

Carsten Carlberg · Ferdinand Molnár

Mechanisms of Gene Regulation

Carsten Carlberg • Ferdinand Molnár

Mechanisms of Gene Regulation



Carsten Carlberg
Institute of Biomedicine
School of Medicine
University of Eastern Finland
Kuopio
Finland

Ferdinand Molnár
Institute of Biopharmacy
School of Pharmacy
University of Eastern Finland
Kuopio
Finland

ISBN 978-94-007-7904-4

ISBN 978-94-007-7905-1 (eBook)

DOI 10.1007/978-94-007-7905-1

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013953581

© Springer Science+Business Media Dordrecht 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Mechanisms of Gene Regulation

Preface

This textbook describes the fascinating area of eukaryotic gene regulation. Gene expression is shaping the phenotype of cells and tissues; its regulation therefore is the essential fundamental aspect of nearly all processes in physiology, both in health and in disease. For this reason not only biologists and biochemists should be aware of the concepts of gene regulation, but all students of biomedical areas would benefit from being introduced into this topic, in order to have a good basis for their specialized disciplines. A complete understanding of transcription factors and the processes that alter their activity is a fundamental goal of modern life science research.

The availability of the whole human genome sequence (and that of other eukaryotic genomes) and the consequent development of next-generation sequencing technologies have significantly changed nearly all areas of bioscience. For example, the genome-wide location of histone modifications and transcription factor binding sites, such as provided by the ENCODE consortium, has largely improved our understanding of gene regulation. Therefore, the focus of this book is the description of the post-genome understanding of gene regulation.

The purpose of this book is to provide in a condensed form an overview on the present understanding of the mechanisms of gene regulation. We are not aiming to compete with more comprehensive books, such as the legendary “Genes” of Ben Lewin, but rather will focus on the essentials. In order to facilitate the latter, we favor a high figure-to-text ratio following the rule “a picture tells more than thousand words”.

The content of the book is based on the lecture course “Mechanism of Gene Regulation”, which is given by one of us (C. Carlberg) continuously since 2001 at the University of Eastern Finland in Kuopio. The book is subdivided into four sections and 13 chapters. Following the Introduction there are three sections, which take a view on gene regulation from the perspective of transcription factors, chromatin and non-coding RNA, respectively.

This course is primarily designed for Master level students of Biosciences, but is also frequented by students of other biomedical disciplines and by PhD students. The course and hence this textbook has four major learning objectives. Students should:

1. have detailed understanding of the structure of genes, chromatin organization, transcription factors and their regulatory mechanisms.
2. recognize the key components, mechanisms and processes in gene expression and the multiple layers of its regulatory complexity
3. show the ability to analyze transcription factors, their co-regulators and non-coding RNA concerning their expression and genome-wide effects, i.e. to judge their impact on the complex regulation of eukaryotic genes.
4. apply knowledge in gene regulation in designing, performing and analyzing respective experiments, such as quantitative PCR, microarrays and ChIP-seq.

We hope the readers will enjoy this rather visual book and get as enthusiastic about the topic of gene regulation as the authors are.

Kuopio, August 2013

Carsten Carlberg and Ferdinand Molnár

Acknowledgements

The authors like to thank Prof. Reinhard Bornemann and Dr. Magdalena Schröder for extensive proofreading and constructive criticism.

List of Abbreviations

1,25(OH) ₂ D ₃	1,25-dihydroxyvitamin D ₃
3C	chromosome conformation capture
AID	activation-induced cytidine deaminase
AP1	activator protein 1 (JUN-FOS heterodimer)
APO	apolipoprotein
AR	androgen receptor
ATF3	activating transcription factor 3
atRA	all- <i>trans</i> retinoic acid
BAF	BRG1- or hBRM-associated factors
bp	base pair
BRE	TFIIB-binding element
BrUTP	5-bromouridine 5'-triphosphate
CAR	constitutive androstane receptor
CREBBP	CREB1-binding protein, also known as CBP
CEBP	CCAAT-binding protein
CHD	chromodomain-helicase-DNA-binding
ChIP	chromatin immunoprecipitation
CREB1	cAMP response element-binding protein
CTCF	CCCTC-binding factor
DBD	DNA-binding domain
DGCR8	DiGeorge syndrome critical region gene 8
DHS	DNase I hypersensitive site
DNMT	DNA methyltransferase
DPE	downstream promoter element
DR	direct repeat
DVL	disheveled
EP300	E1A-binding protein p300
ENCODE	encyclopedia of DNA elements
ER	estrogen receptor
eRNA	enhancer RNA
ES cell	embryonic stem cell
EZH2	enhancer of zeste homolog 2
FAIRE	formaldehyde-assisted isolation of regulatory elements

FRAP	fluorescence recovery after photobleaching
FXR	farnesoid X receptor
GCRP	G-coupled cell surface receptor protein
GLI	glioma-associated oncogene homolog
GO	gene ontology
GR	glucocorticoid receptor
GSK3	glycogen synthesis kinase 3
HAT	histone acetyltransferase
HBB	β -globin
HDAC	histone deacetylase
HDM	histone demethylase
HMG	high-mobility group protein
HMT	histone methyltransferase
HP1	heterochromatin protein 1
HSP	heat-shock protein
ICR	imprinted control region
NFKBI	inhibitor of NFKB
IKBK	NFKBI kinase
IGF2	insulin-like growth factor 2
IL	interleukin
INO80	inositol requiring
iPS cell	induced pluripotent stem cell
IRF	interferon-regulatory factor
ISWI	imitation SWI
JAK	janus kinase
JmjC	Jumonji domain-containing
kb	kilo base pairs (1,000 bp)
LAD	lamina-associated domain
LBD	ligand-binding domain
LCR	locus control region
LINE	long interspersed element
LSD1	lysine specific demethylase 1
LXR	liver X receptor
MAP	mitogen-activated protein
MAR	matrix attachment region
MBD	methyl-DNA-binding domain protein
MBP	methyl-binding protein
MDM2	murine double minute-2
MED	mediator
MECP2	methyl-CpG-binding protein 2
MLL	mixed lineage leukemia
mRNA	messenger RNA
miRNA	micro RNA
ncRNA	non-coding RNA
NEMO	NFKB essential modifier
NFKB	nuclear factor κ B

NICD	NOTCH intracellular domain
nt	nucleotides
PHD	plant homeodomain
PKA	cAMP-dependent protein kinase
Pol II	RNA polymerase II
PPAR	peroxisome proliferator-activated receptor
PRC	Polycomb repressive complex
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PTCH	patched receptor
PXR	pregnane X receptor
PWM	position weight matrix
qPCR	quantitative PCR
RAR	retinoic acid receptor
RE	response element
RISC	RNA-induced silencing complex
RNAi	RNA interference
ROR	RAR-related orphan receptor
rRNA	ribosomal RNA
RSC	remodels the structure of chromatin
RUNX1	runt-related transcription factor 1
RXR	retinoid X receptor
SINE	short interspersed element
siRNA	small interfering RNA
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SP1	specificity protein 1
SREBF1	sterol regulatory element-binding transcription factor 1
SRF	serum response factor
SWI/SNF	switching/sucrose nonfermenting
T ₃	triiodothyronine
TAF	TBP-associated factor
TATA box	TATAWADR core DNA sequence
TBP	TATA box-binding protein
TET	ten-eleven translocation
TFF	trefoil factor 1
TLF	TBP-like factor
TLR	Toll-like receptor
TNF	tumor necrosis factor
TRBP	transactivation-response RNA-binding protein
TSS	transcription start site
TR	thyroid hormone receptor
tRNA	transfer RNA
TP53	tumor protein p53
UTR	untranslated region
VDR	vitamin D receptor

Contents

Part I Introduction

1 Overview: What Is Gene Expression?	3
1.1 The Central Dogma of Molecular Biology	3
1.2 Transcriptional Complexity of Genes	7
1.3 Elements of Transcriptional Regulation	9
1.4 The Role of Chromatin Activity in Gene Regulation	11
1.5 Gene Expression Programs	12
1.6 Key Concepts	14
Further Reading	15
2 The Impact of Chromatin	17
2.1 Eu- and Heterochromatin	18
2.2 Nucleosomes	20
2.3 Histone Modifications in Chromatin Opening and Closing	23
2.4 Genomic Effects of Histone Modifications	24
2.5 Chromatin Architecture	29
2.6 Impact of Epigenetic Signaling in Health and Disease	31
2.7 Key Concepts	33
Further Reading	34

Part II Transcription Factor View

3 The Basal Transcriptional Machinery	37
3.1 The Core Promoter	38
3.2 The TATA Box	42
3.3 Core Promoter Elements	44
3.4 TFIID as a Paradigm of a Multi-Protein Complex	46
3.5 Genome-Wide Approaches on Core Promoter Identification	47
3.6 The Mediator Complex	50
3.7 Key Concepts	53
Further Reading	54

4	Transcription Factors	55
4.1	Site-Specific Transcription Factors	55
4.2	Transcription Factor Domains.....	56
4.3	Dimeric Transcription Factor Complexes: The Example of Nuclear Receptor Heterodimers.....	59
4.4	Bioinformatic Identification of Transcription Factor-Binding Sites.....	62
4.5	Transcription Factor Expression Profile.....	64
4.6	Classification of Transcription Factors	65
4.7	Key Concepts	69
	Further Reading.....	70
5	Linking Signal Transduction and Gene Regulation.....	71
5.1	Activation of Latent Transcription Factors	72
5.2	Transcription Factor Networks.....	74
5.3	Programming Cellular Differentiation by Transcription Factors.....	76
5.4	NFKB Signaling.....	79
5.5	Transcription Factors in the Inflammatory Response	83
5.6	Sensing Cellular Stress: The p53 Pathway.....	85
5.7	Key Concepts	88
	Further Reading.....	89
6	Switching Genes on and off: The Example of Nuclear Receptors.....	91
6.1	The Nuclear Receptor Superfamily.....	92
6.2	Molecular Interactions of Nuclear Receptors	95
6.3	Physiological Role of Nuclear Receptors	97
6.4	Nuclear Receptors and Their Ligands.....	99
6.5	Interaction of Nuclear Receptors with Co-Factors.....	101
6.6	Key Concepts	103
	Further Reading.....	104
7	Mapping the Genome	105
7.1	Phylogenetic Footprinting.....	105
7.2	The ENCODE Project	108
7.3	Exploring Data of the ENCODE Project	111
7.4	Integrating ChIP-seq Data.....	112
7.5	Alternative DNA-Binding Modes of Transcription Factors.....	114
7.6	Assigning Transcription Factors to Their Target Genes.....	116
7.7	Key Concepts	118
	Further Reading.....	121

Part III Chromatin View

8 Chromatin Modifiers	125
8.1 The Histone Code Model	126
8.2 Histone Modifying Enzymes.....	130
8.3 Genome-Wide Analysis of Histone Markers and Their Modifying Enzymes.....	134
8.4 Chromatin Modifiers in Disease	137
8.5 Key Concepts	139
Further Reading.....	141
9 Genomic Imprinting	143
9.1 Insulators	143
9.2 The Genome Regulator CTCF	145
9.3 Genomic Imprinting	147
9.4 Models of Insulator Function	149
9.5 Heritance of CTCF-Mediated Chromatin Structures	151
9.6 Key Concepts	153
Further Reading.....	154
10 The Epigenome	155
10.1 DNA Methylation.....	156
10.2 DNA Methylation on the Genome Scale.....	157
10.3 DNA Methylation, Heterochromatin and Gene Silencing	159
10.4 Epigenomics in Health and Disease.....	163
10.5 Key Concepts	165
Further Reading.....	167
11 Chromatin Remodeling	169
11.1 The Impact of Chromatin Remodeling.....	169
11.2 ATP-Dependent Remodeling Complexes.....	170
11.3 Nucleosome Positioning at Promoters	174
11.4 Pioneer Factors	176
11.5 Transcriptional Dynamics and Noise	178
11.6 Key Concepts	181
Further Reading.....	182
12 Chromatin Architecture	183
12.1 Organization of the Nucleus.....	183
12.2 Larger Order DNA Loop Formation	186
12.3 Compartmentalization of Nuclear Processes	188
12.4 Transcription Factories.....	190
12.5 Nuclear Positioning and Disease.....	191
12.6 Key Concepts	192
Further Reading.....	193

Part IV Non-Coding RNA View

13 Regulatory RNA	197
13.1 Non-coding RNAs.....	197
13.2 miRNAs and their Function	199
13.3 Long ncRNA	203
13.4 eRNAs	204
13.5 Gene Regulation by miRNAs and Transcription Factors.....	207
13.6 Key Concepts	209
Further Reading.....	210

Part I

Introduction



Chapter 1

Overview: What Is Gene Expression?

Abstract The human genome (see Box 1.1) is composed of some 20,000 protein-coding genes and approximately the same number of genes for non-coding RNAs (ncRNAs), being used as a structural backbone of ribosomes (ribosomal RNA (rRNAs)), adaptor molecules in protein translation (transfer RNA (tRNAs)) and regulators of mRNA stability, protein translation and chromatin density (long non-coding (nc) RNAs and micro RNAs (miRNAs), for more details see Sect. 13.2). A gene is considered as “expressed”, when it is transcribed into RNA. In a given tissue or cell type only approximately half of all genes are transcribed. Moreover, the phenotype of every tissue is defined by its own characteristic pattern of transcribed genes, which changes, when the cells are exposed to a signal, such as a dietary or a stress molecule.

Tissue- and signal-specific gene expression is the central mechanism to control the general properties of a cell and its response to environmental perturbations. The latter applies both to healthy and disease conditions. A detailed insight into gene expression therefore bears the potential not only for an understanding of the dysregulation in disease states but also for their therapeutic treatment.

In this chapter, we will provide an overview of the central features of gene expression. We will describe that gene expression is tightly controlled, in particular on the level of gene transcription. There are at least three levels of control, which are the DNA code, the epigenetic code and the transcription factor program.

Keywords Protein-coding genes · Transcriptome · Genome · RNA polymerases · qPCR · Microarray · Chromatin · Nucleosomes · Enhancer · Promoter · Locus control region · Epigenetics · Transcription start site · Gene expression

1.1 The Central Dogma of Molecular Biology

The so-called “central dogma of molecular biology” indicates a clear direction in the flow of information from DNA to RNA to protein (Fig. 1.1). This means that besides a few exceptions, such as reverse transcription of the RNA genome of retroviruses, genomic DNA stores the building plan of all pro- and eukaryotic organisms. Accordingly, genes are defined as those regions of genomic DNA that can be transcribed into RNA. In this traditional view the RNA meant is mRNA, i.e. the template used for protein translation. However, nowadays many other forms of