networks determine cellular response to external signals. Current strategies to reverse engineer such networks fall into two categories. The 'bottom-up' approach uses tissue specific *CRMs* of key regulators of a given tissue as anchors to build block by block the regulatory interactions that operate in that tissue. The 'top-down' approach on the other hand attempts to reconstruct networks from genome-wide expression profiles of a particular tissue and predicted interactions between these expressed genes.

The 'bottom-up' approach

the most advanced haematopoietic stem cell (HSC) network identified in this way took the *Scl/Tal1* gene as the starting point. *Scl/Tal1* is required to specify blood development from mesoder-

mal progenitors. Disruption of this gene during development results in a bloodless embryo and death. A 600 bp enhancer +19kb downstream of the transcription start site was found to be sufficient to target reporter gene expression to HSCs and endothelium in transgenic embryos (Gottgens *et al.*, 2002; Rainis *et al.*, 2005; Sanchez *et al.*, 1999; Sanchez *et al.*, 2001; Silberstein *et al.*, 2005; Sinclair *et al.*, 1999). *In-vivo* activity was shown to be dependent on two conserved Ets-binding sites and one conserved GATA- binding site which were bound by Fli1, Elf1, Erg and Gata2 in blood progenitors. The *Scl* +19 enhancer was therefore the first *in vivo* defined HSC *cis*-regulatory module of any gene required to specify HSCs. Using the Ets/Ets/Gata configuration of the *Scl*+19 enhancer as a template, a subsequent computational screen (Donaldson *et al.*, 2005a) identified 67 clusters that were conserved between the human, mouse and dog genomes with the same spacing and orientation constraints as *Scl*+19. Three of these computationally identified clusters were situated in the *Fli1*, *Hhex* and *Smad6* gene loci and showed HSC activity similar to the parent *Scl*+19 enhancer (Donaldson *et al.*, 2005b, Pimanda *et al.*, 2007a). Using bioinformatics, ChIP-seq and transgenic analyses, additional regulators of haematopoiesis have been progressively added to this nascent gene regulatory network (Chan *et al.*, 2007; Chapman *et al.*, 2003; Landry *et al.*, 2008; Okuno *et al.*, 2005; Wilson *et al.*, 2009).

Therefore reverse engineering networks using a bottom-up approach involves a number of steps. The most important of which is the identification of tissue specific *CRMs* of genes required to specify that tissue. This is followed by the identification of transcription factor binding motifs within these modules that are important for their activity and the use of ChIP (Chromatin Immunoprecipitation) based techniques to identify transcription factors that are expressed in that tissue and are bound to these modules. From this point, one can either work up the developmental hierarchy by identifying tissue specific regulatory modules of transcription factors that are the drivers of the parent CRM or use bio-informatics tools to predict other target CRMs that respond in parallel with the parent CRM. When working up the developmental hierarchy, knowledge of the expression time course of various transcription factors during development of a particular tissue is extremely useful (Patterson *et al.*, 2007; Patterson *et al.*, 2005; Swiers *et al.*, 2006). ChIP-chip analyses in appropriate cells can also be invaluable in identifying potential tissue specific CRMs in regulatory genes that can then be verified as truly targeting HSCs using transgenic assays.

The 'top-down' approach

this approach, which has been used successfully to determine regulatory networks in bacteria and yeasts, has recently been adopted to reverse engineer networks in B lymphocytes and macrophages (Basso *et al.*, 2005; Gilchrist *et al.*, 2006). Broadly, these approaches rely on genome-wide expression data sets in cells following various stimuli or experimental conditions as a starting point and use algorithms to infer gene-gene co-regulation. These methods aim to generate graphical representations of cellular networks where nodes represent genes and edges between them represent interactions, either between the encoded proteins or between encoded proteins and genes (Hartemink *et al.*, 2001). Top down approaches have been variably successful in simple organisms and fall into the following broad categories:

optimization methods, which maximize a scoring function over alternative network models (Gat-Viks and Shamir, 2003; Gat-Viks *et al.*, 2003); regression techniques, which fit the data to *a priori* models (Gardner *et al.*, 2003); integrative bioinformatics approaches, which combine data from a number of independent experimental clues (Ideker *et al.*, 2001); and statistical methods, which rely on a variety of measures of pair wise gene expression correlation (Butte and Kohane, 2000).

An early success in reconstructing mammalian regulatory networks using a 'top-down' approach was in human B- cells. Using a new method called ARACNe (algorithm for the reconstruction of accurate cellular networks), the authors show that a relatively small number of highly connected genes interact with most other genes in the cell, either directly or hierarchically, through highly connected sub-hubs (Margolin *et al.*, 2006). The proto-oncogene *MYC* was identified as one of the largest hubs in the B- cell network (Basso *et al.*, 2005). A top-down approach was also used to show that activating transcription factor 3 (ATF3) negatively regulates the Toll-like receptor transcriptional program in macrophages (Gilchrist *et al.*, 2006). The authors analyzed waves of gene transcription following stimulation of macrophages with a Toll-like receptor 4 agonist and clustered genes based on their temporal patterns of expression. They hypothesized that a cluster of genes that included ATF3 (cluster 6) whose mRNA peaked at 1hr were co-regulated and share *CRMs* that were responsive to one or more members of this cluster. ATF binding sites were noticed to be over represented in these modules. Using Cytoscape, a network analysis and visualization tool to predict protein-protein interactions, ATF3 was predicted to act in concert with NF- κ B and AP1. By scanning DNA sequences for ATF3/ NF- κ B and AP1 binding sites close to the transcription start sites, *IL-6* and *IL12b* were identified as target genes that are silenced following ATF3 binding.

There are approximately 200 transcription factor genes expressed in neural stem cells and possibly a similar number in HSCs (Akashi *et al.*, 2003; Kummerfeld and Teichmann, 2006). To computationally predict and experimentally map every transcription factor- DNA interaction is a gargantuan task. Computational predictions are limited by the fact that DNA binding specificities of only a small fraction of transcription factors are sufficiently characterized to predict the sequences that they can and

cannot bind although attempts are underway to redress this (Berger *et al.*, 2006). The large data sets that inevitably will be generated can be filtered by focusing on phylogenetically conserved binding sites although phylogenetic footprinting will not identify sites that are functional but not conserved (Odom *et al.*, 2007). The data sets can be further filtered by incorporating large scale experimental verification of chromatin accessibility and transcription factor binding events using ChIP- chip (Boyer *et al.*, 2005) or ChIP-sequencing technologies (Marson *et al.*, 2008). A major drawback however is the number of cells that are required to perform these genome-scale experiments, bearing in mind that the E11.5 mouse AGM (Aorta- Gonad- Mesonephros) is estimated to harbour approximately one HSC per embryo. Another limitation is the precise identification of tissue specific stem cells. The sensitivities of current flow cytometry protocols used to identify HSCs are approximately 50% (Weksberg *et al.*, 2008) for mouse and 1% for human (Majeti *et al.*, 2007).

Regulatory networks in haematopoiesis

To determine the regulatory state of a HSC, it would help to first know the identity of the cell or cells from which they arise and the site or sites from which they originate and not just reside. That the precursor cell is a derivative of the mesoderm and has a close association with the endothelial lineage is recognized and is detailed elsewhere in this supplement (Fig. 4). That the AGM is a site of HSC generation is not controversial and the recent demonstration that long-term repopulating HSCs in the AGM originate from VE-cadherin⁺CD45⁺ precursors is helpful in this regard (Taoudi *et al.*, 2008). If the properties of HSCs are uniform, from a purely systems biology point of view, one could argue that wherever or from whatever precursor they arise, the chromatin state and transcription factors that are expressed (with allowance for factor redundancy) within these precursors are similar. However, molecular and phenotypic analysis has clearly shown intrinsic differences between HSC isolated from mouse embryos, juvenile, adult and aged mice (Dykstra *et al.*, 2007; Kim *et al.*, 2006; Sieburg *et al.*, 2006; Yilmaz *et al.*, 2006). Nevertheless, despite these differences, all these HSC populations share the hallmark properties of long-term self-renewal and multilineage differentiation capacity. Hence it is likely that many key aspects of

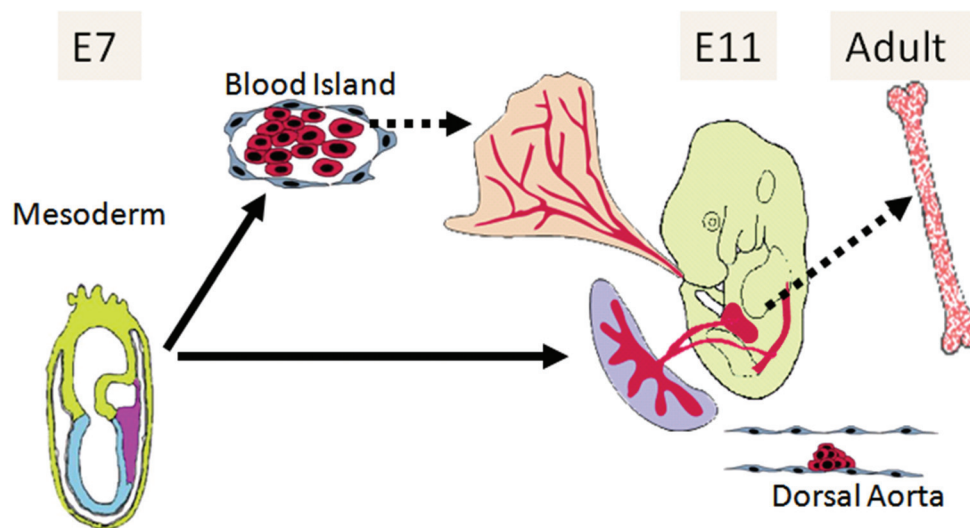


Fig. 4. The ontogeny of blood development in a mouse model. Mesodermal cells from the posterior primitive streak (purple) of day 7 embryos establish extra-embryonic yolk sac endothelial plexus and blood. These cells circulate in the embryo after the establishment of cardiac contractions at day 8. Long-term repopulating blood stem cells originate from and around the dorsal aorta in the AGM region around day 10. These stem cells number is amplified in the fetal liver and placenta and populate the bone marrow around the time of birth. The bone marrow continues as the major haematopoietic organ in adult life.

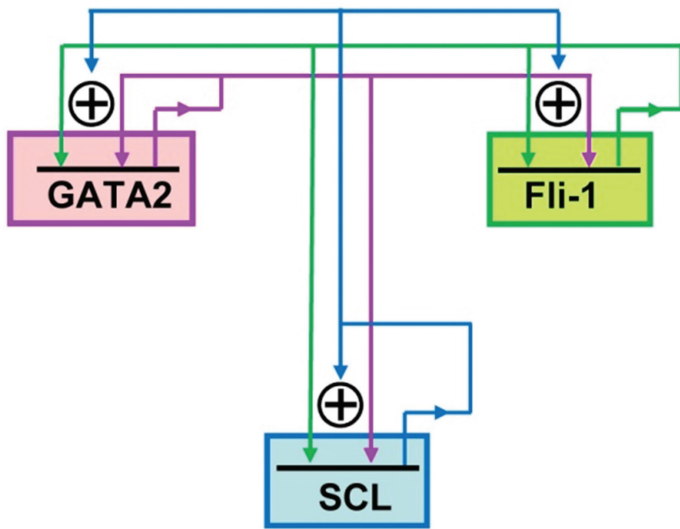


Fig. 5. A fully connected triad of HSC transcription factors. *Scl*, *Fli1* and *Gata2* regulate each other by collectively binding the *Scl*+19, *Fli1*+12 and *Gata2*-3, haematopoietic stem cell enhancers. This triad is probably activated by *Bmp4* signalling and the expression of these essential haematopoietic factors in HSCs can be maintained by each other in the absence of the activator.

GRN architecture will be conserved among these various HSC populations.

We have reported that *Gata2*, *Fli1* and *Scl* transcription factors, all of which are expressed in HSCs, cross-regulate each other in the AGM by binding *CRMs* that are active in the AGM and in LT-HSCs (Pimanda *et al.*, 2007b) (Fig. 5). These transcription factors are required for normal haematopoiesis and are down-stream of *Bmp4*, a morphogen that is required to specify the mesoderm and to specify haematopoiesis within the mesoderm (Maeno *et al.*, 1996). This triad of transcription factors are components of a larger network in HSCs that we have constructed using a bottom-up approach as detailed above. The circuitry between *Gata2*, *Fli1*

and *Scl* and their *CRMs* conforms to a three gene network motif termed a fully connected triad (Pimanda *et al.*, 2007b). This 3-way positive feed-forward loop, which shows a tendency to be locked into an ON state, is rare in the experimentally verified GRNs of bacteria and yeasts. An important feature of the fully connected triad is that following its activation, the members will perpetuate the expression of each other. Therefore the activating signal could be transient but sufficient to initiate a key transcriptional programme which will be retained by an ongoing dynamic interaction (see preceding section on 'memory'). This has important implications for cells such as HSCs which emerge at one site (the AGM) and then travel to another (the fetal liver) for further amplification prior to taking up residency in the bone marrow.

The *Gata2*/*Fli1*/*Scl* triad is also intriguing as it provides a stepping stone to retrace transcriptional events that underlie HSC development (Fig. 6). The development of HSCs in the AGM is regulated by Notch-1 and is impaired in Notch-1 deficient mouse embryos (Kumano *et al.*, 2003). Notch-1 and *Gata2* are co-expressed in the haemogenic endothelium in the floor of the aorta at E10.5 and Notch1 binds the *Gata2* promoter and acts as an upstream regulator of *Gata2* expression (Robert-Moreno *et al.*, 2005). Notch 1 itself is downstream of hedgehog and Vegf signalling (Gering and Patient, 2005; Lawson *et al.*, 2002). However, because activation of the regulatory elements depends on "and" interactions between *Gata2* and *Fli1* (*Scl* likely plays a quantitative role once the *Gata2*/*Fli1* complex is bound), activating *Gata2* alone is unlikely to be sufficient to initiate the triad. *Bmp4* signalling is required for the formation of the dorsal lateral plate mesoderm that gives rise to the dorsal aorta and the subsequent development of blood. Disruption of *Bmp4* results in loss of *Fli1* and *Scl* expression (Walmsley *et al.*, 2002). The *Gata2* promoter responds to *Bmp4* which also initiates *Fli1* expression (Oren *et al.*, 2005). Activation of both *Gata2* and *Fli1* might therefore be sufficient to trigger the circuit that could then maintain expression of each other and *Scl*. As important as maintaining activity of the triad in HSCs is the abrogation of the triad to diminish concentrations of these factors as cells differentiate down

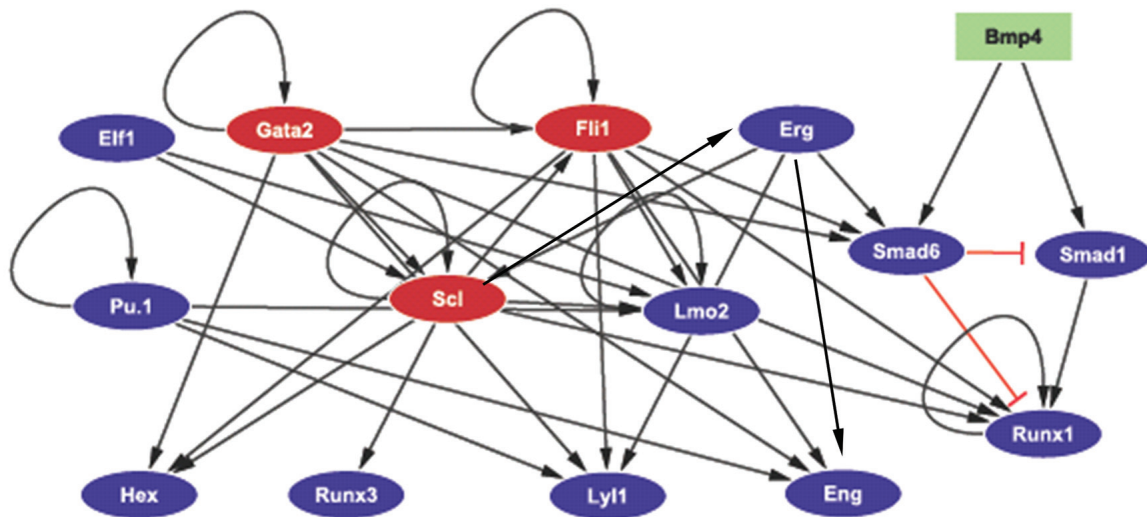


Fig. 6. A bottom-up reconstruction of the haematopoietic stem cell regulatory network. The connections between each of these genes have been experimentally verified.

specific lineages. Interestingly, Gata1 has been shown to disrupt Gata2-positive auto-regulation by binding the Gata2-3 enhancer as cells differentiate down the erythroid-megakaryocytic lineages (Grass *et al.*, 2003). Given the critical “and” interaction between Gata2 and Fli1, shutting down Gata2 may be sufficient to abrogate the triad and one would predict that maintenance of Scl expression in maturing erythroid cells would require an alternative erythroid enhancer. Interestingly, such an enhancer has indeed been identified and shown to depend on Gata1 and Scl as upstream inputs (Ogilvy *et al.*, 2007).

Reconstruction of core transcriptional regulatory circuitry would not be complete without integrating regulatory RNAs. miRNAs appear to regulate expression of a significant proportion of all genes in a variety of mammalian cell types including haematopoietic cells (Garzon and Croce, 2008; Lewis *et al.*, 2005; Stefani and Slack, 2008). The expression of miRNAs is not random. These genes themselves participate in cellular differentiation and are subject to tissue specific regulation (Dore *et al.*, 2008). Incoherent feed-forward loops where X activates Z, but also represses Z by activating a repressor Y is a classic network motif used to generate pulse-like dynamics (Alon, 2007b, Basu *et al.*, 2004). As detailed above, regulatory miRNA genes are ideally placed to act as ‘Y’ type repressors. Indeed by correlating genome-wide tri-methylation of the fourth Lysine residue of Histone H3 subunits (H3K4M3 chromatin marks) with known transcription start sites of primary and mature miRNA transcripts and ESTs (and not promoters of protein coding genes) a directory of transcription start sites of miRNA genes that are expressed in ES cells has been constructed and demonstrate overlapping occupancy of Oct4, Sox2, Nanog and Tcf3 on miRNA promoters (Marson *et al.*, 2008). As promoter occupancy with Oct4, Sox2, Nanog and expression of protein coding genes in ES cells has already been mapped and because experimentally verified targets of certain miRNAs are known, it has been possible to wire miRNA modulation into the gene regulatory network in ES cells.

Regulatory networks and haematopoietic disorders

Diagnosis and classification of haematopoietic malignancies is largely based on morphology and immunophenotype. Although gene expression profiling of haematopoietic malignancies to risk stratify patients based on molecular signatures has been successful (Golub *et al.*, 1999), it is unlikely to enter routine diagnostic and treatment paradigms unless less toxic and better treatment alternatives are available. If targeted therapy is the aim, differential gene expression profiles by themselves are not very informative. If the expression profiles can be organized into a hierarchical interactome that potentially guides the development of targeted therapy, the benefits of expression profiling would be more tangible. Comprehensive analysis of tissue regulatory networks provides a framework to understand perturbations that lead to various disease phenotypes. It also provides a sense of the most important regulatory genes. For example, MYC is one of the most connected hubs in the B- cell interactome (Li *et al.*, 2003) and is the key oncogenic lesion in Burkitt's lymphoma and a potential target for cancer therapy (Soucek *et al.*, 2008). Clearly, if a recurrent cytogenetic

lesion can be identified in a tumour either by conventional karyotyping or by comparative genomic hybridization arrays and the involved gene is either highly connected or is a critical mediator of an important hub within the normal interactome, this would be a target for therapy. Expression profiling has also been used to characterize the molecular signature arising from specific pharmacological interventions (Lamb *et al.*, 2006). Again, it would be far more informative to characterize the networks that are perturbed in response to specific pharmacological interventions for it could guide therapy. By expression profiling a tumour and identifying the key networks that are deranged, specific interventions could be planned based on prior knowledge.

Network analysis of haematopoietic malignancies however will only be as good as the original template. The most informative blood cell template to date, the B- cell interactome, was constructed by integrating gene expression profiles from several stages of B- cell development with known protein-protein and protein-DNA interactions and known modulators of these interactions compiled from various databases (Mani *et al.*, 2008). One drawback of this methodology is the large background population that is necessary for comparison. As dependency metrics like mutual information require a certain sample size to establish significance, this may pose a difficulty in situations where sample sizes are limited (Mani *et al.*, 2008).

Challenges and future directions

By their very nature, HSCs are suspended in transit towards another more differentiated state. HSC purity and adequate cell numbers for *in vitro* analyses remain major issues for network reconstruction. *In vitro* or *in vivo* culture and expansion of HSCs without altering their phenotype would be necessary unless fundamental technical advances in methodology permit gene expression and ChIP on very small numbers of cells. One possible substitute is a robust ES cell differentiation system in which LT-HSCs can be effectively identified.

To reconstruct networks, effective methods to identify protein-DNA interactions are required. The reliance on specific antibodies for ChIP is a hurdle. Methods to express tagged transcription factors at appropriate concentrations in ES cells to reconstruct the ES cell interactome (Kim *et al.*, 2008; Wang *et al.*, 2006) have been successful and could be adapted to HSCs derived from ES cells if the necessary protocols are developed. A limitation that is likely to be overcome in the near future is the lack of a comprehensive database of binding specificities of all the 2,500 or so known transcription factors. Protein binding microarray technology is rapidly increasing the number of known binding specificities (Berger *et al.*, 2008; Berger *et al.*, 2006). High throughput sequencing is also likely to have a major impact on our ability to document genome-wide transcription factor binding events using ChIP-seq technology. Transgenic evaluation of *CRMs* is currently regarded as necessary to determine potential activity in LT-HSCs. This is expensive, labour intensive and time consuming. The evaluation of *CRMs* in ES cell differentiation assays could be an acceptable alternative. As a clearer map of cell signalling cascades that drive developmental processes evolves, a future challenge will be to integrate haematopoietic transcriptional networks with

these cascades that activate the regulatory programme.

Acknowledgements

Work in the authors' laboratories is supported by grants from the National Health and Medical Research Council of Australia; the Leukaemia Foundation, Australia; the Australian Research Council; the Viertel Foundation; the Ramaciotti Foundation; the Leukaemia Research Fund, UK; Cancer Research UK; the Leukemia and Lymphoma Society, the Newton Trust and the Kay Kendall Leukaemia Fund.

References

- AKASHI, K., HE, X., CHEN, J., IWASAKI, H., NIU, C., STEENHARD, B., ZHANG, J., HAUG, J. and LI, L. (2003). Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* 101: 383-389.
- ALLIS, C.D., JENUWEIN, T. and REINBERG, D. (2006). *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- ALON, U. (2007a). *An introduction to systems biology: design principles of biological circuits*. Chapman & Hall/CRC, Boca Raton, Fla.; London.
- ALON, U. (2007b). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8: 450-461.
- AMIT, I., CITRI, A., SHAY, T., LU, Y., KATZ, M., ZHANG, F., TARCIC, G., SIWAK, D., LAHAD, J., JACOB-HIRSCH, J. *et al.* (2007). A module of negative feedback regulators defines growth factor signaling. *Nat Genet* 39: 503-512.
- ARNONE, M.I. and DAVIDSON, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124: 1851-1864.
- BASSO, K., MARGOLIN, A.A., STOLOVITZKY, G., KLEIN, U., DALLA-FAVERA, R. and CALIFANO, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat Genet* 37: 382-390.
- BASU, S., MEHREJA, R., THIBERGE, S., CHEN, M.T. and WEISS, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA* 101: 6355-6360.
- BERGER, M.F., BADIS, G., GEHRKE, A.R., TALUKDER, S., PHILIPPAKIS, A.A., PENA-CASTILLO, L., ALLEYNE, T.M., MNAIMNEH, S., BOTVINNIK, O.B., CHAN, E.T. *et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133: 1266-1276.
- BERGER, M.F., PHILIPPAKIS, A.A., QURESHI, A.M., HE, F.S., ESTEP, P.W., 3RD and BULYK, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* 24: 1429-1435.
- BONIFER, C. and HUME, D.A. (2008). The transcriptional regulation of the Colony-Stimulating Factor 1 Receptor (*csf1r*) gene during hematopoiesis. *Front Biosci* 13: 549-650.
- BOYER, L.A., LEE, T.I., COLE, M.F., JOHNSTONE, S.E., LEVINE, S.S., ZUCKER, J.P., GUENTHER, M.G., KUMAR, R.M., MURRAY, H.L., JENNER, R.G. *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947-956.
- BUTTE, A.J. and KOHANE, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* 418-429.
- CHAN, W.Y., FOLLOWS, G.A., LACAUD, G., PIMANDA, J.E., LANDRY, J.R., KINSTON, S., KNEZEVIC, K., PILTZ, S., DONALDSON, I.J., GAMBARELLA, L. *et al.* (2007). The paralogous hematopoietic regulators *Lyl1* and *Scf* are coregulated by Ets and GATA factors, but *Lyl1* cannot rescue the early *Scf*-phenotype. *Blood* 109: 1908-1916.
- CHAPMAN, M.A., CHARCHAR, F.J., KINSTON, S., BIRD, C.P., GRAFHAM, D., ROGERS, J., GRUTZNER, F., GRAVES, J.A., GREEN, A.R. and GOTTGENS, B. (2003). Comparative and functional analyses of *LYL1* loci establish marsupial sequences as a model for phylogenetic footprinting. *Genomics* 81: 249-259.
- CRIPPS, R.M. and OLSON, E.N. (2002). Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev Biol* 246: 14-28.
- DAVIDSON, E.H. (2006). *The regulatory genome: gene regulatory networks in development and evolution*. Elsevier/Academic Press, Amsterdam; London.
- DAVIDSON, E.H. and ERWIN, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311: 796-800.
- DELABESSE, E., OGILVY, S., CHAPMAN, M.A., PILTZ, S.G., GOTTGENS, B. and GREEN, A.R. (2005). Transcriptional regulation of the SCL locus: identification of an enhancer that targets the primitive erythroid lineage *in vivo*. *Mol Cell Biol* 25: 5215-5225.
- DONALDSON, I.J., CHAPMAN, M. and GOTTGENS, B. (2005a). TFBScluster: a resource for the characterization of transcriptional regulatory networks. *Bioinformatics* 21: 3058-3059.
- DONALDSON, I.J., CHAPMAN, M., KINSTON, S., LANDRY, J.R., KNEZEVIC, K., PILTZ, S., BUCKLEY, N., GREEN, A.R. and GOTTGENS, B. (2005b). Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. *Hum Mol Genet* 14: 595-601.
- DORE, L.C., AMIGO, J.D., DOS SANTOS, C.O., ZHANG, Z., GAI, X., TOBIAS, J.W., YU, D., KLEIN, A.M., DORMAN, C., WU, W. *et al.* (2008). A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci USA* 105: 3333-3338.
- DYKSTRA, B., KENT, D., BOWIE, M., MCCAFFREY, L., HAMILTON, M., LYONS, K., LEE, S.J., BRINKMAN, R. and EAVES, C. (2007). Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* 1: 218-229.
- EICHENBERGER, P., FUJITA, M., JENSEN, S.T., CONLON, E.M., RUDNER, D.Z., WANG, S.T., FERGUSON, C., HAGA, K., SATO, T., LIU, J.S. *et al.* (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol* 2: e328.
- GARDNER, T.S., CANTOR, C.R. and COLLINS, J.J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339-342.
- GARDNER, T.S., DI BERNARDO, D., LORENZ, D. and COLLINS, J.J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301: 102-105.
- GARZON, R. and CROCE, C.M. (2008). MicroRNAs in normal and malignant hematopoiesis. *Curr Opin Hematol* 15: 352-358.
- GAT-VIKS, I. and SHAMIR, R. (2003). Chain functions and scoring functions in genetic networks. *Bioinformatics* 19 Suppl 1: i108-117.
- GAT-VIKS, I., SHARAN, R. and SHAMIR, R. (2003). Scoring clustering solutions by their biological relevance. *Bioinformatics* 19: 2381-2389.
- GERING, M. and PATIENT, R. (2005). Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev Cell* 8: 389-400.
- GILCHRIST, M., THORSSON, V., LI, B., RUST, A.G., KORB, M., ROACH, J.C., KENNEDY, K., HAI, T., BOLOURI, H. and ADEREM, A. (2006). Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* 441: 173-178.
- GOLUB, T.R., SLONIM, D.K., TAMAYO, P., HUARD, C., GAASENBEEK, M., MESIROV, J.P., COLLIER, H., LOH, M.L., DOWNING, J.R., CALIGIURI, M.A. *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537.
- GOTTGENS, B., BARTON, L.M., GILBERT, J.G., BENCH, A.J., SANCHEZ, M.J., BAHN, S., MISTRY, S., GRAFHAM, D., MCMURRAY, A., VAUDIN, M. *et al.* (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. *Nat Biotechnol* 18: 181-186.
- GOTTGENS, B., BROCCARDO, C., SANCHEZ, M.J., DEVEAUX, S., MURPHY, G., GOTHERT, J.R., KOTSPOULOU, E., KINSTON, S., DELANEY, L., PILTZ, S. *et al.* (2004). The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Eif-1. *Mol Cell Biol* 24: 1870-1883.
- GOTTGENS, B., NASTOS, A., KINSTON, S., PILTZ, S., DELABESSE, E.C., STANLEY, M., SANCHEZ, M.J., CIAU-UITZ, A., PATIENT, R. and GREEN, A.R. (2002). Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J* 21: 3039-3050.
- GRASS, J.A., BOYER, M.E., PAL, S., WU, J., WEISS, M.J. and BRESNICK, E.H. (2003). GATA-1-dependent transcriptional repression of *GATA-2* via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci USA* 100: 8811-8816.
- GURDON, J.B. (1988). A community effect in animal development. *Nature* 336: 772-774.

- HARTEMINK, A.J., GIFFORD, D.K., JAAKKOLA, T.S. and YOUNG, R.A. (2001). Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac Symp Biocomput* 422-433.
- HINMAN, V.F., NGUYEN, A.T., CAMERON, R.A. and DAVIDSON, E.H. (2003). Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci USA* 100: 13356-13361.
- IDEKER, T., THORSSON, V., RANISH, J.A., CHRISTMAS, R., BUHLER, J., ENG, J.K., BUMGARNER, R., GOODLETT, D.R., AEBERSOLD, R. and HOOD, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292: 929-934.
- ISTRAIL, S., DE-LEON, S.B. and DAVIDSON, E.H. (2007). The regulatory genome and the computer. *Dev Biol* 310: 187-195.
- JOHNSTON, R.J., JR., CHANG, S., ETCHBERGER, J.F., ORTIZ, C.O. and HOBERT, O. (2005). MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci USA* 102: 12449-12454.
- KALIR, S., MANGAN, S. and ALON, U. (2005). A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol Syst Biol* 1: 2005 0006.
- KIM, I., HE, S., YILMAZ, O.H., KIEL, M.J. and MORRISON, S.J. (2006). Enhanced purification of fetal liver hematopoietic stem cells using SLAM family receptors. *Blood* 108: 737-744.
- KIM, J., CHU, J., SHEN, X., WANG, J. and ORKIN, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132: 1049-1061.
- KUMANO, K., CHIBA, S., KUNISATO, A., SATA, M., SAITO, T., NAKAGAMI-YAMAGUCHI, E., YAMAGUCHI, T., MASUDA, S., SHIMIZU, K., TAKAHASHI, T. *et al.* (2003). Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. *Immunity* 18: 699-711.
- KUMMERFELD, S.K. and TEICHMANN, S.A. (2006). DBD: a transcription factor prediction database. *Nucleic Acids Res* 34: D74-81.
- LAMB, J., CRAWFORD, E.D., PECK, D., MODELL, J.W., BLAT, I.C., WROBEL, M.J., LERNER, J., BRUNET, J.P., SUBRAMANIAN, A., ROSS, K.N. *et al.* (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929-1935.
- LANDRY, J.R., KINSTON, S., KNEZEVIC, K., DE BRUIJN, M.F., WILSON, N., NOTTINGHAM, W.T., PEITZ, M., EDENHOFER, F., PIMANDA, J.E., OTTERSACH, K. *et al.* (2008). Runx genes are direct targets of Scf/Tal1 in the yolk sac and fetal liver. *Blood* 111: 3005-3014.
- LAWSON, N.D., VOGEL, A.M. and WEINSTEIN, B.M. (2002). sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. *Dev Cell* 3: 127-136.
- LEE, T.I., RINALDI, N.J., ROBERT, G., ODOM, D.T., BAR-JOSEPH, Z., GERBER, G.K., HANNETT, N.M., HARBISON, C.T., THOMPSON, C.M., SIMON, I. *et al.* (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298: 799-804.
- LEVINE, M. and DAVIDSON, E.H. (2005). Gene regulatory networks for development. *Proc Natl Acad Sci USA* 102: 4936-4942.
- LEVINE, M. and TJIAN, R. (2003). Transcription regulation and animal diversity. *Nature* 424: 147-151.
- LEWIS, B.P., BURGE, C.B. and BARTEL, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120: 15-20.
- LI, Z., VAN CALCAR, S., QU, C., CAVENEE, W.K., ZHANG, M.Q. and REN, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci USA* 100: 8164-8169.
- MAEDA, Y.T. and SANO, M. (2006). Regulatory dynamics of synthetic gene networks with positive feedback. *J Mol Biol* 359: 1107-124.
- MAENO, M., MEAD, P.E., KELLEY, C., XU, R.H., KUNG, H.F., SUZUKI, A., UENO, N. and ZON, L.I. (1996). The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in *Xenopus laevis*. *Blood* 88: 1965-1972.
- MAJETI, R., PARK, C.Y. and WEISSMAN, I.L. (2007). Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* 1: 635-645.
- MANGAN, S. and ALON, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* 100: 11980-11985.
- MANGAN, S., ZASLAVER, A. and ALON, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* 334: 197-204.
- MANI, K.M., LEFEBVRE, C., WANG, K., LIM, W.K., BASSO, K., DALLA-FAVERA, R. and CALIFANO, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol Syst Biol* 4: 169.
- MARGOLIN, A.A., NEMENMAN, I., BASSO, K., WIGGINS, C., STOLOVITZKY, G., DALLA FAVERA, R. and CALIFANO, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 Suppl 1: S7.
- MARSON, A., LEVINE, S.S., COLE, M.F., FRAMPTON, G.M., BRAMBRINK, T., JOHNSTONE, S., GUENTHER, M.G., JOHNSTON, W.K., WERNIG, M., NEWMAN, J. *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134: 521-533.
- MILO, R., ITZKOVITZ, S., KASHTAN, N., LEVITT, R., SHEN-ORR, S., AYZENSHTAT, I., SHEFFER, M. and ALON, U. (2004). Superfamilies of evolved and designed networks. *Science* 303: 1538-1542.
- MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298: 824-827.
- O'DONNELL, K.A., WENTZEL, E.A., ZELLER, K.I., DANG, C.V. and MENDELL, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435: 839-843.
- ODOM, D.T., DOWELL, R.D., JACOBSEN, E.S., GORDON, W., DANFORD, T.W., MACISAAC, K.D., ROLFE, P.A., CONBOY, C.M., GIFFORD, D.K. and FRAENKEL, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 39: 730-732.
- OGLIVY, S., FERREIRA, R., PILTZ, S.G., BOWEN, J.M., GOTTGENS, B. and GREEN, A.R. (2007). The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. *Mol Cell Biol* 27: 7206-7219.
- OKUNO, Y., HUANG, G., ROSENBAUER, F., EVANS, E.K., RADOMSKA, H.S., IWASAKI, H., AKASHI, K., MOREAU-GACHELIN, F., LI, Y., ZHANG, P. *et al.* (2005). Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol Cell Biol* 25: 2832-2845.
- OREN, T., TORREGROZA, I. and EVANS, T. (2005). An Oct-1 binding site mediates activation of the *gata2* promoter by BMP signaling. *Nucleic Acids Res* 33: 4357-4367.
- PATTERSON, L.J., GERING, M., ECKFELDT, C.E., GREEN, A.R., VERFAILLIE, C.M., EKKER, S.C. and PATIENT, R. (2007). The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. *Blood* 109: 2389-2398.
- PATTERSON, L.J., GERING, M. and PATIENT, R. (2005). Scl is required for dorsal aorta as well as blood formation in zebrafish embryos. *Blood* 105: 3502-3511.
- PIMANDA, J.E., DONALDSON, I.J., DE BRUIJN, M.F., KINSTON, S., KNEZEVIC, K., HUCKLE, L., PILTZ, S., LANDRY, J.R., GREEN, A.R., TANNAHILL, D. *et al.* (2007a). The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *Proc Natl Acad Sci USA* 104: 840-845.
- PIMANDA, J.E., OTTERSACH, K., KNEZEVIC, K., KINSTON, S., CHAN, W.Y., WILSON, N.K., LANDRY, J.R., WOOD, A.D., KOLB-KOKOCINSKI, A., GREEN, A.R. *et al.* (2007b). *Gata2*, *Fli1*, and *Scl* form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci USA* 104: 17692-17697.
- RAINIS, L., TOKI, T., PIMANDA, J.E., ROSENTHAL, E., MACHOL, K., STREHL, S., GOTTGENS, B., ITO, E. and IZRAELI, S. (2005). The proto-oncogene ERG in megakaryoblastic leukemias. *Cancer Res* 65: 7596-7602.
- ROBERT-MORENO, A., ESPINOSA, L., DE LA POMPA, J.L. and BIGAS, A. (2005). RBPjkappa-dependent Notch function regulates *Gata2* and is essential for the formation of intra-embryonic hematopoietic cells. *Development* 132: 1117-1126.
- ROSENFELD, N., ELOWITZ, M.B. and ALON, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* 323: 785-793.
- SANCHEZ, M., GOTTGENS, B., SINCLAIR, A.M., STANLEY, M., BEGLEY, C.G., HUNTER, S. and GREEN, A.R. (1999). An SCL 3' enhancer targets developing

- endothelium together with embryonic and adult haematopoietic progenitors. *Development* 126: 3891-3904.
- SANCHEZ, M.J., BOCKAMP, E.O., MILLER, J., GAMBARDELLA, L. and GREEN, A.R. (2001). Selective rescue of early haematopoietic progenitors in Scl(-/-) mice by expressing Scl under the control of a stem cell enhancer. *Development* 128: 4815-4827.
- SHEN-ORR, S.S., MILO, R., MANGAN, S. and ALON, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31: 64-68.
- SIEBURG, H.B., CHO, R.H., DYKSTRA, B., UCHIDA, N., EAVES, C.J. and MULLER-SIEBURG, C.E. (2006). The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. *Blood* 107: 2311-2316.
- SILBERSTEIN, L., SANCHEZ, M.J., SOCOLOVSKY, M., LIU, Y., HOFFMAN, G., KINSTON, S., PILTZ, S., BOWEN, M., GAMBARDELLA, L., GREEN, A.R. *et al.* (2005). Transgenic Analysis of the SCL +19 Stem Cell Enhancer in Adult and Embryonic Haematopoietic and Endothelial Cells. *Stem Cells* 23:1378-88.
- SINCLAIR, A.M., GOTTGENS, B., BARTON, L.M., STANLEY, M.L., PARDANAUD, L., KLAINÉ, M., GERING, M., BAHN, S., SANCHEZ, M., BENCH, A.J. *et al.* (1999). Distinct 5' SCL enhancers direct transcription to developing brain, spinal cord, and endothelium: neural expression is mediated by GATA factor binding sites. *Dev Biol* 209: 128-142.
- SOUCEK, L., WHITFIELD, J., MARTINS, C.P., FINCH, A.J., MURPHY, D.J., SODIR, N.M., KARNEZIS, A.N., SWIGART, L.B., NASI, S. and EVAN, G.I. (2008). Modelling Myc inhibition as a cancer therapy. *Nature* 455: 679-683.
- STEFANI, G. and SLACK, F.J. (2008). Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* 9: 219-230.
- SWIERS, G., PATIENT, R. and LOOSE, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev Biol* 294: 525-540.
- TAGOH, H., HIMES, R., CLARKE, D., LEENEN, P.J., RIGGS, A.D., HUME, D. and BONIFER, C. (2002). Transcription factor complex formation and chromatin fine structure alterations at the murine c-fms (CSF-1 receptor) locus during maturation of myeloid precursor cells. *Genes Dev* 16: 1721-1737.
- TAGOH, H., INGRAM, R., WILSON, N., SALVAGIOTTO, G., WARREN, A.J., CLARKE, D., BUSSLINGER, M. and BONIFER, C. (2006). The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction. *EMBO J* 25: 1070-1080.
- TAOUDI, S., GONNEAU, C., MOORE, K., SHERIDAN, J.M., BLACKBURN, C.C., TAYLOR, E. and MEDVINSKY, A. (2008). Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell* 3: 99-108.
- TOKUOKA, M., IMAI, K.S., SATOU, Y. and SATOH, N. (2004). Three distinct lineages of mesenchymal cells in *Ciona intestinalis* embryos demonstrated by specific gene expression. *Dev Biol* 274: 211-224.
- WADDINGTON, C.H. (1957). *The Strategy of the Genes. A discussion of some aspects of theoretical biology .. With an appendix by H. Kacser*. pp. ix. 262. George Allen & Unwin: London.
- WALMSLEY, M., CIAU-UITZ, A. and PATIENT, R. (2002). Adult and embryonic blood and endothelium derive from distinct precursor populations which are differentially programmed by BMP in *Xenopus*. *Development* 129: 5683-5695.
- WANG, J., RAO, S., CHU, J., SHEN, X., LEVASSEUR, D.N., THEUNISSEN, T.W. and ORKIN, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444: 364-368.
- WEKSBERG, D.C., CHAMBERS, S.M., BOLES, N.C. and GOODELL, M.A. (2008). CD150⁺ side population cells represent a functionally distinct population of long-term hematopoietic stem cells. *Blood* 111: 2444-2451.
- WILSON, N.K., MIRANDA-SAAVEDRA, D., KINSTON, S., BONADIES, N., FOSTER, S.D., CALERO-NIETO, F., DAWSON, M.A., DONALDSON, I.J., DUMON, S., FRAMPTON, J. *et al.* (2009). The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood* 113: 5456-5465.
- XU, J., POPE, S.D., JAZIREHI, A.R., ATTEMA, J.L., PAPATHANASIOU, P., WATTS, J.A., ZARET, K.S., WEISSMAN, I.L. and SMALE, S.T. (2007). Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci USA* 104: 12377-12382.
- YILMAZ, O.H., KIEL, M.J. and MORRISON, S.J. (2006). SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. *Blood* 107: 924-930.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

See our recent Special Issue **Placenta** edited by Joan S. Hunt and Kent L. Thornburg at:
<http://www.ijdb.ehu.es/web/contents.php?vol=54&issue=2-3>

Dynamic expression of Endoglin, a TGF-beta co-receptor, during pre-circulation vascular development in chick

Cantas Alev, Brendan A.S. McIntyre, Kanako Ota and Guojun Sheng
Int. J. Dev. Biol. (2010) 54: 737-742 (doi: 10.1387/ijdb.092962ca)

Identification of *hoxb1b* downstream genes: *hoxb1b* as a regulatory factor controlling transcriptional networks and cell movement during zebrafish gastrulation

Willem M.R. van den Akker, Antony J. Durston and Herman P. Spalink
Int. J. Dev. Biol. (2010) 54: 55-62

Transcriptional regulation by Pax3 and TGFbeta2 signaling: a potential gene regulatory network in neural crest development

Hirohichi Nakazaki, Yueh-Wei Shen, Beth Yun, Anvesh Reddy, Varun Khanna, Barbara Mania-Farnell, Shunsuke Ichi, David G. Mclone, Tadanori Tomita and C. Shekhar K. Mayanil
Int. J. Dev. Biol. (2009) 53: 69-79

Epiblast-derived stem cells in embryonic and adult tissues

Maria P. De-Miguel, Francisco Arnalich-Montiel, Pilar Lopez-Iglesias, Alejandro Blazquez-Martinez and Manuel Nistal
Int. J. Dev. Biol. (2009) 53: 1529-1540

A history of Evo-Devo research in Spain

Jaume Baguña
Int. J. Dev. Biol. (2009) 53: 1205-1217

A novel role of the glial fate determinant glial cells missing in hematopoiesis

Cécile Jacques, Laurent Soustelle, István Nagy, Céline Diebold and Angela Giangrande
Int. J. Dev. Biol. (2009) 53: 1013-1022

Reptile scale paradigm: Evo-Devo, pattern formation and regeneration

Cheng Chang, Ping Wu, Ruth E. Baker, Philip K. Maini, Lorenzo Alibardi and Cheng-Ming Chuong
Int. J. Dev. Biol. (2009) 53: 813-826

The Hox Complex - an interview with Denis Duboule

Michael K. Richardson
Int. J. Dev. Biol. (2009) 53: 717-723

Genomic control of patterning

Isabelle S. Peter and Eric H. Davidson
Int. J. Dev. Biol. (2009) 53: 707-716

5 yr ISI Impact Factor (2009) = 3.253

