Preface

In our time, the sheer volume of published experimental measurements, their scope and technical sophistication, compounded with a proliferation of diverse approaches, objectives, and model systems, has made it particularly difficult to see the conceptual forest for the trees. Yet all of the elegant and sophisticated though disparate and unconnected data sets with which we are confronted represent biological output of the same fundamental operating principles. Each experimental system provides a different window which offers a pathway to these principles. In this book we provide a conceptual framework that we hope will make accessible the principles by which the genomic control system operates developmental and evolutionary process. This framework grows from the realization that the most fundamental causal principles in biology, which distinguish biology from all other sciences, emerge from the existence and function of genomic information. From the genomic sequence are to be recovered the determinants of body plan development in animals. Of course, the processes of biology are subject to the same laws of physics and chemistry as are those of the inanimate world, but it is the genome that mandates biological organization. This is not a metaphor, it is a description of mechanisms that we can now begin to perceive as an unbroken chain of causal connections, leading from the A's, C's, G's and T's of the genomic DNA to the developmental formulation of the elements of the organism.

The new field that is coalescing around the concepts of genomic information processing partakes of principles and evidence from systems biology, developmental molecular biology, various aspects of body plan evolution and phylogenetics, as well as biological engineering and computational modeling. We found it useful to select and incorporate insights from all of these fields, where these illuminate the genomic control of development, without operating wholly within the paradigms of any one of them. In this book we focus on the main characteristics of the genomic control system, which include its hierarchy, its logic processing functions and its structural organization in the form of gene regulatory networks. Such networks encompass at a system level the recognition interactions between transcription factors and DNA sequence that lie at the heart of the whole regulatory process. The general operational properties of genomic regulatory systems are shared across the Bilateria, while diversity in animal forms directly reflects diversity in genomic developmental programs. Focus on the genomic programs controlling development provides a single conceptual lens through which the most disparate phenomena of development and evolution can be viewed, causally understood and interpreted.

A brief précis of the trajectory of our treatment follows:

Chapter 1 is about the molecular biology of sequence-dependent regulation of gene expression in animal development. Here we consider three major levels at which gene expression is controlled, with respect to their roles in the developmental process. The first is transcriptional regulation by sequence-specific transcription factors and their interaction with *cis*-regulatory modules. Secondly, miRNAs modulate transcript prevalence, and a third level operates by means of chromatin modifications installed upon transcription factor-DNA interactions. While the expression of many genes is affected successively by all these mechanisms, control of developmental complexity is executed by transcription factors and their sequence-specific interactions with the regulatory genome. We further conclude that the primary key to the informational process of development lies in the control system regulating the expression of genes encoding transcription factors.

Chapter 2 focusses on the transcriptional control apparatus that operates animal development. Gene regulatory network (GRN) theory defines the principal structural and functional properties of genomic control programs in animals. Here we provide an introductory overview, specifying the components of GRNs, and focusing on higher level design features such as hierarchy, modular organization, and the unidirectionality of these encoded regulatory systems. Two major aspects of GRN output are their generation

of regulatory states that in turn determine all downstream genetic functions, and the Boolean nature of spatial gene expression. The genomic regulatory transactions linked together in GRNs are executed by *cis*-regulatory modules, and their combinatorial information processing functions deeply affect GRN organization. This Chapter further includes a first principles quantitative treatment of network dynamics, which rationalizes the measurable kinetics of accumulation of transcriptional products and permits computational assessment of the outputs of regulatory gene cascades. Current GRN theory devolves from multiple earlier roots which we very briefly trace.

Chapter 3 focusses on the means by which the transcriptional regulatory system is deployed in the pre-gastrular development of many kinds of bilaterian animal. The fundamental task in early embryogenesis is to install in the descendants of a single cell initial patterns of spatial regulatory gene expression in respect to the axes of the future body plan. This poses control challenges unlike any others encountered in bilaterian life. We introduce in this Chapter general principles of development couched in terms of regulatory logic, such as the regulatory properties of the egg, inductive signaling, differentiation and cellular morphogenesis. Beyond these, bilaterian evolution has given rise to several distinct developmental strategies for pre-gastrular embryogenesis, specifically those of regularly cleaving small embryos with early onset of transcription; of large embryos undergoing rapid cell division with delayed onset of transcription and cell motility; of embryos utilizing a transcriptionally active syncytium. Here, for each of these strategies, we focus on particular regulatory design features of GRNs encoding embryonic specification, response to maternal spatial inputs, territorial pattern formation, and global control of zygotic gene expression.

Chapter 4 is about animal body part development. Bilaterian body parts are formed during development according to a common scheme. The basic principles, irrespective of body part or organism, emerge from studies that reveal the genomically encoded mechanisms underlying body part formation. Here we consider partial gene regulatory networks that have been solved for a diverse variety of body parts including brains, hearts, limbs, and guts, in *Drosophila* and/or vertebrates. In all of these cases, the process of body part formation begins with the allocation of a progenitor field, defined by installation of an initial regulatory state, and its placement in respect to the axes of the body plan. There follows the progressively finer subdivision of this field into appropriately positioned regulatory state domains that generate the subparts and ultimately the cell types of the body part. Circuitry commonly encountered in these GRNs controls signaling interactions, exclusion functions and boundary formation, and often involves feedback and feed forward regulation.

Chapter 5 considers the differentiation of cell types as the final readout of the preceding spatial specification processes in development of the body plan. Cell types represent the installation of biological and molecular effector functions within the matrix of spatial regulatory states. Here the focus is on the specific genomically encoded wiring utilized for deployment of cohorts of effector genes in terminal differentiation. Two objectives require to be fulfilled: one is the regional activation and permanent maintenance of expression of a small set of driver regulatory genes; and the other is control of large sets of effector genes by these drivers. The first is accomplished by positive feedback circuitry among the driver genes, usually accompanied by exclusion of drivers of alternative cell fates, and the second depends on effector gene *cis*-regulatory architecture. Our examples include *Drosophila* photoreceptors, and differentiation of erythrocytes, somitic muscle, neural crest and placodes in vertebrates. Diverse body parts often deploy similar cell types and this is mediated by modularity of genomic regulatory systems.

In Chapter 6 we analyze diverse approaches to construction of quantitative and logic models of gene regulatory networks. The structural and functional properties of GRNs can be accessed only by use of models. In this Chapter we consider diverse forms of GRN model, focusing on insights into the biology of developmental GRNs that accrue from the generation of abstract mathematical, logical or topological models. Topological models provide genome-centered maps of regulatory linkages, and thus graphically represent the architecture of causal interaction networks. We address the significance of network topology

in an extensive comparative structure/function analysis of GRN subcircuit architecture. Insights otherwise unattainable emerge from dynamical mathematical treatment of GRN function. We review informative applications of ODE analysis to developmental GRNs operating in *Drosophila* embryos and in mammalian hematopoietic cells, and consider the regulatory interpretation of morphogen gradients. We turn then to Boolean models which are focused on GRN logic transactions. Several asynchronous models are considered and we end with a summary of the new insights deriving from a comprehensive Boolean logic model of the sea urchin embryo GRN.

Chapter 7 provides interpretations of major macro-evolutionary processes in terms of GRN structure and function. Evolution of the animal body plan is the outcome of change in the encoded genomic regulatory program for development. Major features of Phanerozoic animal evolution relate directly to developmental GRN hierarchy. We first consider rapid and continuously adaptive changes occurring at the species level, in terms of regulation of effector gene expression at the periphery of developmental GRNs. We then discuss developmental processes responsible for the generation of definitive characters of the body plan shared amongst all members of given phyla or classes, which occur at upper levels of GRN hierarchy. Regulatory mechanisms also account for the evolutionary stasis of such developmental characters. Distinct evolutionary mechanisms operate at different levels of GRN hierarchy, ranging from *cis*-regulatory adaptation at individual genes to co-optive redeployment of whole regulatory circuits. Conservation of regulatory circuitry within GRN hierarchy leads to an explanation for the nested organization of shared character sets underlying animal phylogeny.

We have maintained the discussion in this book at two different levels. Throughout, we have stressed conceptual themes while, at the same time, tying these concepts to the detailed experimental results that underpin them (these accounts might well exceed the concerns of some readers and could be skipped over without losing the thread of the discussion). Our hope has been to make this book accessible to a wide range of readers. Those with expert knowledge for whom the details may harbor particular fascination will find the specific information in the figure captions more important. For those entering this area from without, whose preoccupation may be to obtain a sense of the overall conceptual structure of genomic control process, we have included a brief introduction into each of the developmental contexts considered. Although having deep historical roots, much of our conceptual framework is only of very recent vintage, as its formulation has depended on the ever increasing accumulation of high quality experimental data such as we have included here.

It is our great pleasure to acknowledge the help we received from some of the world's most renowned scientists in their respective fields, who were kind enough to devote their time to reviewing various of our chapters, thus providing us with invaluable comments, corrections, suggestions and insights. Our reviewers were as follows: for Chapters 1 and 2, Gary Felsenfeld of the National Institutes of Health, and Ellen Rothenberg of Caltech; for Chapter 3, Ellen Rothenberg; for Chapter 4, Marianne Bronner of Caltech, and Roger Patient of the University of Oxford; for Chapter 5, Marianne Bronner and Oliver Hobert of Columbia University; for Chapter 6, Bertie Göttgens of the University of Cambridge, Ellen Rothenberg, Eric Siggia of Rockefeller University, and Stas Shvartsman of Princeton University; for Chapter 7, Ellen Rothenberg and Doug Erwin of the Smithsonian Institution. We are indebted to Ellen Rothenberg, not only for reviewing well over half of our book and providing continuing scientific feedback, criticism and discussion, but also for her ongoing encouragement and enthusiastic support for this project from the time of its inception to its final stages. We are particularly grateful to Yi Krooss for her skillful and professional graphic design, which can be found in every figure of this book. We appreciate the hundreds of permissions we received, allowing us to reproduce the results constituting these figures and our deepest appreciation goes to all the scientists whose intelligence, hard work and discoveries have led to the state of knowledge that this book concerns. We would like to thank Deanna Thomas in our laboratory, who extracted endless references for our use. At Elsevier, it has been a privilege to work with our editor Christine Minihane, who has been endlessly supportive, with Halima Williams, who cheerfully confronted all of our requirements and demands regarding the practical organization of the project, and finally with Julia Haynes, our Project Manager.

In the long view, this book grows out of many years of work on gene regulatory networks at Caltech. In this sense, all of the many participants of the network project over the last 15 years in this lab and elsewhere have contributed to the foundation on which this work stands. For decades now, Eric Davidson has worked intimately with Jane Rigg, whose perspicacity and insights are as always a pleasure to acknowledge. Our experimental and conceptual research on gene regulatory networks has over the years been primarily funded by the National Institutes of Child Health and Development, and in this context we are particularly grateful for the longstanding support of our Program Director, James Coulombe.

On our cover we have reproduced a painting made about a half century ago by Morris Davidson (Eric Davidson's father, 1898-1979), who was a well-known American artist. In the abstract relations of its patterns, this painting for us evokes the beautiful organization of the natural world.

Isabelle S. Peter and Eric H. Davidson
December 2014
Caltech, Pasadena

Dedications

I would like to dedicate this work to my son Milad, whose existence gave me the inspiration and the courage to follow my dreams further than what I would ever have considered possible.

Isabelle S. Peter

I would like to dedicate this book to those scientists whose work has most deeply shaped my intellectual world, none of whom are still with us. They are Theodor Boveri, Alfred Mirsky, and my partner for those many years, Roy J. Britten.

Eric H. Davidson