Carsten Carlberg · Ferdinand Molnár

Mechanisms of Gene Regulation

Springer

Carsten Carlberg • Ferdinand Molnár

Mechanisms of Gene Regulation

Springer

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

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 there 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 Schräder for extensive proofreading and constructive criticism.

List of Abbreviations

$1,25(OH)_2D_3$	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-*trans* 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	embyronic 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 κB

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 III Chromatin View

Part IV Non-Coding RNA View

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

C. Carlberg, F. Molnár, *Mechanisms of Gene Regulation*,
DOI 10.1007/978-94-007-7905-1_1,