Physiology of the Gastrointestinal Tract

Sixth Edition

Physiology of the Gastrointestinal Tract

Sixth Edition
Volumes 1 and 2

Editor-in-Chief

Hamid M. Said
Departments of Medicine and Physiology/Biophysics
University of California-School of Medicine
Irvine, CA, United States

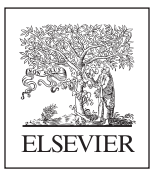
Associate Editors

Fayez K. Ghishan
Department of Pediatrics
Steele Children's Research Center
University of Arizona Health Sciences Center
Tucson, AZ, United States

Jonathan D. Kaunitz
Greater Los Angeles VA Healthcare System
Departments of Medicine and Surgery
UCLA School of Medicine
Los Angeles, CA, United States

Juanita L. Merchant
Departments of Internal Medicine and Molecular and Integrative Physiology
University of Michigan
Ann Arbor, MI, United States

Jackie D. Wood
Departments of Physiology and Cell Biology and Internal Medicine
College of Medicine, The Ohio State University
Columbus, OH, United States



ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-809954-4 (Set)
ISBN: 978-0-12-815901-9 (Volume 1)
ISBN: 978-0-12-815902-6 (Volume 2)

Publisher: John Fedor
Acquisition Editor: Stacy Masucci
Editorial Project Manager: Sam W. Young
Production Project Manager: Punithavathy Govindaradjane
Cover Designer: Matthew Limbert

Typeset by SPi Global, India

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

VOLUME 1

Yasutada Akiba (709), Greater Los Angeles Veterans Affairs Medical Center; UCLA School of Medicine, Los Angeles, CA, United States

Denise Al Alam (71), Children's Hospital Los Angeles, Los Angeles, CA, United States

Rana Al-Sadi (587), Penn State University College of Medicine, Hershey, PA, United States

Qasim Aziz (373), Queen Mary University of London, London, United Kingdom

Eric J. Battaglioli (755), Mayo Clinic, Rochester, MN, United States

Adil E. Bharucha (517), Mayo Clinic College of Medicine, Rochester, MN, United States

Richard S. Blumberg (767), Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Natacha Bohin (141), The University of Michigan, Ann Arbor, MI, United States

Joel C. Bornstein (429), University of Melbourne, Parkville, VIC, Australia

Stuart M. Brierley (387), Flinders University, SAHMRI, Adelaide, SA, Australia

Simon J.H. Brookes (517), Flinders University, Adelaide, SA, Australia

Joel Castro (387), Flinders University, SAHMRI, Adelaide, SA, Australia

Eugene B. Chang (795), University of Chicago, Chicago, IL, United States

Benoit Chassaing (775), Georgia State University, Atlanta, GA, United States

Mary Cheung (485), Long Island Jewish Hospital, New Hyde Park; Hofstra Northwell Medical School, Hempstead, NY, United States

Matthew A. Ciorba (221), Washington University School of Medicine in St. Louis, St. Louis, MO, United States

Sheila E. Crowe (663), University of California San Diego, La Jolla, CA, United States

Michael Czerwinski (185), University of Michigan Medical School, Ann Arbor, MI, United States

Soula Danopoulos (71), Children's Hospital Los Angeles, Los Angeles, CA, United States

Soumita Das (663), University of California San Diego, La Jolla, CA, United States

Peter J. Dempsey (141), The University of Colorado Denver – Anschutz Medical Campus, Aurora, CO, United States

Gerco den Hartog (663), University of California San Diego, La Jolla, CA, United States

Hideki Enomoto (273), Kobe University, Kobe, Japan

Andelain Erickson (387), University of Adelaide, SAHMRI, Flinders University, Adelaide, SA, Australia

Peter B. Ernst (663), University of California San Diego, La Jolla; School of Veterinary Medicine, UC Davis, Davis, CA, United States

Adam D. Farmer (373), Queen Mary University of London, London; University Hospitals of North Midlands, Staffordshire, United Kingdom

Jaime P.P. Foong (429), University of Melbourne, Parkville, VIC, Australia

Mark R. Frey (71,683), Children's Hospital Los Angeles; University of Southern California, Los Angeles, CA, United States

Andrew T. Gewirtz (775), Georgia State University, Atlanta, GA, United States

Fayez K. Ghishan (235), Steele Children's Research Center, University of Arizona Health Sciences Center, Tucson, AZ, United States

Fiona M. Gribble (31), University of Cambridge, Cambridge, United Kingdom

Luke Grundy (387), Flinders University, SAHMRI, Adelaide, SA, Australia

Marlene M. Hao (273), University of Melbourne, Melbourne, VIC, Australia

Andrea M. Harrington (387), University of Adelaide, SAHMRI, Flinders University, Adelaide, SA, Australia

Grant W. Hennig (469), University of Vermont, Burlington, VT, United States

Hongzhen Hu (337), Washington University School of Medicine, Saint Louis, MO, United States

Jan D. Huizinga (305), McMaster University, Hamilton, ON, Canada

Shankar S. Iyer (767), Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Izumi Kaji (671), Vanderbilt University Medical Center, Nashville, TN, United States

Purna C. Kashyap (755), Mayo Clinic, Rochester, MN, United States

Jonathan D. Kaunitz (671,709), Greater Los Angeles VA Healthcare System; UCLA School of Medicine, Los Angeles, CA, United States

Jennifer S. Labus (419), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Brigitte Lavoie (453), University of Vermont, Burlington, VT, United States

Cambrian Y. Liu (683), Children's Hospital Los Angeles; University of Southern California Keck School of Medicine, Los Angeles, CA, United States

Thomas Y. Ma (587), Penn State University College of Medicine, Hershey, PA, United States

Xiaoya Ma (131), University of Michigan Medical School, Ann Arbor, MI, United States

Gary M. Mawe (453), University of Vermont, Burlington, VT, United States

Juanita L. Merchant (3), University of Michigan, Ann Arbor, MI, United States

Jeannette S. Messer (795), University of Chicago, Chicago, IL, United States

Larry Miller (485), Long Island Jewish Hospital, New Hyde Park; Hofstra Northwell Medical School, Hempstead, NY, United States

Bruce D. Naliboff (419), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Mark T. Nelson (453), University of Vermont, Burlington, VT, United States

Donald F. Newgreen (273), Royal Children's Hospital, Parkville, VIC, Australia

Prashant Nighot (587), Penn State University College of Medicine, Hershey, PA, United States

Monica Passi (485), Long Island Jewish Hospital, New Hyde Park; Hofstra Northwell Medical School, Hempstead, NY, United States

D. Brent Polk (683), Children's Hospital Los Angeles; University of Southern California Keck School of Medicine, Los Angeles, CA, United States

Maria J. Pozo (453), University of Extremadura, Cáceres, Spain

Frank Reimann (31), University of Cambridge, Cambridge, United Kingdom

Geoffrey P. Roberts (31), University of Cambridge, Cambridge, United Kingdom

Bani C. Roland (485), Lenox Hill Hospital, New York, NY, United States

James K. Ruffle (373), Queen Mary University of London, London, United Kingdom

Hyder Said (709), UC Irvine School of Medicine, Irvine, CA, United States

Linda C. Samuelson (141), The University of Michigan, Ann Arbor, MI, United States

Michael A. Schumacher (71), Children's Hospital Los Angeles; University of Southern California Keck School of Medicine, Los Angeles, CA, United States

Yatrik M. Shah (131), University of Michigan Medical School, Ann Arbor, MI, United States

Terez Shea-Donohue (641), University of Maryland School of Medicine, Baltimore, MD, United States

Noah F. Shroyer (185), Baylor College of Medicine, Houston, TX, United States

Jason R. Spence (185), University of Michigan Medical School, Ann Arbor, MI, United States

Nick J. Spencer (337,469), Flinders University of South Australia, Bedford Park, SA, Australia

Stephanie N. Spohn (737), University of Michigan Medical School, Ann Arbor, MI, United States

Lincon A. Stamp (273), University of Melbourne, Melbourne, VIC, Australia

William F. Stenson (221), Washington University, St Louis, MO, United States

Miyako Takaki (289), Nara Medical University, Kashihara, Nara, Japan

Kirsten Tillisch (419), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Toshihiro Uesaka (273), Kobe University, Kobe, Japan

Gijs R. van den Brink (103), Tytgat Institute for Liver and Intestinal Research; Academic Medical Center, Amsterdam, The Netherlands

Willemijn A. van Dop (103), Tytgat Institute for Liver and Intestinal Research; Academic Medical Center, Amsterdam, The Netherlands

Anil Vegesna (485), Long Island Jewish Hospital, New Hyde Park; Hofstra Northwell Medical School, Hempstead, NY, United States

Jakob von Moltke (721), University of Washington, Seattle, WA, United States

Arnold Wald (565), University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

B. Florien Westendorp (103), Tytgat Institute for Liver and Intestinal Research; Academic Medical Center, Amsterdam, The Netherlands

Mathew Whitson (485), Long Island Jewish Hospital, New Hyde Park; Hofstra Northwell Medical School, Hempstead, NY, United States

Jackie D. Wood (361), College of Medicine, The Ohio State University, Columbus, OH, United States

Hua Xu (235), Arizona Health Sciences Center, University of Arizona, Tucson, AZ, United States

Vincent W. Yang (197), Stony Brook University School of Medicine, Stony Brook, NY, United States

Heather M. Young (273), University of Melbourne, Melbourne, VIC, Australia

Vincent B. Young (737), University of Michigan Medical School, Ann Arbor, MI, United States

VOLUME 2

Nada A. Abumrad (1087), Washington University School of Medicine in St. Louis, St. Louis, MO, United States

Waddah A. Alrefai (1317), University of Illinois at Chicago; Jesse Brown VA Medical Center, Chicago, IL, United States

Rana Ammoury (1669), Children's Hospital of The King's Daughters, Norfolk, VA, United States

Gregory J. Anderson (1451), QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

Shinji Asano (831), Ritsumeikan University, Shiga, Japan

Giorgos Bamias (1579), National and Kapodistrian University of Athens, Athens, Greece

Yangzom D. Bhutia (1063), Texas Tech University Health Sciences Center, Lubbock, TX, United States

Niviann M. Blondet (981), University of Washington and Seattle Children's Hospital, Seattle, WA, United States

Patrick Borel (1181), Aix Marseille Univ, INSERM, INRA, C2VN, Marseille, France

Vincenza Cifarelli (1087), Washington University School of Medicine in St. Louis, St. Louis, MO, United States

James F. Collins (1451), University of Florida, Gainesville, FL, United States

Robert J. Cousins (1485), University of Florida, Gainesville, FL, United States

Nicholas O. Davidson (1109), Washington University School of Medicine in St. Louis, St. Louis, MO, United States

Paul A. Dawson (931), Emory University, Atlanta, GA, United States

Guillaume de Lartigue (1501), The John B. Pierce Laboratory; Yale Medical School, New Haven, CT, United States

Charles Desmarchelier (1181), Aix Marseille Univ, INSERM, INRA, C2VN, Marseille, France

Pradeep K. Dudeja (1317), University of Illinois at Chicago; Jesse Brown VA Medical Center, Chicago, IL, United States

Shireen R.L. Flores (1451), University of Florida, Gainesville, FL, United States

Vadivel Ganapathy (1063), Texas Tech University Health Sciences Center, Lubbock, TX, United States

Chiara Ghezzi (1051), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Fayez K. Ghishan (1273,1405,1669), Steele Children's Research Center, University of Arizona Health Sciences Center, Tucson, AZ, United States

Ravinder K. Gill (1547), University of Illinois at Chicago, Chicago, IL, United States

Fred S. Gorelick (869), Yale University School of Medicine, West Haven, CT, United States

Matthew B. Grisham (1579), Texas Tech University Health Sciences Center, Lubbock, TX, United States

Earl H. Harrison (1133), The Ohio State University, Columbus, OH, United States

Gail A. Hecht (1547), Loyola University Medical Center and Stritch School of Medicine, Maywood, IL, United States

Dawn A. Israel (1517), Vanderbilt University School of Medicine, Nashville, TN, United States

James D. Jamieson (869), Yale University School of Medicine, West Haven, CT, United States

Bryson W. Katona (1615), University of Pennsylvania, Philadelphia, PA, United States

Pawel R. Kiela (1273,1405), Steele Children's Research Center; University of Arizona, Tucson, AZ, United States

Rachel E. Kopec (1133), The Ohio State University, Columbus, OH, United States

Kris V. Kowdley (981), Swedish Medical Center, Seattle, WA, United States

Nicholas F. LaRusso (1003), Mayo Clinic, Rochester, MN, United States

Seong M. Lee (1153), University of Wisconsin-Madison, Madison, WI, United States

Rodger A. Liddle (895), Duke University Medical Center; Veterans Affairs Medical Center, Durham, NC, United States

Juan P. Liuzzi (1485), Florida International University, Miami, FL, United States

Donald D.F. Loo (1051), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

John P. Lynch (1615), University of Pennsylvania, Philadelphia, PA, United States

Anatoliy I. Masyuk (1003), Mayo Clinic, Rochester, MN, United States

Tatyana V. Masyuk (1003), Mayo Clinic, Rochester, MN, United States

Donald J. Messner (981), Bastyr University, Kenmore, WA, United States

Mark B. Meyer (1153), University of Wisconsin-Madison, Madison, WI, United States

Karen F. Murray (981), University of Washington and Seattle Children's Hospital, Seattle, WA, United States

Ebba Nexo (1201), Aarhus University Hospital, Aarhus, Denmark

Curtis T. Okamoto (831), University of Southern California, Los Angeles, CA, United States

Bernardo Ortega (1027), SUNY Brockport, Brockport, NY, United States

Stephen Pandol (869), Cedars-Sinai Medical Center and University of California, Los Angeles, CA, United States

Richard M. Peek Jr. (1517), Vanderbilt University School of Medicine, Nashville, TN, United States

J. Wesley Pike (1153), University of Wisconsin-Madison, Madison, WI, United States

Shubha Priyamvada (1317), University of Illinois at Chicago, Chicago, IL, United States

Gordon B. Proctor (813), King's College London, London, United Kingdom

Vazhaikkurichi M. Rajendran (1363), West Virginia University School of Medicine, Morgantown, WV, United States

Helen E. Raybould (1501), University of California Davis, Davis, CA, United States

Jesus Rivera-Nieves (1579), University of California San Diego, La Jolla, CA, United States

Hamid M. Said (1197,1201), University of California-School of Medicine, Irvine, CA, United States

Hideki Sakai (831), University of Toyama, Toyama, Japan

Seema Saksena (1317), University of Illinois at Chicago; Jesse Brown VA Medical Center, Chicago, IL, United States

Monica Sala-Rabanal (1051), Washington University School of Medicine, Saint Louis, MO, United States

Jörg-Dieter Schulzke (1363), Institute of Clinical Physiology, Berlin, Germany

Ursula E. Seidler (1363), Hannover Medical School, Hannover, Germany

Abeer K. Shaalan (813), King's College London, London, United Kingdom

Irshad A. Sheikh (1669), Steele Children's Research Center, University of Arizona Health Sciences Center, Tucson, AZ, United States

Jay R. Thiagarajah (1249), Harvard Medical School, Boston, MA, United States

Alan S. Verkman (1249), University of California, San Francisco, CA, United States

Xiaoyu Wang (1451), University of Florida, Gainesville, FL, United States

Paul A. Welling (1027), University of Maryland School of Medicine, Baltimore, MD, United States

Allan W. Wolkoff (957), Marion Bessin Liver Research Center; Albert Einstein College of Medicine, Bronx, NY, United States

Jackie D. Wood (1643), College of Medicine, The Ohio State University, Columbus, OH, United States

Ernest M. Wright (1051), David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

Preface

This is the sixth edition of *Physiology of the Gastrointestinal Tract,* a book that was started by the distinguished Dr. Leonard R. Johnson some 38 years ago with an aim of presenting a source for a most up-to-date and comprehensive knowledge of the field of gastrointestinal physiology. Our aim in this edition was to maintain the same outstanding level of knowledge that Dr. Johnson established by involving the most prominent and recognized authorities in the different fields of investigation in gastrointestinal physiology. This project has been the product of a collective effort with my distinguished colleagues Drs. Ghishan, Kaunitz, Merchant, and Wood and the valuable advices provided by Dr. Johnson all along. These editors have been responsible for recruiting authors and for the scientific editing of most chapters. Their enthusiasm and dedication are indeed what made this project possible. I am also especially grateful to Sam Young, Stacey Masucci, and Anita Mercy Vethakkan for contacting our distinguished authors and extending a friendly line of help to them in all issues related to the preparation and submission of their chapters. Overall, the sixth edition is organized to be useful for a wide spectrum of gastrointestinal interests ranging from graduate students and postdoctoral fellows, active investigators, and clinicians. The volume emphasizes the broad scope of research in gastrointestinal physiology. The editors hope that the volume conveys the excitement and significance of modern research in the many aspects of gastrointestinal function.

Hamid M. Said

Chapter 1

Transcription and Epigenetic Regulation

Juanita L. Merchant

With the human genome sequencing project completed in 2001, perhaps the most important piece of information that we have learned is that the clues to our genetic destiny are contained in more than just the primary sequence of DNA encoding 20,500 proteins.[1] Apparently, what distinguishes man from other life forms and most interestingly other mammals lies in the complex modifications, organization, and function of the 3.3 billion nucleotides (nt). Not only are these ~20,500 genes alternatively spliced, but their DNA, RNA, and protein products are chemically modified so as to change gene function. Therefore, as opposed to our genetic template being composed of a mere 20,500 genetic units, we are actually controlled by 20,500 to the nth power. The exponent has yet to be determined, but likely results in an enormous, perhaps infinite, combination of genetic events. This chapter will briefly summarize our basic understanding of gene expression, but will focus primarily on the new concepts and technologies of gene regulation in the postgenomic era. Arguably, the major advances since the 5th edition of this textbook continue to be the explosion in our understanding of noncoding RNAs, the impact of epigenetics, chromatin topology, and the refinement of high-throughput techniques.

1.1. OVERVIEW OF GENE ORGANIZATION

1.1.1 Nucleic Acids

The molecular definition of a eukaryotic gene is complex, but in the simplest terms, it is a nucleic acid sequence that encodes one polypeptide and one messenger ribonucleic acid molecule (mRNA).[2] Genes are comprised of "two intertwining polymers" of deoxyribonucleic acids (DNAs) that are noncovalently attached to a variety of proteins, including histones and specialized proteins (e.g., polymerases and various accessory proteins). The association of DNA, histones, and specialized nuclear proteins collectively is called chromatin. Chromosomes are comprised of continuous strands of chromatin that have been compacted by supercoiling and looping so as to fit into the nucleus (Fig.1.1). The steps governing the compacting and location of chromatin are now an area of intense investigation and will be discussed in Section 1.2. Chromosomes are the basic heritable unit in the mammalian cell. In humans, there are 46 individual chromosomes or 23 chromosome pairs. The smallest unit of the DNA polymer is a nucleotide—a base attached to the first carbon of a five-carbon sugar phosphorylated at its fifth carbon (Fig. 1.2). Nucleosides do not contain phosphates linked to the pentose sugar, thus differing from nucleotides, which contain one, two, or three phosphate groups. The type of base distinguishes the 4 nt found in DNA: adenine (A), thymine (T), cytosine (C), or guanine (G). They are bases because of the nitrogen groups contained within their single-ring (thymine, cytosine, or uracil) or double-ring structures (adenine or guanine). DNA contains the sugar *deoxy*ribose, whereas RNA contains the sugar ribose and the base uracil (U) instead of thymine. CpG islands are dinucleotides consisting of a deoxycytidine in the 5′ position adjacent to deoxyguanosine. These dinucleotides are "hot spots" for enzymes (e.g., DNMTs = DNA methyl transferases), which add a methyl group to the 5th carbon of the cytosine ring. The "p" indicates that one phosphate group separates these two nucleosides. This epigenetic mark blocks the expression of DNA and is a mechanism used frequently by gastrointestinal (GI) cancers to silence genes that block their ability to proliferate.[3]

1.1.2 Nucleic Acid Polymers: DNA, RNA

Polymers of nucleotides or nucleic acids (also called nucleoside mono-, di-, or triphosphates) are formed when the free phosphate group attached to the fifth carbon of an adjacent nucleotide of the pentose sugar condenses with the hydroxyl group on the third pentose carbon to produce two ester bonds and water (phosphodiester bond). Accordingly, the proximal end of each DNA strand (5′ end) contains a phosphate group at the fifth carbon of the deoxyribose sugar residue. The terminal nucleic acid at the 3′ end of each DNA strand contains a free hydroxyl group at the third carbon of the deoxyribose ring. By convention, nucleotide sequences are written from 5′ to 3′ reading from left to right with the sense strand presented as the upper strand. The

Physiology of the Gastrointestinal Tract. https://doi.org/10.1016/B978-0-12-809954-4.00001-3

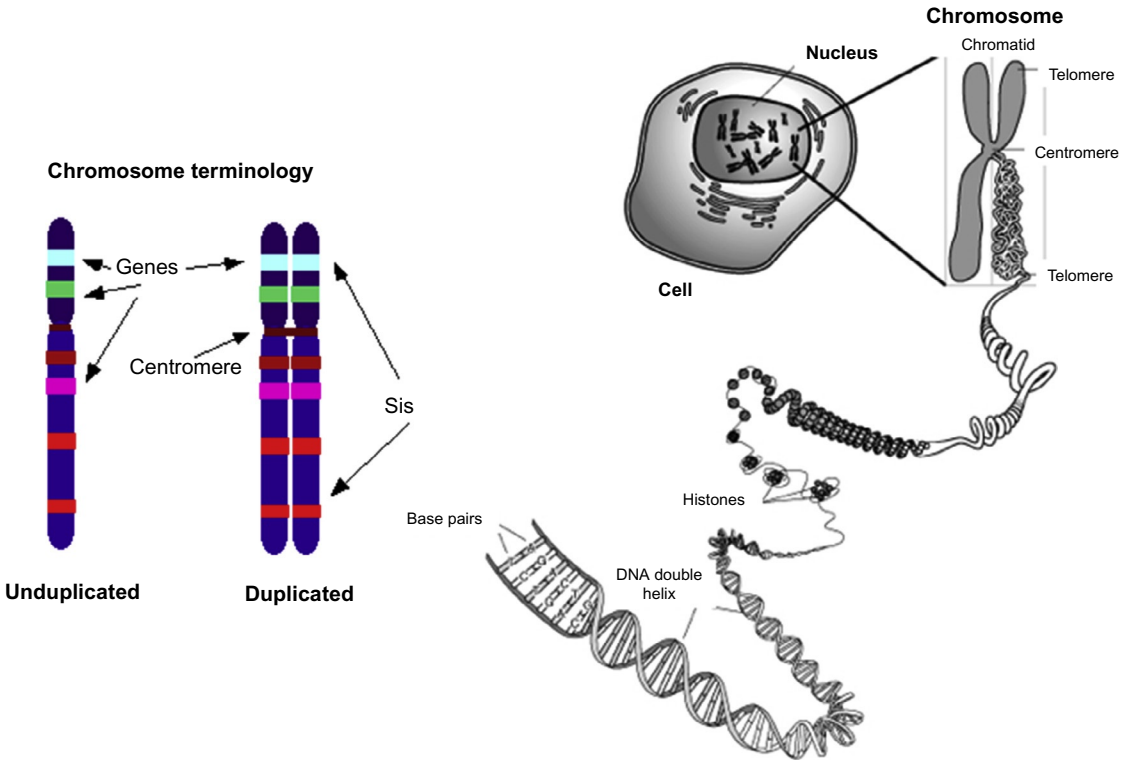


FIG. 1.1 Chromatin structure and organization. Each chromosome exists in a haploid (germ cells) or diploid/tetraploid state depending on their stage in the cell cycle. The short arm of the chromosome relative to the centromere is the "p" arm and the long arm is the "q" arm. Chromosomes represent compressed, compacted DNA double strand helix wrapped around core histones. *(From the* Language of medicine. *4th ed.)*

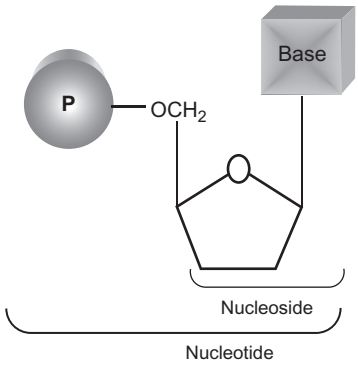


FIG. 1.2 Nucleic acid structure. A nucleoside consists of a purine or pyrimidine base covalently linked to the firs carbon of the pentose ring. The addition of one, two, or three phosphate groups is a nucleotide mono-, di-, or triphosphate. The type of sugar determines the type of nucleic acid: ribose in ribonucleic acids (RNAs) and deoxyribose in deoxyribonucleic acids (DNAs). *(Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

antisense strand, written on the bottom, is antiparallel and complementary to the sense strand so that the 5′ to 3′ direction proceeds from right to left. Each nucleotide within the polymer is base paired with a particular nucleotide on the opposing strand by hydrogen bonds: adenine with thymine and guanine with cytosine. The DNA strand containing the same sequence as the messenger RNA (mRNA) is designated the *sense strand*, and the strand that it pairs with is designated the *antisense strand*. The antisense strand becomes the template sequence that will be transcribed by RNA polymerase II (Pol II) into mRNA and subsequently translated into amino acids.[4]

Most of the studies on transcriptional control focus on genes transcribed by the 12-subunit enzyme Pol II and thus are designated as class II genes.[5,6] It is Pol II that is responsible for transcribing gene sequences into protein-encoding messenger RNA (mRNA). Less than 2% of total RNA in the cell is mRNA. Many of these initial primary transcripts (hnRNA for heterogeneous nuclear RNA) are further processed as discussed below. Therefore, 98% of the nucleotides in the human genome do not reside in exons (sequences that encode proteins). Nevertheless, at least 50% of the noncoding RNA is transcribed and serves a function. Nine percent of cellular RNA is hnRNA, the bulk of which are small nuclear RNAs (snRNA, e.g., U2 involved in RNA splicing, 4%) and small nucleolar RNAs, for example, U22 snoRNA, comprising 1%. The other 4% of hnRNA is mRNA. An additional 1% of total cell RNA is microRNA (miRNA), previously called guide RNA (gRNA), which edits its mature mRNA transcripts.[7] RNA polymerase I (Pol I) transcribes all of the ribosomal genes except for the 5S

gene. Ribosomal RNA represents about 75% of the RNA in the cell and is therefore essential for translation.[8,9] RNA polymerase III (Pol III) transcribes the 5S ribosomal gene and the genes-encoding transfer RNA (tRNA).[10] Transfer RNA represents about 15% of the total RNA in the cell. Pol I and III transcribe genes that will not be further translated into peptides and noncoding RNA transcripts, although their primary transcripts are also processed before reaching the cytoplasm. Since Pol II transcribes genes-encoding proteins, peptides, long noncoding RNA (lncRNA), and miRNAs, Pol II-regulated genes will be the primary focus of this chapter.

1.1.3 Gene Composition

A gene is analogous to a long sentence read from left to right and comprised of letters organized into words separated by spaces and punctuations. Specific DNA sequences "punctuate" the gene with important start and stop signals for transcription and translation. Several hundred to several thousand DNA base pairs (bp) may comprise one gene. These bp (the alphabet) are organized into functional groups (phrases) on the basis of whether a particular sequence is untranscribed, only transcribed (RNA), or both transcribed and translated (RNA and protein) (Fig. 1.3). Exons are DNA sequences that are transcribed into mRNA by Pol II and exit the nucleus. Within the cytoplasm, exons may or may not be translated into peptides. Those exons that are transcribed *and* translated form the coding sequences (coding exon). In general, the term intron is used to describe the intervening DNA sequence that is transcribed but is subsequently removed from the primary transcript by RNA splicing (RNA processing) before exiting the nucleus as a mature transcript. However, it is now clear that many transcribed DNA sequences generate small noncoding RNA transcripts such as miRNAs or lncRNAs that can inhibit or modulate protein-coding genes in "cis or trans." LncRNAs are commonly defined as transcripts that are >200 nt that do not encode a protein compared to the significantly shorter miRNAs.

DNA sequences or elements that regulate transcription and are not transcribed into mRNA usually reside in the 5′ portion of a gene upstream (to the left) of the promoter. The promoter is a cluster of DNA sequences that binds Pol II in concert with accessory proteins to initiate the synthesis of mRNA. Accessory proteins control the accuracy and rate of polymerase binding. The first nucleotide transcribed into mRNA is assigned the number 1 with subsequent nucleotides (downstream or to the right of the promoter) assigned positive numbers as transcription proceeds toward the 3′ end. Nucleotides preceding the promoter (upstream or 5′) are assigned negative numbers. DNA sequences that encode a polypeptide (open reading frame) begin with the translational start site codon ATG (encoding methionine) and end with one of the three stop codons: TAA, TAG, and TGA. Thus, the translational start and three stop codons, respectively, are transcribed into mRNA as AUG, UAA, UAG, and UGA. Since there are four different DNA bases and it takes

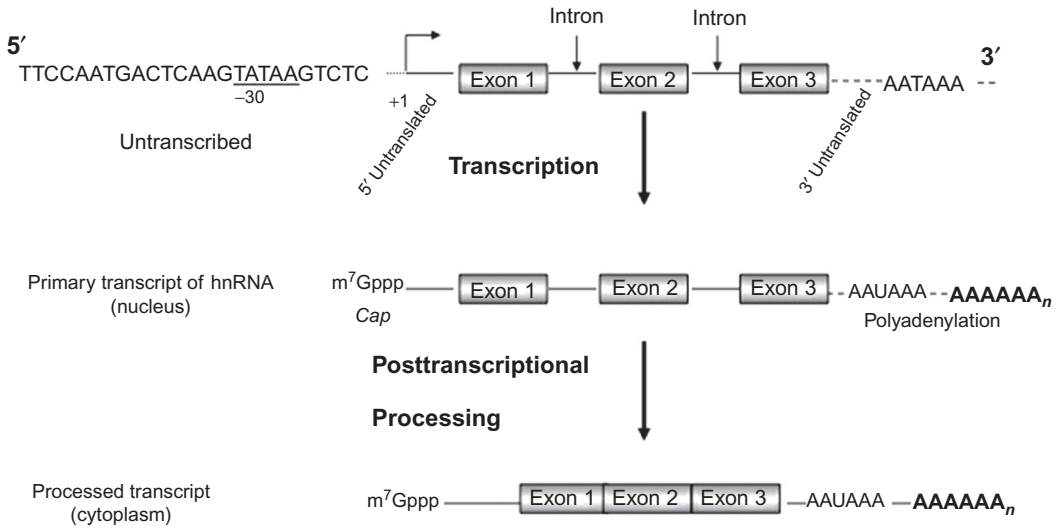


FIG. 1.3 Gene structure, transcription, and posttranscriptional processing. A gene is comprised of several hundred to several thousand bp, subdivided into functional elements. The locations of 5′ and 3′ untranslated sequences, exons, and introns are shown. The 5′ flanking sequences contain specific DNA elements (e.g., TATA box). RNA polymerase II transcribes DNA into heterogeneous nuclear RNA (hnRNA) during *transcription.* Twenty bp after the sequence AATAAA is transcribed to AAUAAA, mRNA is cleaved and the polyadenylate (poly(A)) tail is added to the 3′ end. A methylated guanylate residue is added to the 5′ end of the mRNA through a triphosphate linkage. Prior to exiting the nucleus, intron segments are removed by splicing factors during *posttranscriptional processing. (Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

only three bases (a triplet) to encode an amino acid. There are $4^3 = 64$ possible codons for 20 amino acids. In this way, the nucleotide code for proteins is considered "degenerate." The redundant genetic code protects against the deleterious effects of mutations as detailed in the next paragraph. In addition, two or three peptides can be encoded by overlapping codons simply by shifting the reading frame by 1 or 2 nt. Regulatory sequences that are transcribed but not translated reside at both the 5′ and 3′ ends of the mature RNA transcript. Both 5′ and 3′ untranslated regulatory sequences, which range from 10 to several thousand nucleotides, participate in the fidelity of translation and mRNA stabilization or destabilization.

The degeneracy of the genetic code (several codon triplets encoding one amino acid) is what makes some bp changes (mutations) within an exon exhibit no deleterious phenotype. The bp change is designated synonomous if the same amino acid is substituted (also known as a silent mutation) or nonsynonomous if a different amino acid is substituted. Strictly speaking, mutations mean that there has been a bp change whether or not the change affects the type of amino acid inserted into a peptide. Despite a nonsynonomous mutation in the coding sequence, the amino acid substitution might not exhibit a change in the physical characteristics (phenotype) of the organism nor render phenotypic advantages or disadvantages to the organism. Changes in the genetic code that put an organism at a disadvantage and contribute to disease are what we commonly call "mutations." BP changes in DNA that are neutral or impart a positive or negative advantage to the organism are also known as single nucleotide polymorphisms (SNPs). These SNPs can render subtle differences in the way an organism responds to its environment or other genetic influences (Fig. 1.4). SNPs are a focus of intense investigation due to their use in genome-wide scans to identify genes contributing to common multigene disorders, for example, diabetes, hypertension, etc.[11,12]

1.1.4 RNA Species

RNA molecules that encode proteins (except most histone proteins) are distinguished from ribosomal and transfer RNA by the series of adenosines added to the 3′ end of the molecule commonly referred to as the poly(A) RNA tail (Fig. 1.3). This feature is a useful means to isolate mRNA from more abundant RNA species (transfer and ribosomal RNA) and also designates the functional termination of the protein-encoding portion of the gene.[13] During transcription, the primary RNA transcript is cleaved 20 bp downstream of the AAUAAA site at the 3′ end, and ~150–200 adenine nucleotides are added to form the poly(A) tail.[14–16] The 5′ end of the mRNA transcript receives a protective "cap" after synthesis of the first 30 nt that consists of a guanylate residue methylated at the seventh position and linked to the first nucleotide of RNA by three phosphates. The RNA cap is a high-affinity binding site for ribosomes.[17,18] It should be noted that the element AATAA indicates the site of the poly A tail, but is not necessarily the functional end of the gene. Rather, the 3′ untranslated region (3′UTR) and 3′ untranscribed regions may still contain regulatory elements that modulate gene expression. In fact, most mRNAs bind sequences in the 3′UTR. Therefore, like the 5′ end of a gene, the 3′ end of the gene must be determined empirically.

Two classes of noncoding RNAs transcribed by Pol II have motivated the current expanded interest in RNA biology—mRNAs and long noncoding RNAs.[19] mRNAs (miRNAs) are a class of noncoding RNAs generated primarily from DNA sequences between genes (intergenic) within introns or at the 3′ end of the gene. They were originally identified in plants and worms as posttranscriptional regulators of gene silencing.[20–22] Pol II and sometimes Pol III transcribe DNA to produce primary miRNA transcripts.[23–25] In addition, transcription factors modulate the expression of these mRNAs as for protein-encoding genes. For instance,

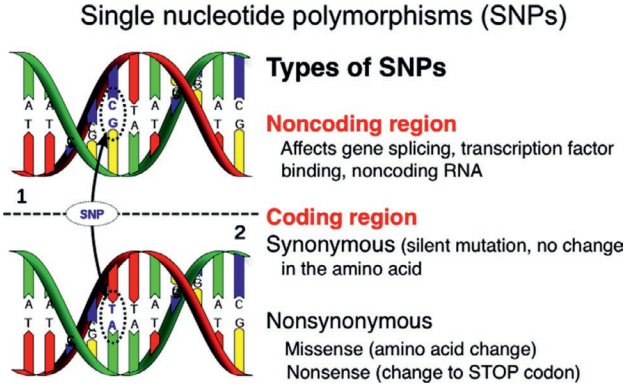


FIG. 1.4 Single nucleotide polymorphism (SNP). Schematic diagram of a SNP in which a protein encoding gene sequence differs between two individuals by one nucleotide.

extracellular signaling via typical signal transduction pathways and epigenetic mechanisms regulate the expression of mRNAs.[26] The gene product is RNA rather than protein and exerts its effect on its own locus as well as multiple loci due to their small size and less stringent binding requirements.[7,27] In this way, miRNAs are thought to regulate at least one-third of all human genes.

miRNAs are synthesized in the nucleus as a primary transcript (pri-miRNA) capable of forming several hairpin structures through internal complementarity (Fig. 1.5). The microprocessing complex containing a nuclear RNase III endonuclease called Drosha and the DiGeorge syndrome critical region 8 protein (DGCR8) cleaves the pri-miRNA transcript. The Drosha protein complex removes flanking segments and an ~11 bp stem region. This step converts the pri-miRNA to precursor miRNAs (pre-miRNAs). Pre-miRNAs are typically 60–70-nt long hairpin RNAs with 2-nt overhangs at the 3′ end. The nuclear export receptor exportin-5 and RanGTP transport the pre-miRNA into the cytoplasm where it is further processed by a complex containing another RNase III endonuclease called Dicer. Dicer partners with RNA-binding proteins to cleave the pre-miRNA into 21–25 nt duplexes. The miRNA/miRNA* duplex consists of a guide RNA strand and a passenger strand indicated by an asterisk (miRNA*) that is discarded upon assembly of the *R*NA-*i*nduced *s*ilencing *c*omplex (RISC). Loading the miRNA/miRNA* duplex into RISC is a four-step process requiring ATP hydrolysis and the major RISC protein component called Argonaute (Ago proteins). Upon unwinding of the duplex, the miRNA* strand is discarded leaving a single strand 21–25 nt RNA molecule available for silencing specific clusters of genes by hybridizing to their 3′UTRs. Ago protein coat miRNAs and along with exosomes protect miRNAs from degradation in biofluids such as blood and urine rendering them potential biomarkers.[28,29]

Long noncoding RNAs are nucleic acids that do not encode a protein and are at least 200-nt long or greater.[30] They are distinguished from miRNAs by their size (lncRNA >200 nt versus miRNAs ~22 nt) and the ability to exhibit more diverse functions. miRNAs typically suppress multiple gene targets, whereas lncRNAs typically regulate the gene from which they are transcribed, albeit by multiple mechanisms. The advent of whole genome sequencing has identified more noncoding transcripts than coding complicating our ability to define their function. lncRNAs can function in "cis" or "trans," can circularize or remain linear. Moreover, lncRNAs can function as protein scaffolds by recruiting regulatory complexes to genes, or behave as decoys, signaling molecules or as antisense interference transcripts. Therefore, through these diverse behaviors, lncRNAs exhibit pleomorphic functions such as genomic imprinting, chromosome shaping, and allosterically enzyme regulation. The function of most lncRNAs is unknown and thus the transcripts have simply been named numerically. Those lncRNAs that have been assigned a function include XIST (X chromosome inactivation), HOTAIR (Hox transcript antisense RNA), and TERC (telomerase elongation).

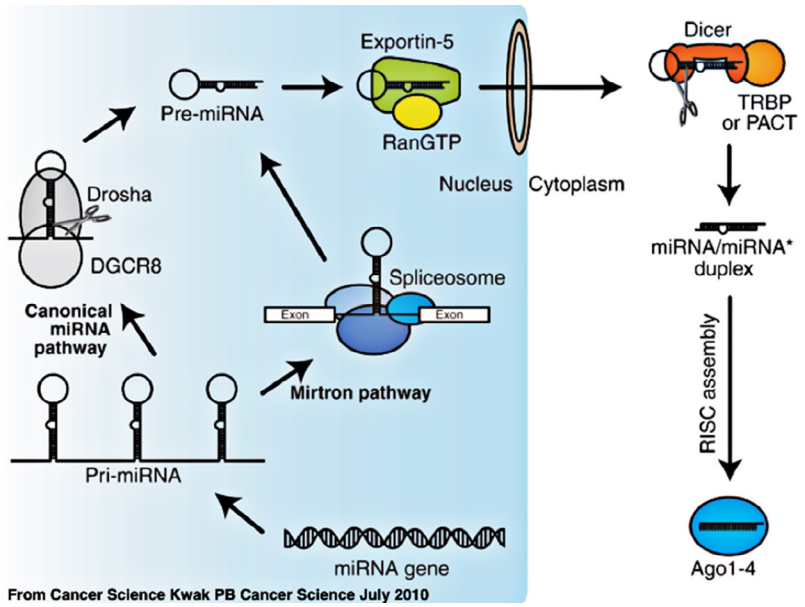


FIG. 1.5 Synthesis of microRNAs (miRNA). miRNAs are synthesized from the primary miRNA (pri-miRNA), which are then edited to the pre-miRNA. The RAN-GTP/Exportin 5 complex transports the Pre-RNA to the nucleus where the pre-miRNA is further processed to the miRNA/miRNA* duplex. *miRNA indicates the passenger strand that is discarded upon assembly of the *R*NA-*i*nduced *s*ilencing *c*omplex (RISC). The Argonaute (Ago) protein are the major protein component of the RISC. TRBP = TAR RNA-binding protein (aka PACT). *(Reproduced from Kwak PB, Iwasaki S, Tomari Y. The microRNA pathway and cancer.* Cancer Sci *2010;**101**(11):2309–15. doi: 10.1111/j.1349-7006.2010.01683.x)*

1.1.5 Linking Gene Structure to Function

Previously the 5′ border of a gene was identified by the promoter region (functionally determined) and by the first nucleotide transcribed into mRNA (cap site) determined empirically by various reverse transcriptase methods—for example, primer extension analysis or anchored polymerase chain reaction (PCR and DNAse1 hypersensitivity sites).[31] These techniques used reverse transcriptase to synthesize complementary or copy DNA (cDNA) (Fig. 1.6). Radiolabeled primers complementary to the 5′ end of the DNA sequence to be copied were allowed to anneal to mRNA. Reverse transcriptase then adds deoxynucleotides to the primer in the 3′ to 5′ direction. Synthesis of the cDNA terminates when the 5′ end of the mRNA is reached. Template mRNA molecules were removed by ribonucleases (RNases), and the synthesis of a double-stranded cDNA was completed through the action of DNA polymerase. Because the newly synthesized cDNA was radiolabeled at the 5′ end, the length of the cDNA (and hence the transcriptional start site) was determined by resolving the fragments on a denaturing polyacrylamide gel and comparing the length observed in bp to the known cDNA sequence.

In the age of whole genome analysis, the characterization of gene function has lagged behind the generation of transcript mapping. In other words, the biochemical assays such as DNase-seq, ATAC-seq (assay for transposase-accessible chromatin), ChIP-seq, and 3C (chromatin conformation capture) genome-based methods do not provide an assessment of function.[32–36] This has led to the development of high-throughput methods to identify changes in gene transcription levels (both coding and noncoding).[37] These include RNA-seq and STARR-seq (self-transcribing active regulatory region sequencing.[38] In addition, CRISPR/Cas9 methods of activating or silencing gene in situ have permitted the development of functional readouts for enhancer modification within its endogenous environment.[12,39,40]

We now know that these additional DNA sequences might encode noncoding RNA that regulates gene expression in addition to the well-described enhancer sequences. Specific DNA elements called insulator elements mark the boundary of genes.[41–44] These elements, originally identified on the globin gene, bind an 11-zinc finger transcription factor called CTCF, which is capable of blocking histone acetylation spreading between adjacent genes.[45,46] More recently, it is now understood that gene expression occurs in insulated neighborhoods generated by chromosomal loops formed by the binding site for CTCF and the cohesion complex.[47] Thus, enhancer or repressor sequences that are kilobases away from the transcriptional start site (TSS) can brought closer to the genes that they regulate by forming gene-enhancer/repressor "neighborhoods" called topologically associated domains (TADs).[48] It has recently been shown that CTCF-binding site mutations that prevent the formation of TADs can cause disease.[49]

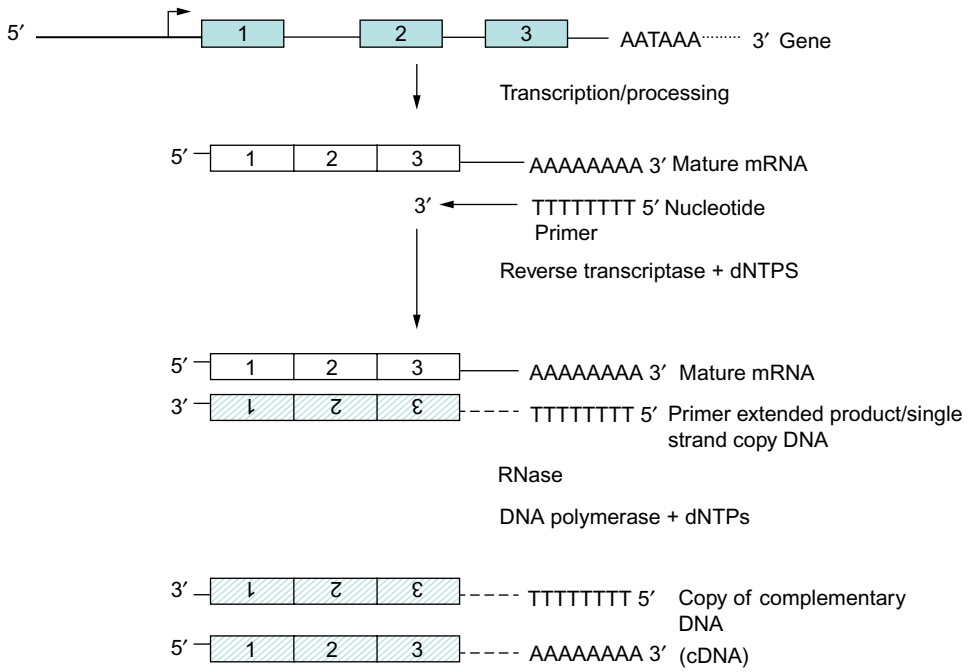


FIG. 1.6 Complementary DNA (cDNA). Primers complementary to a portion of the mRNA are allowed to anneal. For unknown sequences, as in the synthesis of cDNA libraries, a primer complementary to the poly (A) tail is used, i.e., poly (dT). Reverse transcriptase added along with all four deoxynucleotides (dNTPs) will transcribe mRNA in the 3′ to 5′ direction to make copy DNA. The mRNA template is removed by RNases, and double-stranded cDNA is made using DNA polymerase. In primer extension analysis, the 5′ end of mRNA (the cap site) is identified by annealing primers of a known sequence near the 5′ end of mRNA. *(Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

Given the requirement for larger and larger pieces of DNA to recapitulate native expression in transgenic mouse models, techniques have been developed to clone and manipulate large pieces of DNA (over 50 kilobases), for example, yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs).[50,51] Recombineering is a powerful technique performed in bacteria that permits the introduction of foreign DNA or point mutations into these large plasmids that are eventually introduced into transgenic mice,[52–54] but has been superceded by a powerful new technology called CRISPR-Cas9.[55,56]

CRISPR/Cas represents the latest and to date the most powerful breakthrough in our ability to modify or manipulate the genome with precision. The term CRISPR stands for clustered regularly interspaced short palindromic repeats and Cas is the abbreviation for CRISPR-associated protein. Cas9 is a nuclease that uses guide RNA to direct the enzyme to the specific DNA sequence to be modified by forming Watson-Crick base pairing. Thus, the technique is a simple, RNA-guided method by which bacteria and *Archaea* defend themselves from the DNA of invading bacteriophages (adaptive immune mechanism). In short, the technology originates from studying the bacterial immune system and consists of two parts: a DNA-binding domain that recognizes the sequence to be modified and an effector domain that mediates double-strand DNA breakage. These two steps activate the host cell's sequence-specific endonucleases to repair the break by nonhomologous recombination resulting in modification of the targeted sequence. The specificity of the technology lies in the ability to program the guide RNA. Prior to CRISPR/Cas, zinc-finger nucleases (ZFNs)[57,58] and transcription activator-like effector nucleases (TALENs) were the primary methods used to execute programmable genome editing.[59,60]

1.2. EPIGENETIC INFLUENCES

Epigenetics, literally means "outside of or beyond genetics," refers to the "study of genetic modifications that are mitotically and/or meiotically heritable yet do not change the DNA sequence".[61] Thus, mutations or deletions can alter the length of a gene that in turn alters the primary sequence of the protein. By contrast, epigenetic influences chemically modify the nucleotide or amino acid structure that in turn changes how that particular DNA or (histone) protein is recognized by nuclear proteins without changing the sequence itself. Although it is now clear from the completed sequence of the human genome that there are only about 20,500 gene loci,[2] the complexity of the genetic information encoded in human chromosomes must enlist other features of chromatin.[62] The epigenetic influences on chromatin appear to be one of the critical features that enhance genomic complexity. A major target of epigenetic changes is histones, basic proteins that coat the naked DNA double helix. The N-terminal tails of histones (H1, H2A, H2B, H3, H4) are positively charged due to the basic amino acid lysine. The positively charged histones attach to DNA because of the negatively charged phosphate groups comprising the DNA backbone. The ionic interaction is reduced if the positive charge on the lysines is removed. Specific enzymes called histone acetyltransferases (HATs) acetylate the lysine side group effectively eliminating the positive charge (Fig. 1.7). The loss of the ionic interaction between the histones and phosphate groups on DNA permit greater access to the DNA helix by accessory proteins such as polymerases, transcription factors, and coactivators or repressors. Chromatin becomes "open," accessible and readily transcribed. By contrast, there are enzymes called histone deacetylases (HDACs) that "close" chromatin by removing the acetyl groups from the lysines at the N-terminal tails of histone proteins. These enzymes are called histone deacetylates (HDACs). Removal of the acetyl group restores the positive charge to histones allowing the ionic interaction between histones and DNA to be restored. Consequently, nonhistone proteins such as polymerases and transcription factors become excluded from DNA, transcription is silenced, and chromatin becomes inactive.

Collectively, histones and accessory proteins associated noncovalently with DNA are what forms chromatin. Chromatin exists in two forms—euchromatin and heterochromatin.[63] Euchromatin contains actively transcribed genes that decondense during DNA replication. Euchromatin is also centrally located in the nucleus. By contrast, heterochromatin contains transcriptionally silent genes that remain condensed at the periphery of the nucleus. The DNA sequences within heterochromatin are repetitive and only 15% of nuclear chromatin is heterochromatin. The major forms of epigenetic modifications in mammalian cells occur on DNA and histones and include such covalent modifications as methylation and acetylation, but also the addition of other organic residues. The most common epigenetic change is DNA methylation. In addition, methylation is currently the only epigenetic change known to occur on DNA. By contrast, histone proteins undergo over 100 types of epigenetic modifications, of which the most common include acetylation, methylation, and phosphorylation.[64] Histones are frequently the target of changes, but nuclear regulatory proteins, for example, transcription factors can also be covalently modified, most commonly by phosphorylation. Epigenetic changes affect such events as chromatin folding, gene expression, X-chromosome inactivation, and genomic imprinting.[65,66] They are essential for development and differentiation in which clusters of genes must be activated or silenced at precisely timed intervals during an organism's growth and maturation. In addition, epigenetic changes provide mechanisms by which the environment affects the genome, for example, microbiota, immune disorders, and cancer.[67–70]

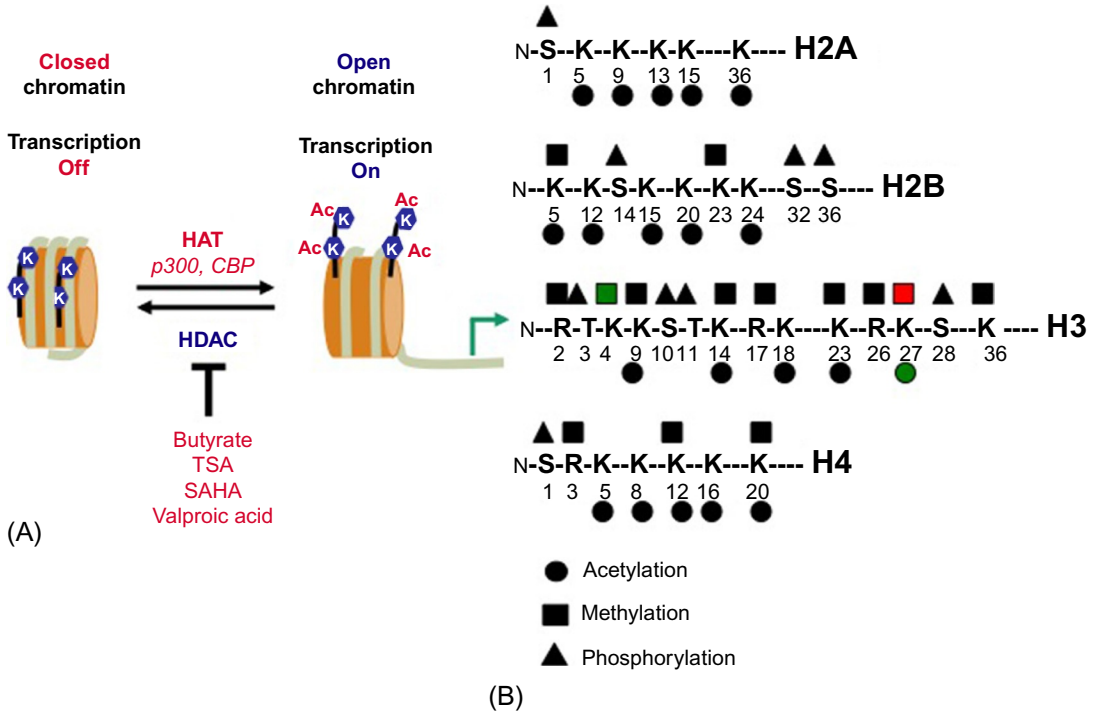


FIG. 1.7 Nucleosome structure and histone modifications on histone tails. (A) The double-strand DNA helix winds twice around a complex of the four core histones assembled as dimmers. Unacetylated histones are positively charged and adhere tightly to the negatively charged DNA preventing access by transcription regulatory proteins. Histones that are acetylated are less positively charged and do not adhere as tightly to chromatin allowing access of regulatory proteins to the DNA. The addition or removal of acetyl groups to the ends of histones is regulated by acetyltransferase (HATs) and deacetylase enzyme complexes (HDACs). The short chain fatty acid butyrate inhibits the activity of HDACs. (B) Shown are the amino-terminal histone residues modified by acetylation, methylation and phosphorylation. *(Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

1.2.1 DNA Methylation

DNA methylation is a postsynthesis modification that normal DNA undergoes after each replication. This modification is catalyzed by DNA methyltransferases (DNMTs) and occurs on the C-5 position of cytosine residues within CpG dinucleotides located primarily in the promoter of a gene. There are three major DNMTs (DMNT1, 3A, 3B). Each DNMT plays a distinct and critical role in cells. Murine knockouts of DNMT1 and DNMT3b exhibit embryonic lethality.[71] The DNMT3a homozygous mouse appeared normal at birth but died by 4 weeks of age.[71,72] In humans, mutations of DNMT3b are linked to ICF syndrome (*I*mmunodeficiency, *C*entromere instability, *F*acial anomalies).[71,73] DNMTI functions as the "maintenance" methyltransferase since it functions during cell division to methylate the newly synthesized DNA strand as dictated by the hemi-methylated complementary strand.[74] DNMT3a plays a central role in the methylation of neural specific genes.[75,76]

Sixty percent of human genes contain a CpG island.[77] While methylation can also occur in other parts of the gene, CpG dinucleotides tend to be underrepresented in the genome and when they are found appear in clusters ranging from 0.5 to several kilobases with GC content greater than 55%.[78] About 15% of CpG dinucleotides cluster in short DNA segments known as CpG islands.[79] The remaining 85% of the islands are spread throughout the genome in repetitive hypermethylated segments that are transcriptionally silent.[80] Methylation of "CpG islands" is a late evolutionary development and functions to maintain genome stability by repressing transposons and repetitive DNA elements.[81]

DNA methylation is an important event in many processes, including transcriptional repression, X chromosome inactivation and genomic imprinting. CpG islands locate in the promoter region of genes about 60% of the time[78,82,83] and are normally hypomethylated particularly in the germ cells. Collectively, these CpG clusters or islands cover only about 0.7% of the entire genome, which is still equivalent to several million nucleotides. Hypermethylation at CpG islands induces transcriptional silencing that in turn is stably inherited. Thus as cells differentiate, a significant percentage of these CpG islands become methylated in a tissue specific manner. Typically these would be genes involved in cell renewal. As observed with HDACs and deacetylation, the methylation status of cancers might seem contradictory. Yet, aberrant de novo hypermethylation of CpG islands is a hallmark of some human cancers and occurs early during carcinogenesis.[84–86] Tumor suppressor genes are locally hypermethylated by some cancers to silence their expression; whereas, oncogenes might be hypomethylated.[78]

The DNA of tumor cells is globally hypomethylated, a process that is linked to nutritional status, for example, B_{12} or cobalamin absorption.[87] Cobalamin is required for the synthesis of *S*-adenosylmethionine, the primary methyl donor in the cell.[88] In this way, reduced cobalamin absorption as sometimes observed in Crohn's or pernicious anemia would provide an environment favorable to cancer.[89] Niacin required to form NAD, which is necessary for ADP-ribosylation of histones, also affects chromatin structure.[90]

The most precise approach to assessing DNA methylcytosines is through bisulfite sequencing. Treating DNA with sodium bisulfite converts unmethylated cytosines to uracil that when subjected to conventional DNA sequencing are read as thymines. Methylated cytosines are still read as cytosines. Although bisulfite sequencing is not as easy to scale up as a genome-wide analysis by methylation-sensitive restriction enzyme (MSRE) analysis, sequencing is the most accurate way to determine the methylated sites in DNA or the methylome.[91]

Genomic imprinting occurs in gametogenesis and is necessary for development. One of the X chromosomes in females is not expressed due to the heavy methylation of the inactive X chromosome. The epigenetic phenomenon whereby expression of a gene depends on whether it is inherited from the mother or the father is called imprinting and is due to differential methylation of specific cytosine bases on the maternal versus the paternal genes.[92] Recent genome-wide analysis of genomic imprinting in the mouse identified 1300 loci that exhibit parental bias in the expression of specific mRNA transcripts. The gene loci identified control neural systems associated with feeding and behavior.[93] In addition, the authors in a separate article showed preferential selection of the X chromosome inherited from the mother as opposed to the one from the father in glutamatergic neurons of the female cortex.[94] The interleukin-18 gene was identified as an important locus controlling sex-specific preferences.

1.2.2 Histone Modifications

The basic repeating unit of chromatin is the nucleosome. Each nucleosome is composed of 147 bps of DNA wrapped twice around a histone protein octamer consisting of two molecules of each of the four core histones (H2A, H2B, H3, and H4). The linker histone H1 sits alone between each core nucleosome facilitating further compaction.[95] Each histone contains a structured globular domain with a histone-fold motif important for nucleosome assembly, and a highly charged unstructured amino-terminal tail of 25–40 residues, which protrudes from the body of the nucleosome to latch onto the phosphate backbone. The amino-termini are the major sites for histone modifications.[96] Histones can be modified by acetylation, methylation, phosphorylation, ADP-ribosylation, ubiquitination, and sumoylation (Table 1.1).[97] The mixture of these covalent modifications create a "code" on the surface of the histone molecule that is subsequently recognized by a class of chromatin-binding proteins, for example, bromo- and chromodomain-containing

TABLE 1.1 Enzymes, Targets, and Effect of Epigenetic Modifications

Target Covalently Modified	Group	Adds	Removes	Effect on Gene Expression[a]	Enzyme Inhibitors
DNA	Methyl	DNMT	Gadd45	↑ increases or ↓ decreases	Azacytadine; RG-108
Histone	Acetyl	(KAT)/HAT	HDACs	↑ increases or ↓ decreases	Butyrate, SAHA, trichostatin A, valproic acid
Add to lysines (K)	Methyl	KMT (SETs, PCG1, 2, TrG)	KDM Jumonji (JMjC, Jarid)	↑ if H3K4me3; H3K36me3; H3K79me3 ↓ if H3K9me2,3; H3K27me3	BIX-01294
Add to arginines (R)	Methyl	PRMTs (CARM1, PRMT1)	PADI4	?	
Add to S10H3	Phosphate	AurB	PP1	↑	
Add to lysines (K)	Ubiquitin 76 aa peptide	Ub ligases (Ring 2)	Ub protease (USP)	↑	
Add to lysines (K)	Sumo = small ubiquitin-like modifiers, ~76 aa	Ubc9	Ub protease (SUSP)	↑	

[a] *"?" means unknown.*

proteins that mediate chromatin compaction, transcription, and DNA repair.[98] Acetylation, methylation, ubiquitination, and sumoylation occur on the lysine residues while methylation also occurs on arginine residues.[99] Phosphorylation occurs on serines and threonines, ADP-ribosylation on glutamic acids. Most of these modifications, particularly acetylation, alters the charge distribution on the amino-terminus and alters nucleosome structure, which can in turn regulate chromatin structure.[100,101] Some covalent modifications act as molecular switches, enabling or disabling subsequent covalent modifications, which explains the functional complexity of epigenetic modifications.[102] Each modification correlates with a specific physical status of chromatin. The next several sections will highlight the most common histone modifications.

1.2.2.1 Histone Acetylation

Acetylation of histones occurs at the ε-amino side group of specific lysines within the N-termini of histones. HATs transfer an acetyl group from the donor acetyl-CoA to the histone terminal lysines.[103] In hypoacetylated chromatin, the positive charges on unacetylated lysines are attracted to the negatively charged DNA, producing compact, closed chromatin, which represses transcription.[104] By contrast, acetylation of the lysines removes their positive charges resulting in a less compact, open chromatin structure, which facilitates gene transcription. Therefore, HAT activity and subsequently histone acetylation are linked mainly to transcriptional activation[105] (Fig.1.7). Removal of the acetyl group (deacetylation) by HDACs restores the positive charge on lysines, chromatin becomes compacted and less accessible to regulatory proteins required for transcription. Thus, HDACs and deacetylation are primarily associated with transcriptional repression (Fig. 1.7).

The HATs are divided into five families. These include the p300/CBP HATs (p300 and CBP), *Gcn*5-related *a*cetyl-*t*ransferases (GNATs, including Gcn5, PCAF, etc.), MYST (*MOZ*, *Y*bf2/Sas3, *S*as2 and *T*ip60)-related HATs, the general transcription factor HATs (TFIID subunit TAF250 and TFIIIC), and the nuclear hormone-related HATs (SRC1 and ACTR).[106] The most consistent functional characteristic of HATs is that they are transcriptional coactivators. These proteins are components of large multisubunit complexes that do not bind DNA directly, but instead form protein-protein interactions with DNA-binding transcription factors.[107] The MYST proteins are the largest family of acetyltransferases.[108] More recently, the Gcn5-related acetyltransferases are considered to be part of a complex called SAGA for *S*pt-*A*da-*G*cn5-*A*cetyltransferase.[109] SAGA preferentially acetylates several N-terminal lysines within H3 and H2B in response to cellular stress, for example, low glucose, hypoxia, and UV damage.[110] Moreover, in addition to its HAT activity, SAGA also has deubiquitinase activity.[110] In summary, the themes that are consistently emerging are first that these histone-modifying enzymes are components of large complexes and second for every enzymatic complex that adds an organic residue to histones, there is a complementary enzymatic complex that can remove them (Table 1.1).

The more numerous mammalian HDACs have been grouped into three protein classes.[111] Class I includes HDACs 1, 2, 3, and 8; class IIA includes HDACs 4, 5, and 7; class IIB includes HDACs 6 and 10; and class IV is comprised of HDAC 11. HDACs 1–11 are zinc-dependent.[112] The class III HDAC family consists of the conserved nicotinamide adenine dinucleotide (NAD)-dependent Sir2 family of deacetylases or sirtuins of which there are 7. The sirtuins are not zinc dependent. Like HATs, HDACs do not bind directly to DNA, but are recruited to genes by large multisubunit complexes to function primarily as corepressors of transcription.[113]

The function of HATs and HDACs is of particular relevance in the GI tract due to the effect of butyrate, a by-product of colonic bacterial fermentation, on histone acetylation (Fig. 1.7). Epidemiologic studies uniformly concur that a diet high in fiber is protective against colon cancer.[114] The short-chain fatty acid butyrate is one of several fiber-derived fermentation products capable of maintaining epithelial cell differentiation.[115] The differentiation effects were initially revealed after treatment of erythroleukemic cells with butyrate.[116] Subsequently, it was discovered that the induction of differentiation by butyrate correlated with histone hyperacetylation[117–119] due to suppression of HDACs.[120–124] Thus, the HDAC inhibitory effects of butyrate and resulting histone hyperacetylation might, in fact, be one mechanism by which dietary fiber exerts its anticancer effects.[125,126] While butyrate is normally used by colonocytes as a carbon source under low glucose conditions, colon cancers use the Warburg effect when glucose is in abundance to generate ATP via glycolysis.[127] The butyrate that is not converted by fatty acid oxidation in the mitochondria to produce ATP is taken up by the nucleus where it suppresses HDACs.[128] Thus, the HDAC inhibitory effect of butyrate depends on the metabolic state of the cell.[127]

Most reviews support the viewpoint that butyrate and HDAC inhibitors are potent anticancer agents.[129–133] Collectively, early studies emphasized the global effects of butyrate on chromatin remodeling, but the molecular basis for the gene-specific effects of butyrate remains poorly defined. HDAC inhibitors regulate less than 10% of actively transcribed genes. Most of those are upregulated through GC-rich sites.[134,135] In addition to histone acetylation, it is now known that DNA-binding proteins can become acetylated.[106] Thus, a possible mechanism by which hyperacetylation induced by butyrate might target specific genes is through acetylation of specific transcription factors. The proposed function of acetylated transcription factors varies and includes increased or decreased DNA binding as well

as protein stability.[136] In many instances, the genetic targets of butyrate are GC-rich sequences that bind Sp1 and Sp3.[137–139] Gamma glutamyl transferase,[140] IGF-binding protein 3,[141] G alpha (i2),[142] galectin,[143] Cox 1,[144] and intestinal alkaline phosphatase[145] are all upregulated by butyrate through Sp1 sites. Sp1-binding sites are also implicated in the butyrate induction of $p21^{WAF1}$ gene expression.[146] HAT p300 recruited to the $p21^{WAF1}$ promoter cooperates with Sp1 and Sp3 to mediate the effects of butyrate.[147] However, Sp1 does not cooperate directly with p300, but instead binds the histone deacetylase HDAC1.[148,149] The Sp1-HDAC1 complex in turn forms complexes with other corepressors such as Sin3A.[150] Thus, Sp1 appears to be the factor that confers $p21^{WAF1}$ promoter repression by recruiting HDAC4 and corepressor complexes.[151]

HDACs have opposing functions especially in cancer. On the one hand, HDACs can prevent the activation of tumor suppressor genes and block the ability of a cancer cell to undergo apoptosis.[152] However, HDAC2 silencing triggers apoptosis.[153] Another important feature of HDACs is their interaction with DNA methylation. HDACs cooperate with DNMTs by removing the acetyl groups blocking methylation targets on histones or DNA.[154–156]

1.2.2.2 Histone Methylation

There are two types of histone methylation, targeting either lysine or arginine residues. Histone methyltransferases (HMTs) perform these modifications utilizing *S*-adenosylmethioine as the methyl group donor. Lysine methylation is implicated in changes in chromatin structure and gene regulation; whereas, arginine methylation correlates with the active state of transcription, like acetylation.[157]

1.2.2.2.1 Histone Methylation at Lysines

Methylation of lysine residues (K) occurs on histone H3 primarily at K4, K9, and K27 and on H4 at K20 (Fig. 1.7B). The lysine residue can be mono-, di-, or trimethylated at the episilon amino group. Methylation of H3 especially on lysine 4 and 36 (H3K4 and H3K36) is associated with an open chromatin configuration and transcribed chromatin.[83,158,159] In contrast, the methylation of H3 at K9, K27, and H4K20 is associated with condensed, repressed chromatin.[160] Thus, the overall effect of histone methylation on gene expression depends on which lysine is methylated and to what degree (mono-, di-, trimethylated).

In general, there are at least four families of lysine methyltransferases. All of the lysine methyltransferases (KMT or HMT for histone methyltransferases) are distinguished by the presence of SET domains. One family of these methyltransferases is further distinguished by the presence of an additional protein domain separate from the SET domain and will be discussed further in Section 1.2.3 on chromatin-binding proteins. SET protein domains are approximately 130 residues homologous to amino acid segments in Su(var)3-9, Enhancer of Zeste and Trithorax, three *Drosophila* proteins with intrinsic methyltransferase activity.[161,162] The mammalian form of the prototypical lysine methyltransferase (or KMT) Su(var)3-9 is Suv39h and is involved in stabilizing heterochromatin by trimethylation of histone H3 at lysine K9. Histone methylation at K9 is recognized by a subgroup of E2F-related transcription factors called HP1α, β, or γ.[163] These HP1 proteins use chromodomains to recognize the trimethylated atomic feature or imprint on H3.[164] The methylated or acetylated imprints on DNA or two classes of proteins—those with chromodomains that recognize methyl group imprints, and those with bromodomains that recognize acetyl group imprints. Transcriptional coactivators such as CBP, p300, and PCAF are HATs that contain bromodomains. They acetylate histones and other nuclear proteins and so not surprisingly also recognize an acetyl group imprint. These proteins are discussed in greater detail below under Section 1.2.3. So, in summary, the initial and prototypical KMT protein family are the SET domain-containing proteins, which target H3K9 and are recognized by HP1 factors.

Proteins involved in histone demethylation underscore the fact that like acetylation, protein methylation is a dynamic process. The Jumonji domain-containing proteins demethylate histone lysines. The Jumonji C protein family (JmjC) catalyzes the removal of methyl groups from lysines while the Jumonji D family (JmjD) removes methyl residues from arginines.[165,166] In addition, JARID2 (a Jumonji C and ARID-domain-containing protein) catalyzes the removal of methyl groups from H3K4me3 and H3K4me2 and can function as corepressors[167] or balances stem cell self-renewal with differentiation by affecting methylation at H3K27.[168] It has been recently shown that JARID2 is a component of the PRC2 and mediates transcriptional repression by recruiting PCR2, an H3K27 methylase, to gene promoters.[169–171]

1.2.2.2.2 Histone Methylation at Arginines

Methylation at arginines occurs within the tails of histone H3 (R2, R17, and R26) and H4 (R3) and is catalyzed by CARM1 (coactivator-associated arginine methyltransferase 1) and PRMT1 (protein arginine *N*-methyltransferase 1 (PRMT1), respectively, in mammalian cells (Fig. 1.7). Like lysines, arginines can be either mono- or dimethylated (asymmetric or symmetric additions) on the guanidino nitrogen and this process is antagonized by human PADI4 (peptidylarginine deiminase 4), which converts methyl-Arg to citrulline.[172,173] Less is known about the fate of histones methylated at arginines. However, initial studies indicate that the methylated arginines create an imprint recognized by coregulatory molecules, for example, p300 and SWI/SNF.[174,175] CARM1 has been shown to inhibit alveolar cell proliferation and promote differentiation.[176]

1.2.2.3 Histone Phosphorylation

Histone phosphorylation occurs on all four core histones: H2A (S1), H2B (S14), H3 (S10 and S28), and H4 (S1)[177,178] (Fig. 1.7A). The phosphorylation of S10 in H3 is associated with transcriptional activation[179] and chromosome condensation during mitosis.[180] In addition, phosphorylation of S10 in H3 is also associated with the transduction of external signals to chromatin leading to the transient expression of immediate-early (IE) genes.[181,182] The phosphorylation of H3 is mediated by several specific kinases, activated by distinct pathways. For example, mammalian mitotic H3 phosphorylation is associated with Aurora B/IPL kinase.[183–185] H3 phosphorylation by IKKα is important for the activation of NF-κB,[186] and the IE gene response is mediated mainly by mitogen and stress-activated kinases MSK1 and MSK2.[187] Histone H2B phosphorylation condenses chromatin and is involved in apoptosis.[188,189] The downstream effect of H2A phosphorylation by Bub1 kinase is apparently required for chromosome stability.[190] By contrast, the effect of H4 phosphorylation is unknown.

Most of the covalent modifications of histones are known to be reversible. Consequently, if the presence of a modification influences transcription in a particular way, its removal might have the opposite effect. In this way, the cell could effectively respond to changes in environmental cues. Moreover, some histone modifications are mechanistically linked. For example, phosphorylation of S10 on H3 enhances histone acetylation by Gcn5 (part of the SAGA complex),[191,192] while H3 K9 methylation inhibits phosphorylation at H3 S10.[161] Given the number of sites and the variety of possible covalent modifications, the combinatorial possibilities are extremely large. The combinatorial pattern of N-terminal modifications results in a heterogeneous identity for each nucleosome that the cell interprets as a readable code from the genome to the cellular machinery directing various processes to occur. This concept is commonly referred to as the "histone code hypothesis".[98,193] The precise modification status of a specific histone tail on a given gene can also change during the process of transcriptional regulation and each of these different combinations of histone modifications may elicit distinct downstream transcriptional signals.[98,177]

1.2.3 Chromatin-Binding Proteins

The remaining histone methyltransferase families also recognize methyl groups on regulatory proteins other than histones and therefore are discussed here. The second group of SET domain methyltransferase proteins is related to the *Drosophila* protein Enhancer of Zeste, with the prototypical mammalian protein named Ezh2. Ezh2 is part of a complex of proteins called the Polycomb group (PcG). The two variations of these complexes have been designated Polycomb repression complexes 1 (PRC1) and 2 (PRC2). Ezh2 belongs to the PRC2 complex that also includes Eed, Suz12; whereas, PRC1 includes the Ring finger proteins (Ring1a,b, Rnf, Hpc, Edr, and Bmi1). Conditional deletion of Eed in the intestinal crypt resulted in crypt degeneration.[194] Therefore, the PRC2 complex is required for normal stem cell maintenance. Ezh2 has recently garnered significant attention due to its overexpression and therefore oncogeneic function in several epithelial cancers including prostate and breast,[195–197] compared to its tumor suppressor function in some hematopoietic cancers.[198,199] Consistent with its oncogenic role in epithelial cancers, Ezh2 is also overexpressed in colon, gastric, and liver cancer.[200–202] A genome-wide analysis of prostate cancers recently revealed an androgen-dependent fusion protein called TMPRSS2-ERG that in Chip-Seq analysis was found to transcriptionally target Ezh2.[203] Bmi1 has received increased attention because it is an important marker of normal and cancerous hematopoeitic stem cells.[204–206] Bmi1 is also associated with the +4 reserve stem cell in the intestinal crypt zone.[207–209] In addition, the PRC1 complex contains proteins that have E3 ubiquitin ligase activity.[210,211] The Polycomb group of proteins with their SET domains not only participates in histone lysine methylation, but both PRC1 and 2 complexes are also important in recognizing the methylated protein imprint.

The human homolog of the *Drosophila* Trithorax (Trx) protein is the mixed leukemia gene (MLL1). There are four human MLL homologs. MLL1 has been shown to be a specific methyltransferase of H3 at K4.[212] It in turn forms protein-protein interactions with coactivators, for example, CBP and corepressor chromatin remodelers, for example, SWI/SNF.[213,214] Other Trithorax homologs, for example, Ash1, Trx, form complexes with different coregulatory complexes. Collectively, the Trithorax group (TrG) of proteins can either activate or repress transcription depending on the coregulator with which it associates. Nevertheless, the TrG proteins characteristically oppose the activity of the Polycomb group (PcG).[215] The tumor suppressor protein menin (positionally cloned gene product of the MEN type 1 locus) interacts with MLL1 and normally induces the cyclin-dependent inhibitor $p27^{Kip}$.[216,217]

RIZ (retinoblastoma protein-interacting zinc finger protein), SMYD3, and MDS-EVI1 form a fourth family of SET-domain proteins because they have two isoforms that exhibit opposing functions. The isoform containing the SET domain has tumor suppressor function while the isoform missing the SET domain is cancer promoting. This "ying-yang" theory put forth by Huang is especially true for RIZ and MDS-EVI1 in which the cancer by an unclear mechanism disturbs the normal ratio between the two isoforms.[89] The SMYD3 protein contains another DNA-binding domain called MYND in addition to a SET domain and is overexpressed in colorectal and hepatocellular carcinomas.[218]

Crosstalk between DNA methylation and the histone modifications exists.[219] These interactions were revealed by the observation that HDAC1 forms a complex with DNMT1 and 5-methyl-cytosine-binding protein (MBP) on a methylated promoter to silence gene expression.[220] Similar crosstalk occurs between HDACs, Suv39, and HP1; HDACs, PRC2, and PRC1; HATs, MLL1, and BRM.[101] The enzymes that epigenetically modify the genome are categorized in terms of those that "erase" chromatin marks; add chromatin marks ("writers") or "read" chromatin marks (Table 1.2).

1.2.4 Epigenetics and Development

The epigenetic control of gene expression is a fundamental feature of mammalian development, as indicated by developmental arrest or abnormalities in methylation or acetylation-deficient mutants.[221] X-chromosome inactivation is an example of sequence-identical alleles being stably maintained in different functional states. In humans, X-linked inactivation serves to normalize the level of expression of X-linked genes in females (XX) and males (XY). Mutations in genes that affect global epigenetic profiles can give rise to human diseases. For example, the Fragile X syndrome results when a CGG repeat in the FMR1 (fragile X mental retardation gene 1) 5′ regulatory region expands and becomes methylated de novo, causing the gene to be silenced and creating a visible "fragile" site on the X chromosome under certain conditions.[222] On a more global level, mutations in the DNMT3b (which regulates DNA methylation) gene lead to ICF syndrome[71,73] and CBP (with acetyltransferases activity) mutations cause RSTS (Rubinstein-Taybi syndrome).[223,224] Discovered in 2004, lysine demethylases (LSDs) appear to play an essential role in stem cell pluripotency versus lineage specification.[225]

1.2.5 Epigenetics and Cancer

Epigenetic changes play an important role in tumorigenesis. The major epigenetic changes that take place during cancer development are generally the aberrant DNA methylation of tumor suppressor genes and histones. Genomic methylation patterns are frequently altered in tumor cells, with global hypomethylation accompanying region-specific hypermethylation events. When hypermethylation events occur within the promoter of a tumor suppressor gene, this can silence expression of the associated gene and provide the cell with a growth advantage in a manner similar to deletions or mutations. Although cancer cells are hypomethylated in the genome compared to normal tissues, many tumor-suppressor genes are silenced in tumor cells due to hypermethylation. This aberrant methylation event occurs early in tumor development and increases progressively, eventually leading to the malignant phenotype. For example, a high percentage of patients with sporadic colorectal cancers with a microsatellite instability phenotype show methylation and silencing of the gene-encoding MLH1 (MutL protein homolog 1).[226] Other methylated tumor suppressors loci include *CDKN2A* ($p16^{INK}$), *p14ARF*, *Rb*, *E-cadherin*, and *BRCA1*. Deregulation of genomic imprinting can also play a role in cancer development, as exemplified by loss of *IGF2* gene imprinting in Wilms' tumor.[227]

Chromatin remodeling also plays an important role during tumorigenesis. Loss or misdirection of HATs has been linked to embryonic aberrations in mice[228,229] and to human cancers.[230,231] Misdirection of HAT activities as a result of chromosomal translocations is associated with multiple human leukemias.[232–234] In acute promyelocytic leukemia, the oncogenic fusion protein PML-RARα (promyelocytic leukemia-retinoic acid receptor-α) recruits an HDAC to repress genes essential for the differentiation of hematopoietic cells.[235] Similarly, in acute myeloid leukemia, AML1-ETO fusions recruit the repressive N-CoR-Sin3-HDAC1 complex that in turn inhibits normal myeloid development.[236]

More recently, noncoding RNAs transcribed from intervening (intronic) sequences have been linked to epigenetic changes cell cycle regulation, immune surveillance and cancer.[237] These large intervening noncoding RNAs (originally called lincRNAs, currently called lncRNAs) in some instances redirect the repressive polycomb repressor complex 2 (PRC2) to genes that promote cell renewal.[238] In particular, the lncRNA called HOTAIR is overexpressed in breast cancer and redirects the PRC2 complex, which methylates H3K27, an epigenetic change that tends to condense chromatin.[239,240] Perhaps not surprising, many epigenetic marks target the developmentally relevant homeobox class of transcription factors (Hox genes),[241] which in turn are master regulators of embryonic development and stem cell pluripotency that when altered can lead to disease.[242]

TABLE 1.2 Erasers, Writers, Readers

Categories	Erasers	Writers	Readers
	HDACs	HATs	Bromodomain
	NURD	PRC1, PRC2	Chromodomain
		PARP	

The fact that many human diseases, including cancer, have an epigenetic etiology has encouraged the development of a new therapeutic option called "epigenetic therapy".[64] Many agents have been discovered that alter methylation patterns on DNA or the modification of histones, and several of these agents are currently being tested in clinical trials.[243]

1.3. ANATOMY OF A GENE PROMOTER

The major advances in the area of transcriptional initiation since the prior edition of this textbook have occurred primarily in the explosion of information on epigenetic modifications.[244,245] The prototypical epigenetic therapies include the use of demethylating agents, for example, 5-azacytidine, for myelodysplastic disorders and histone deacetylase inhibitors, for example, SAHA, to treat a number of epithelial cancers.[246–249] A major focus of transcriptional elongation has been the role of the enzymes involved in epigenetic changes, for example, the HAT Gcn5.[250] Moreover, the trithorax group of epigenetic factors, specifically MLL1, forms a lysine methyltransferase complex with the elongation factor ELL.[251] Thus, the following section will briefly summarize the historical basis of gene promoter structure and transcriptional initiation. For more details, the reader is referred to recent reviews[12] and the prior edition of this chapter.

1.3.1 DNA Elements

RNA polymerase II (Pol II) and its accessory factors bind to a DNA sequence called the *promoter* located upstream of protein-coding sequences to direct RNA transcription.[252,253] Without the promoter, the genetic sequences that encode the information to make a functional peptide product will not be transcribed. Other 5′ flanking sequences or DNA elements that participate in transcription are sequence-specific binding sites for proteins that regulate the fidelity, rate, and timing of Pol II binding, formation of the preinitiation complex, and initiation of transcript elongation under basal and regulated conditions.[254–256] These sequences are defined as *cis-acting* elements because they are a part of the same (cis) gene.[257–260] DNA elements are categorized according to their ability to regulate transcription as a function of their distance and orientation from the promoter. Sequences that are contained within the first 30–100 bp of the promoter are considered promoter-dependent cis-acting elements. If they are *positive-acting* elements and increase the rate of transcription, they are considered *activating DNA elements*, whereas if they are *negative-acting* DNA elements and decrease or repress the rate of transcription, they are considered *repressor or silencer elements.*[261–263]

The RNA core promoter consists of two types—focused and dispersed.[264] Focused promoters contain either one or a tight cluster of start sites over a few bp; whereas, a dispersed promoter contains several start sites over about 100 bp and are typically found at CpG dinucleotide sites. Critical promoter elements include TATA elements, which lie upstream of the transcription start site, the initiator sequence (Inr) that spans the start site, upstream regulatory elements that bind either transcriptional activators or repressors and finally downstream poly(dA-dT) elements.[265] The TATA element or "TATA box" is an element whose DNA sequence is TATA or variants thereof.[258,266–269] This sequence resides at a fixed distance 25–30 bp upstream from the transcriptional start site in many Pol II promoters, and its location relative to the start site is position- and distance dependent.[270–272] However, many genes do not have TATA sequences. These "TATA-less promoters" still remain dependent on the TATA-binding protein (TBP) to assemble at the promoter to form the preinitiation complex (PIC) but the recruitment of TBP is not rate limiting.

Initiator elements (Inr) although initially identified at the "TATA-less promoters"[273,274] have subsequently been found at both TATA-containing and TATA-less promoters. Their role appears to be in directing the accuracy of Pol II initiation.[275] These Inr elements reside within the first 60 bp of the transcriptional start site, directly overlap the start site itself, but do not have a clearly defined consensus sequence.[276] Many of the genes-encoding gastrointestinal peptides (e.g., gastrin, somatostatin, cholecystokinin, glucagons, and secretin) contain TATA elements[277–281]; however, the gene-encoding the growth factor, transforming growth factor alpha (TGFα), does not.[282]

1.4. METHODOLOGY

This section summarizes some of the molecular techniques used to study the transcriptional control of genes. These methods are used to study either genetic structure or function. Three systems have been used to study function: reconstituted cell-free transcription assays, cell and tissue culture models, and whole-animal studies. Methods that analyze the structural interactions include those techniques that assess DNA-protein interactions and those that assess protein-protein interactions. More recently, studies of noncoding RNAs involve understanding RNA-DNA, RNA-RNA and RNA-protein interactions.

1.4.1 Structural Methods

Once functional regulatory DNA elements have been identified, assays that assess DNA-protein interactions are performed.[283] Indeed, in circumstances where a long sequence (>50 bp) must be analyzed, it is simpler to identify DNA-protein interactions first and then determine if these DNA elements are involved in transcriptional regulation. DNase I footprinting assays are used to identify DNA-binding

elements that interact with crude or purified nuclear proteins by protecting them from chemical or enzymatic cleavage.[284,285] Such assays are particularly well suited for studying cooperative interactions among proteins bound to adjacent DNA elements. The technique can be carried out in vivo or in vitro.[31] However, in vivo footprinting has been superceded by chromatin immunoprecipitation (ChIP) assays described below. Electrophoretic gel mobility shift assays (EMSA, gel shift, gel delay, or band shift assays) permit a more detailed analysis of (a) the type of protein complexes that bind to individual DNA elements and (b) the specificity of the protein interaction with specific bp[286–288] (Fig. 1.8). This assay is also rapid and easier to perform than footprinting assays. Methylation interference assays extend the power of the gel shift assay by identifying specific nucleotide contacts that are required for DNA binding.[289] DNA affinity precipitation (DAPA) is a DNA-protein interaction assay that uses a biotinylated DNA-binding site to identify the proteins that are recruited to an element.[290] The assay uses the DNA element to isolate the protein factors along with immunoblots to identify the proteins that form both the protein-DNA and protein-protein interactions. Southwestern blot analysis takes advantage of specific DNA elements that are used to detect nuclear proteins separated on a denaturing gel and transferred to nitrocellulose or produced by a phage expression library.[291–293]

1.4.1.1 Chromatin Immunoprecipitation Assays

ChIP analysis is now the most effective method to document an in vivo interaction at DNA.[294–296] First, a fixative, usually formaldehyde, is used to crosslink proteins to DNA. Antibodies are then used to immunoprecipitate the DNA-binding proteins. After a series of extractions to remove the protein from DNA, specific primers are used to PCR amplify the DNA-binding element precipitated with the protein and antibody. Variations of this method are used to identify preferred in vivo binding sites of known DNA-binding proteins. The latest iterations of ChIP analyses include genome-wide analysis of the genomic DNA precipitated in the protein-antibody complex. After genomic DNA purification, adapters are added to permit fluorescent labeling of the DNA for microarray chip analysis (ChIP-chip) or to permit primer binding and size fractionation prior to direct DNA sequencing (ChIP-Seq).[297] Using ChIP-Seq, binding sites for the protein of interest are analyzed across the entire genome in 25–35 nt reading frames. However, 35 nt reads over 3 billion will result in an enormous amount of sequence data to be analyzed requiring significant computing power to establish genomic linkage and identify specific genes. The computational capabilities as opposed to sequencing costs tend to be the major limitations to these genome-wide approaches. In addition, a

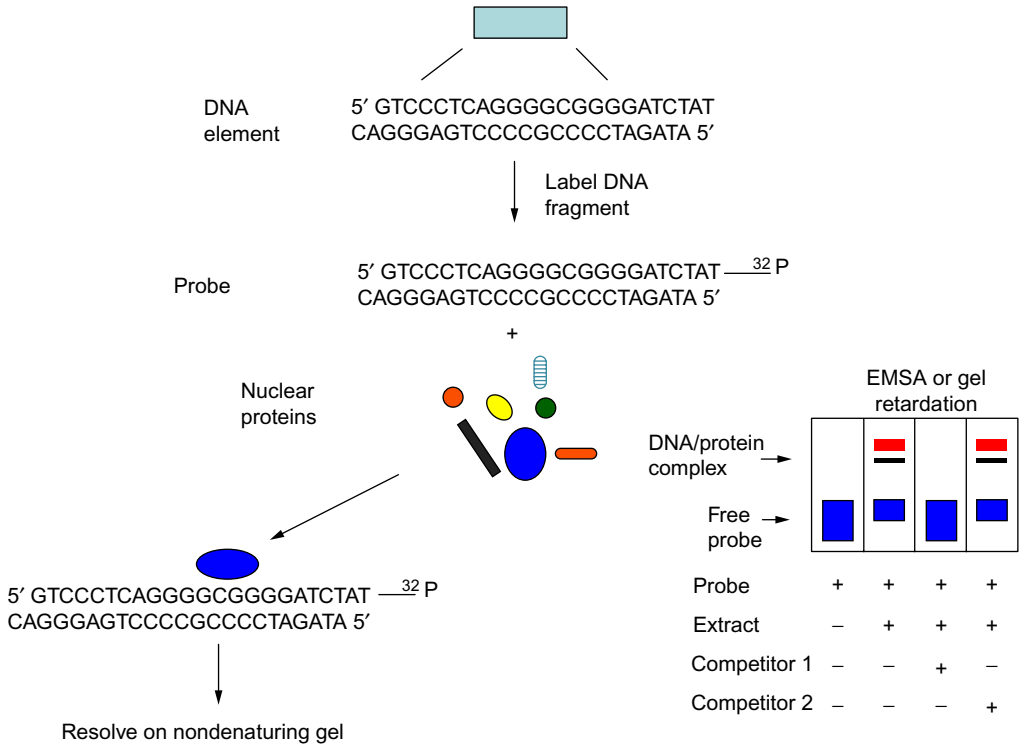


FIG. 1.8 Electrophoretic mobility shift assay (EMSAs, gel shift). A DNA element ~30–100 bp in length is labeled and then incubated with crude nuclear extract or purified protein. A band on the autoradiogram is detected if the radiolabeled probe is retarded and does not migrate to the bottom of the gel. The specificity of binding is determined by competing with unlabeled DNA sequences. Competitor 1 is related to the probe sequence, whereas Competitor 2 is unrelated to the probe sequence. *(Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

similar analysis can be applied to RNA-binding proteins.[298] Alternatively, the immunoprecipitate is resolved on an SDS gel, and mass spectroscopy can be used to identify the proteins that coprecipitate and are likely involved with protein-protein interactions with the DNA-binding proteins. ChIP techniques completely depend upon the quality of the antibodies, the quality and quantity of genomic DNA precipitated, and primer specificity. ChIP assays complement in vitro DNA-protein interaction assays such as EMSAs or footprinting. To demonstrate functional significance, expression vectors or a cell-based knockout strategies using dominant negative constructs, antisense technology, or RNA interference may be used.[299] These approaches are rapid and useful to perform prior to using transgenic mouse approaches.

1.4.1.2 Microarray Technology

At the transcription level, DNA array technology increases by several orders of magnitude the number of genes that can be examined simultaneously under different conditions.[300–302] The number of genes that are either stimulated or inhibited under various conditions can be studied simultaneously with the limitations being the number of genomic sequences that are spotted on the glass slide. A glass slide is able to hold the genomic sequences of the 25,500 genes, the current estimate of the total number of genes in the human genome. Two types of arrays are available—EST/cDNA and oligonucleotide Affymetrix-based. The EST microarray chips use *e*xpressed *s*equence *t*ags that are fragments of DNA corresponding to segments of the genome that encode mRNA. The Affymetrix gene chips spot commercially designed oligonucleotide sequences. These DNA fragments are subsequently "arrayed" onto glass slides. In most instances, several regions of the genomic sequence unique to that gene are spotted in multiple copies to ensure reproducibility. Different genetic domains are plated because of differences in hybridization affinity. The RNA is isolated from cells or tissues after treatment with an extracellular molecule or from cells at different stages of development or transformation. Copy DNA is then generated and fluorescently tagged and then hybridized under stringent conditions to the DNA arrayed on the glass slide followed by analysis by a special plate reader. Computer-generated algorithms are required to interpret the fluorescent signals and rank the degree of change from baseline fluorescence. The technology is being used to study the gene expression pattern found in various tissues at under various conditions, for example, development, inflammation, and transformation.[303–305] The significance of the findings must be confirmed by alternative methods including Northern blot analysis or quantitative polymerase chain reaction.

1.4.1.3 Whole Genome Sequencing

Thanks to the intense focus on sequencing the genomes of various species, especially human, there are now high-throughput methods to batch sequence genomic DNA. Genomic DNA is prepared after treating cells with growth factors, pharmaceuticals, etc., or after ChIP. Primer tags are ligated to the raw ends of the extracted genomic DNA generated after sonication (ChIP-Seq) or restriction digest, then submitted for multiple automated rounds of sequencing.[306] Depending on the Sequencer used, an average of 70–3000 nt can be sequenced over 0.1–3 gigabytes. The longer the sequencing run, the fewer the total number of bases that are sequenced.[307] RNA-Seq methods are becoming more frequent as the focus on alternative splice products and noncoding RNAs expands.[308,309] The expectation within the next 5–10 years is that the cost of genomic sequencing will be sufficiently inexpensive to permit sequencing the DNA of participants in all clinical studies or of animal models. As a result, the storage of such data is beginning to spawn new industries focused on storage, access, privacy, and ethical issues as well as repetitive data mining.

1.4.1.4 Proteomics

Analogous high-throughput approaches have been developed to study protein modifications.[310] However, the techniques used to detect protein posttranslational modifications are more complex and use more labor-intensive technology. Protein is extracted from the cell or organelle of interest and resolved by two-dimensional gel electrophoresis in which proteins are separated by both size and ionic charge (along a pH gradient). The proteins are visualized with a dye either directly on the gel or transferred to a paper substrate. Both substrates (gel or paper) can be used for further analysis. However, proteins transferred to a paper substrate permit several options for analysis. Resolved proteins that are transferred to paper can be submitted for analysis with an antibody (immunoblot) that might recognize phosphorylated or acetylated peptides. Computers are used to analyze differences in the size of the spot corresponding to the amount of a particular protein modification (phosphorylated, acetylated). Proteins that cannot be identified by antibody can be analyzed by mass spectroscopy. Therefore, proteomics allows regulatory changes that occur because of posttranslational modifications to be followed and quantified for a large number of proteins simultaneously. Taking advantage of information on the same technology used to develop DNA arrays, companies are now developing protein arrays that will be applied to new drug discovery.[311]

1.4.1.5 Bioinformatics and Computational Biology

Perhaps not surprisingly, these high-throughput technologies have generated enormous amounts of information that require sophisticated computers to analyze. As computer technology logarithmically improves, so does desktop computing such that individual investigators can manipulate

the data generated with the assistance of sophisticated programs.[312,313] In most instances, investigators will need to recruit the assistance of a computational core facility or specially trained statisticians to analyze the reams of data. Since a discussion of complex computing algorithms is beyond the scope of this chapter, the author refers the reader to several recent reviews on the topic.[314–316]

1.5. POSTTRANSCRIPTIONAL PROCESSING

1.5.1 Polyadenylation

Three major events occur at the end of transcription: (i) the poly(A) tail is added, (ii) adenine bases are methylated, and (iii) hnRNA is processed by removing introns prior to exiting the nucleus (Fig. 1.3).[317] All mRNA except those encoding most histone proteins have poly(A) tails. The length of the poly(A) tail that added ranges from 200 to 250 bp and is quite uniform among eukaryotic organisms. Once the transcript reaches the cytoplasm, the length of the poly(A) sequence decreases with the age of the transcript.[318] Thus, polyadenylation contributes to mRNA stability and translational activation, processes that also involve a synergistic interaction with the cap site.[319–321] Because there is no poly(dT) sequence within DNA, addition of the poly(A) tail represents a posttranscriptional modification of the newly synthesized mRNA. The AATAAA site in DNA is transcribed as AAUAAA and signals endonuclease cleavage of hnRNA ~20 bp after this RNA element.[322] Several factors are required for specific recognition of the AAUAAA element before the addition of adenylate residues by poly(A) polymerase.[323,324] Polyadenylation occurs in two phases: (i) an AAUAAA-dependent phase marked by addition of the first 10 residues and (ii) an AAUAAA-independent phase marked by rapid elongation and catalyzed by a poly(A)-binding protein.[325] In addition, endonuclease cleavage of poly adenylated histone H1 transcripts has also been shown to require the presence of small nuclear ribonucleoproteins (U7 snRNP, pronounced "snurp"), trans-acting factors that participate in RNA splicing reactions.[326] Transcription can proceed for up to 2 kb past the polyadenylation site and may terminate prematurely 30% of the time. Adenylate residues within exons are methylated at the sixth nitrogen and are thought to serve a protective role for those sequences that will eventually be translated.[327] It is now known that formation of the preinitiation complex is linked to the assembly of factors involved in polyadenylation.[328]

1.5.2 RNA Splicing

1.5.2.1 The Spliceosome

Soon after the termination of transcription, most vertebrate hnRNA (pre-mRNA) will be posttranscriptionally processed after exiting the nucleus into a form that can be translated (Fig. 1.3). This involves removing intervening sequences that in some transcripts contain transcriptional regulatory signals (cis-acting elements) some of which are now known to encode miRNAs. Comparing the genomic sequence with the cDNA prepared from an RNA template identifies splice sites. The cis-acting elements within the intron that regulate RNA splicing are GU (GT in the genomic sequence) at the 5′ splice border, AG at the 3′ splice border, and a pyrimidine-rich element that defines the area of the branchpoint 20 bp upstream from the 3′ splice junction (Fig. 1.9). The branchpoint lies just upstream of the pyrimidine-rich region $(PyPy)_n$ and is a highly conserved sequence in yeast (UACUAAC) but much less so in vertebrates.

Five small nuclear RNAs—U1, U2, U5, U4, and U6 (snRNAs)—combine with subsets of about 10 different proteins to form small nuclear ribonucleoproteins (snRNPs; pronounced "snurps").[329,330] The snRNAs, ranging in size from 56 to 217 nt, are quite abundant in the nucleoplasm and contain a trimethylguanylate cap. Some proteins are components of all five major snRNPs, while others are unique to one snRNP. The U7 snRNP, which is present in low concentrations, participates in the 3′ posttranscriptional processing of hnRNA (poly(A).[331] The five major snRNPs assemble into large multicomponent complexes called *spliceosomes* to carry out the splicing reactions.[332] There reactions occur in three steps: cleavage at the 5′ exon-intron border with formation of a branchpoint, excision of the branchpoint as a lariat, and joining of the exons. Splice site selection can be influenced by subtle changes in flanking exon sequences.[333–335]

The basic steps in RNA processing illustrated in Fig. 1.9 are as follows[331]: U1 snRNP binds in a sequence-specific manner to the 5′ exon-intron junction of capped pre-mRNA.[336] An U2 snRNP accessory factor (U2AF) then binds to the pyrimidine-rich element prior to sequence-specific recognition of the branchpoint element by U2 snRNP.[337,338] The 5′ exon is released by cleavage of the 5′ exon junction. This allows the freed 5′-guanylate residue to form a phosphodiester bond at the 2′ site of an adenylate residue within the branchpoint. U4 and U6 snRNPs are paired together by complementary bases and function as a single snRNP complex.[339] The recruitment of the U4/U6 snRNPs to the spliceosome is essential to the last excision step and final removal of the intron from the pre-mRNA. U4/U6 snRNP cooperates with the U2 branchpoint complex without direct contact with RNA.[340] U5 snRNP binds just upstream of the 3′ splice junction to initiate cleavage of the 3′ intron border. Finally, the intron is removed as a lariat and the two exons are joined. More recent evidence indicates that small RNAs themselves catalyze the splicing reactions without the presence of specific enzymes.[341,342] As observed for polyadenylation, the splicing events coincide with transcriptional events.[343] It is therefore

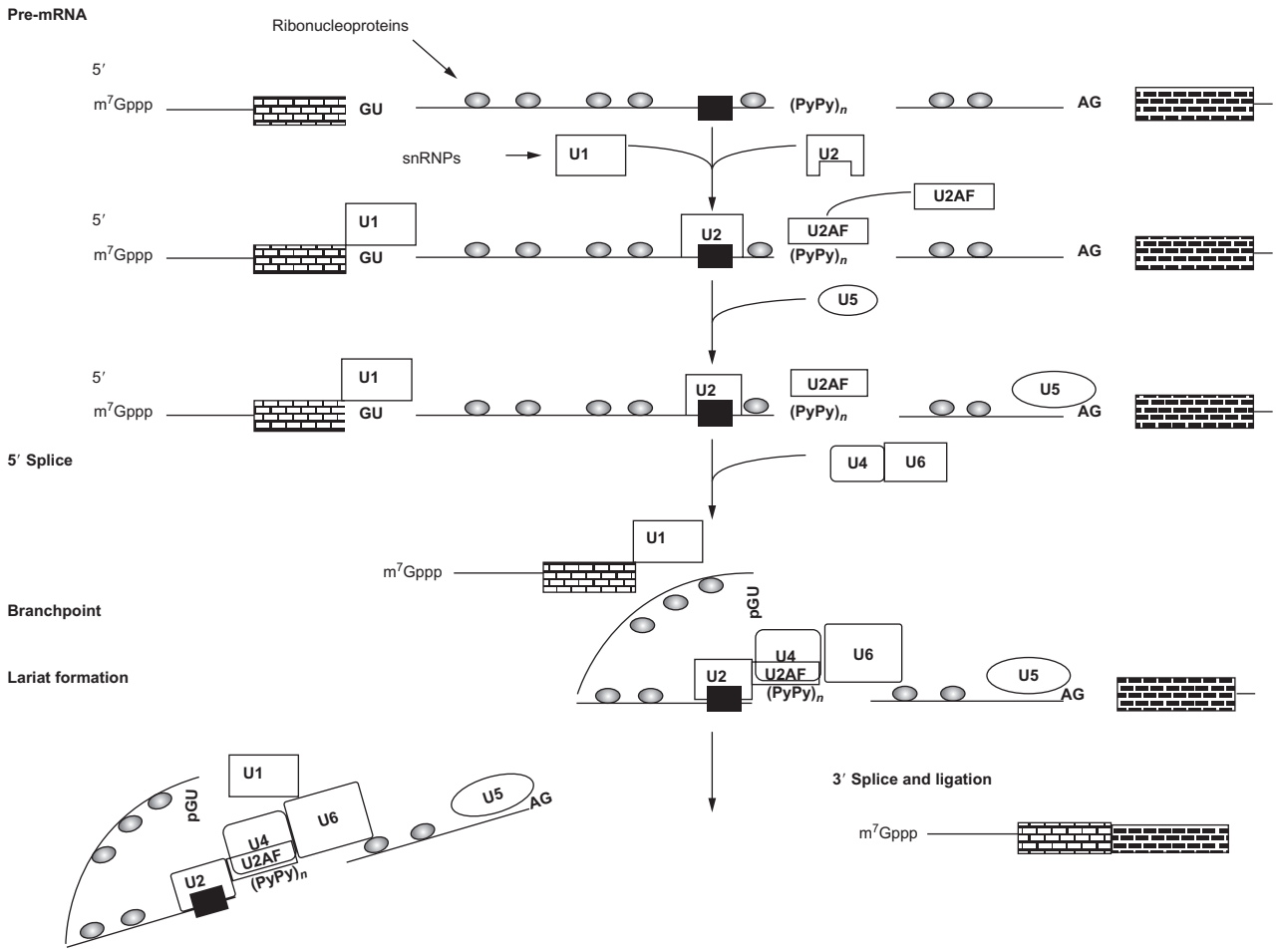


FIG. 1.9 RNA splicing reactions. First, small ribonucleoproteins (snRNPs or "snurps") and accessory factors (U2 accessory factor, U2AF) bind in a sequence-specific manner to the branchpoint and intron-exon borders. Second, the 5′ exon-intron border is cleaved and a "lariat" is formed by the free end of the intron at the branch point. Third, the 3′ intron-exon border is cleaved, the exons are joined, and the excised intron is removed in the form of a "lariat." *(Reprinted from* Physiology of the gastrointestinal tract. *4th ed. 2006.)*

somewhat surprising that the events involved in splicing are not better understood. Nevertheless, with the understanding that the complexity of the human genome lies beyond the DNA sequence and at the level of epigenetics, alternative-splice products and noncoding RNA, the next decade will likely witness heightened attention to these other nuclear processes.[343,344]

1.5.2.2 Alternative Splicing

Eukaryotic cells have applied the mechanics of RNA splicing to generate the protein diversity necessary to meet their multiple demands. Thus, in contrast to the original definition of a gene in which only one transcript is produced, complex genes can generate multiple protein isoforms from multiple RNA transcripts through alternative splicing.[345] This can be achieved by altering which introns and exons are included in or excluded from the mature mRNA transcript that is used as the template for peptide chain elongation. Accordingly, the definition of introns and exons for each gene is actually a fluid concept because an intron for one gene product might become an exon within another transcript. Alternative splicing is a mechanism used by many protein classes, including muscle-related genes, hormones, and transcription factors.[346–350]

1.5.2.3 Regulated Posttranscriptional Mechanisms—mRNA Stability

In addition to cis-acting DNA elements, the cis and trans models of regulation also occur at the posttranscriptional level.[351] Ferritin and the transferrin receptor (TfR), which regulate the storage and uptake of iron, were the best-known examples of regulated posttranscriptional control[352] until the discovery of noncoding RNAs. cis-Acting RNA elements, responsible for conferring iron regulation to both proteins (iron response elements or IREs), reside in the 5′ and 3′ untranslated regions (UTRs) of ferritin and TfR mRNA transcripts, respectively. The same iron-binding protein (IRE-BP), which binds to the IRE in the 5′ UTR of

ferritin to block translation, will also bind to the 3′ UTR of TfR to block mRNA degradation.[351,353,354] Therefore, regulation of iron levels ultimately depends upon posttranscriptional mechanisms that either block translation or increase mRNA stability.

1.5.2.4 Regulated Posttranscriptional Mechanisms—mRNAs

Due to the explosion of interest in these molecules, a separate section has been devoted to this specialized type of posttranscriptional regulation.[355] miRNAs have become especially important in understanding GI development due to their role in embryonic stem cells[356] and in GI cancers,[25,357] especially hepatocellular[358] and gastric.[359,360] How miRNAs repress translation has not been firmly established. Kwak and Tomari proposed at least six different mechanisms several of which include interference with elongation factors or ribosome assembly.[25] miRNAs exhibit broader regulation of gene expression by coincidently targeting the multiple 3′ UTRs of different genes. miRNAs are categorized by their "seed region" in positions 2–8 and 12–16 in the 3′ region. This permits mismatching of the remaining 12 or so nucleotides and subsequently the ability to target several gene loci. By contrast short-interfering or siRNA molecules are quite stringent in their sequence recognition and primarily target the 5′ UTR of the same gene from which it is transcribed. Thus, genome-wide high-throughput approaches are now being used to identify and catalog the miRNAs of different tissues under various conditions, for example, cancer, inflammation.

1.6. TRANSPORT ACROSS THE NUCLEAR MEMBRANE

As noted earlier, RNA is synthesized initially as a much larger primary transcript molecule that in many instances undergoes posttranscriptional modification (e.g., splicing, degradation). However, for any mature RNA transcript to be translated, it must be transported from the nucleus to the cytoplasm.[13] In contrast, nuclear regulatory proteins are translated in the cytoplasm and are eventually returned to the nucleus, either immediately following synthesis or following a dormant state from which they are activated in response to signals.[361] This bidirectional shuttling of macromolecules between the cytoplasm and the nucleus occurs through the nuclear pore complex (NPC), a specialized compartment of the nuclear membrane regulated by a group of transport receptors called karyopherins.[362] Both import and export processes through the nucleus require energy in the form of the Ras-related GTPase Ran and specific targeting signals on the cargo to be transported (nuclear localization and export signals or NLS, NES, respectively).[363,364] The three-dimensional structure of the NPC reveals a doughnut-shaped structure comprised of eight subunits.[365] From the eight subunits emanate "spoke-like" structures, which radiate inward to form a central plug.[366,367] Its cytoplasmic surface is closely associated with ribosomes. Its nuclear surface is thought to participate in the organization of the genome by binding to specific DNA sequences within transcribed genes whose products may be destined for export from the nucleus (gene-gating hypothesis).[368] Perturbations in the nuclear envelope affect chromosome organization, DNA repair, and the cell cycle.[369] Consequently, there are several rare clinical disorders associated with a defective nuclear envelope, for example, muscular dystrophic and premature aging (Hutchinson-Gilbert progeria syndrome).[370,371]

1.7. CONCLUDING REMARKS

In summary, the goals of this initial chapter are to introduce the reader to the basic molecular building blocks of the cell and the techniques used to study them. Some of the sections provide a historic perspective tinged with recently established or evolving concepts. In no way are the sections meant to be exhaustive reviews. Rather, they will hopefully arm the reader with sufficient background to understand the current molecular biology literature and apply the concepts to the study of the GI tract. Molecular physiology continues to be strongly impacted by the explosion of knowledge about the epigenome and noncoding RNAs. Although our ability to access the genetic basis for cellular structure and function using genome-wide approaches has become routine, the challenges that lie ahead will exist primarily at the level of data management, function (phenotype) characterization, and translation to whole organisms, for example, in vivo animal models and ultimately to human physiology and disease.

ACKNOWLEDGMENT

The work was supported by Public Health Service NIH Grant R01-DK55732 and R37-DK45729 to JLM.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**(6822):860–921.
2. Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A* 2007;**104**(49):19428–33.
3. Jia M, Gao X, Zhang Y, Hoffmeister M, Brenner H. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. *Clin Epigenetics* 2016;**8**:25.
4. Sawadogo M, Sentenac A. RNA polymerase B (II) and general transcription factors. *Annu Rev Biochem* 1990;**59**:711–54.
5. Zhang Y, Najmi SM, Schneider DA. Transcription factors that influence RNA polymerases I and II: to what extent is mechanism of action conserved? *Biochim Biophys Acta* 2017;**1860**(2):246–55.

6. Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, et al. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 2008;**37**:337–52.
7. Novina CD, Sharp PA. The RNAi revolution. *Nature* 2004;**430**(6996):161–4.
8. Belogurov GA, Artsimovitch I. Regulation of transcript elongation. *Annu Rev Microbiol* 2015;**69**:49–69.
9. Kusnadi EP, Hannan KM, Hicks RJ, Hannan RD, Pearson RB, Kang J. Regulation of rDNA transcription in response to growth factors, nutrients and energy. *Gene* 2015;**556**(1):27–34.
10. Orioli A. tRNA biology in the omics era: stress signalling dynamics and cancer progression. *Bioessays* 2017;**39**(3).
11. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008;**118**(5):1590–605.
12. Chatterjee S, Ahituv N. Gene regulatory elements, major drivers of human disease. *Annu Rev Genomics Hum Genet* 2017;**18**:45–63.
13. Bjork P, Wieslander L. Integration of mRNP formation and export. *Cell Mol Life Sci* 2017;**74**(16):2875–97.
14. Fitzgerald M, Shenk T. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 1981;**24**:251–60.
15. Birnstiel M, Busslinger M, Strub K. Transcription termination and 3′ processing: the end is in site. *Cell* 1985;**41**:349–59.
16. McDevitt MA, Gilmartin GM, Nevins JR. Multiple factors are required for poly (A) addition to a mRNA 3′ end. *Genes Dev* 1988;**2**:588–97.
17. Shatkin AJ. Capping of eukaryotic mRNAs. *Cell* 1976;**9**:645–54.
18. Rozen F, Sonenberg N. Identification of nuclear cap-specific proteins in HeLa cells. *Nucleic Acids Res* 1987;**15**:6489–500.
19. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;**10**(3):155–9.
20. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 1999;**286**(5441):950–2.
21. Ha I, Wightman B, Ruvkun G. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev* 1996;**10**(23):3041–50.
22. Neilson JR, Sharp PA. Small RNA regulators of gene expression. *Cell* 2008;**134**(6):899–902.
23. Adlakha YK, Seth P. The expanding horizon of microRNAs in cellular reprogramming. *Prog Neurobiol* 2017;**148**:21–39.
24. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* 2002;**21**(17):4663–70.
25. Kwak PB, Iwasaki S, Tomari Y. The microRNA pathway and cancer. *Cancer Sci* 2010;**101**(11):2309–15.
26. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 2008;**134**(3):521–33.
27. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol* 2005;**3**(3):e85.
28. Makarova JA, Shkurnikov MU, Wicklein D, Lange T, Samatov TR, Turchinovich AA, et al. Intracellular and extracellular microRNA: an update on localization and biological role. *Prog Histochem Cytochem* 2016;**51**(3–4):33–49.
29. Liu X, Liu X, Wu Y, Wu Q, Wang Q, Yang Z, et al. MicroRNAs in biofluids are novel tools for bladder cancer screening. *Oncotarget* 2017;**8**(19):32370–9.
30. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 2016;**17**(1):47–62.
31. Ausubel FM, Brent R, Moore DD, Seidman JG, Smith JA, Struhl K. *Current protocols in molecular biology*. New York: John Wiley and Sons; 2004.
32. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**(7561):486–90.
33. Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle* 2014;**13**(18):2847–52.
34. Sewitz SA, Fahmi Z, Lipkow K. Higher order assembly: folding the chromosome. *Curr Opin Struct Biol* 2017;**42**:162–8.
35. Schmitt AD, Hu M, Ren B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol* 2016;**17**(12):743–55.
36. Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* 2016;**30**(12):1357–82.
37. Chung IM, Ketharnathan S, Kim SH, Thiruvengadam M, Rani MK, Rajakumar G. Making sense of the tangle: insights into chromatin folding and gene regulation. *Genes (Basel)* 2016;**7**(10).
38. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 2013;**339**(6123):1074–7.
39. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**(7539):317–30.
40. Sanjana NE, Wright J, Zheng K, Shalem O, Fontanillas P, Joung J, et al. High-resolution interrogation of functional elements in the noncoding genome. *Science* 2016;**353**(6307):1545–9.
41. Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* 2007;**17**(5):400–7.
42. Ghirlando R, Giles K, Gowher H, Xiao T, Xu Z, Yao H, et al. Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta* 2012;**1819**(7):644–51.
43. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, et al. The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A* 2002;**99**(Suppl. 4):16433–7.
44. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, et al. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* 2002;**99**(10):6883–8.
45. Hou C, Dale R, Dean A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 2010;**107**(8):3651–6.
46. Zhang Y, Liang J, Li Y, Xuan C, Wang F, Wang D, et al. CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen. *J Biol Chem* 2010;**285**(37):28604–13.
47. Hnisz D, Day DS, Young RA. Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 2016;**167**(5):1188–200.
48. Acemel RD, Maeso I, Gomez-Skarmeta JL. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdiscip Rev Dev Biol* 2017;**6**(3).
49. Kaiser VB, Semple CA. When TADs go bad: chromatin structure and nuclear organisation in human disease. *F1000Res* 2017;**6**:.
50. Moreira PN, Giraldo P, Cozar P, Pozueta J, Jimenez A, Montoliu L, et al. Efficient generation of transgenic mice with intact yeast artifi-

cial chromosomes by intracytoplasmic sperm injection. *Biol Reprod* 2004;**71**(6):1943–7.
51. Krzywinski M, Wallis J, Gosele C, Bosdet I, Chiu R, Graves T, et al. Integrated and sequence-ordered BAC- and YAC-based physical maps for the rat genome. *Genome Res* 2004;**14**(4):766–79.
52. Newman RJ, Roose-Girma M, Warming S. Efficient conditional knockout targeting vector construction using co-selection BAC recombineering (CoSBR). *Nucleic Acids Res* 2015;**43**(19):e124.
53. Holmes S, Lyman S, Hsu JK, Cheng J. Making BAC transgene constructs with lambda-red recombineering system for transgenic animals or cell lines. *Methods Mol Biol* 2015;**1227**:71–98.
54. Cotta-de-Almeida V, Schonhoff S, Shibata T, Leiter A, Snapper SB. A new method for rapidly generating gene-targeting vectors by engineering BACs through homologous recombination in bacteria. *Genome Res* 2003;**13**(9):2190–4.
55. Jiang F, Doudna JA. CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys* 2017;**46**:505–29.
56. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem* 2016;**85**:227–64.
57. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* 1985;**4**(6):1609–14.
58. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 2010;**11**(9):636–46.
59. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009;**326**(5959):1509–12.
60. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 2013;**14**(1):49–55.
61. Wu C, Morris JR. Genes, genetics, and epigenetics: a correspondence. *Science* 2001;**293**(5532):1103–5.
62. Golubovsky M, Manton KG. Genome organization and three kinds of heritable changes: general description and stochastic factors (a review). *Front Biosci* 2005;**10**:335–44.
63. Sarma K, Reinberg D. Histone variants meet their match. *Nat Rev Mol Cell Biol* 2005;**6**(2):139–49.
64. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004;**429**(6990):457–63.
65. Isles AR, Holland AJ. Imprinted genes and mother-offspring interactions. *Early Hum Dev* 2005;**81**(1):73–7.
66. Bartolomei MS, Webber AL, Brunkow ME, Tilghman SM. Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev* 1993;**7**:1663–73.
67. Bhat MI, Kapila R. Dietary metabolites derived from gut microbiota: critical modulators of epigenetic changes in mammals. *Nutr Rev* 2017;**75**(5):374–89.
68. Kuriakose JS, Miller RL. Environmental epigenetics and allergic diseases: recent advances. *Clin Exp Allergy* 2010;**40**(11):1602–10.
69. Morel D, Almouzni G, Soria JC, Postel-Vinay S. Targeting chromatin defects in selected solid tumors based on oncogene addiction, synthetic lethality and epigenetic antagonism. *Ann Oncol* 2017;**28**(2):254–69.
70. Salas LA, Johnson KC, Koestler DC, O'Sullivan DE, Christensen BC. Integrative epigenetic and genetic pan-cancer somatic alteration portraits. *Epigenetics* 2017;**12**(7):561–74.
71. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;**99**(3):247–57.
72. Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol* 2000;**225**(2):294–303.
73. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 1999;**96**(25):14412–7.
74. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 2005;**74**:481–514.
75. Feng J, Wilkinson M, Liu X, Purushothaman I, Ferguson D, Vialou V, et al. Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol* 2014;**15**(4):R65.
76. LaPlant Q, Vialou V, Covington 3rd HE, Dumitriu D, Feng J, Warren BL, et al. Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat Neurosci* 2010;**13**(9):1137–43.
77. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 1993;**90**(24):11995–9.
78. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;**3**(6):415–28.
79. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;**16**(1):6–21.
80. Liang G, Weisenberger DJ. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. *Epigenetics* 2017;**12**:416–32.
81. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997;**13**(8):335–40.
82. Fazzari MJ, Greally JM. Epigenomics: beyond CpG islands. *Nat Rev Genet* 2004;**5**(6):446–55.
83. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell* 2007;**128**(4):669–81.
84. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 2002;**99**(6):3740–5.
85. Lee JH, Park SJ, Abraham SC, Seo JS, Nam JH, Choi C, et al. Frequent CpG island methylation in precursor lesions and early gastric adenocarcinomas. *Oncogene* 2004;**23**(26):4646–54.
86. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 2004;**36**(5):497–501.
87. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;**301**(5895):89–92.
88. Geisel J, Schorr H, Bodis M, Isber S, Hubner U, Knapp JP, et al. The vegetarian lifestyle and DNA methylation. *Clin Chem Lab Med* 2005;**43**(10):1164–9.
89. Huang S. Histone methyltransferases, diet nutrients and tumour suppressors. *Nat Rev Cancer* 2002;**2**(6):469–76.
90. Kirkland JB. Niacin status impacts chromatin structure. *J Nutr* 2009;**139**(12):2397–401.
91. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;**38**(12):1378–85.
92. Ho E, Zempleni J. Overview to symposium "Nutrients and epigenetic regulation of gene expression". *J Nutr* 2009;**139**(12):2387–8.
93. Gregg C, Zhang J, Butler JE, Haig D, Dulac C. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 2010;**329**(5992):682–5.
94. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 2010;**329**(5992):643–8.

95. Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. *Cell* 2004;**116**(2):259–72.
96. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997;**389**(6648):251–60.
97. Klenova E, Ohlsson R. Poly(ADP-ribosyl)ation and Epigenetics: is CTCF PARt of the Plot? *Cell Cycle* 2005;**4**(1):96–101.
98. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature* 2000;**403**(6765):41–5.
99. Kouzarides T. Chromatin modifications and their function. *Cell* 2007;**128**(4):693–705.
100. Wu J, Grunstein M. 25 Years after the nucleosome model: chromatin modifications. *Trends Biochem Sci* 2000;**25**(12):619–23.
101. Lund AH, van Lohuizen M. Epigenetics and cancer. *Genes Dev* 2004;**18**(19):2315–35.
102. Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. *Nature* 2003;**425**(6957):475–9.
103. Wolffe A. *Chromatin: structure and function*. San Diego: Academic Press; 1998.
104. Workman JL, Kingston RE. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* 1998;**67**:545–79.
105. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* 2000;**64**(2):435–59.
106. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Annu Rev Biochem* 2001;**70**:81–120.
107. Utley RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, et al. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* 1998;**394**(6692):498–502.
108. Voss AK, Thomas T. MYST family histone acetyltransferases take center stage in stem cells and development. *Bioessays* 2009;**31**(10):1050–61.
109. Baker SP, Grant PA. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* 2007;**26**(37):5329–40.
110. Koutelou E, Hirsch CL, Dent SY. Multiple faces of the SAGA complex. *Curr Opin Cell Biol* 2010;**22**(3):374–82.
111. Marks PA, Miller T, Richon VM. Histone deacetylases. *Curr Opin Pharmacol* 2003;**3**(4):344–51.
112. Marks PA, Xu WS. Histone deacetylase inhibitors: potential in cancer therapy. *J Cell Biochem* 2009;**107**(4):600–8.
113. Khochbin S, Verdel A, Lemercier C, Seigneurin-Berny D. Functional significance of histone deacetylase diversity. *Curr Opin Genet Dev* 2001;**11**(2):162–6.
114. Marlett JA, McBurney MI, Slavin JL. Position of the American Dietetic Association: health implications of dietary fiber. *J Am Diet Assoc* 2002;**102**(7):993–1000.
115. Hinnebusch BF, Meng S, Wu JT, Archer SY, Hodin RA. The effects of short-chain fatty acids on human colon cancer cell phenotype are associated with histone hyperacetylation. *J Nutr* 2002;**132**(5):1012–7.
116. Leder A, Leder P. Butyric acid, a potent inducer of erythroid differentiation in cultured erythroleukemic cells. *Cell* 1975;**5**(3):319–22.
117. Riggs MG, Whittaker RG, Neumann JR, Ingram VM. *n*-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. *Nature* 1977;**268**(5619):462–4.
118. Sealy L, Chalkley R. The effect of sodium butyrate on histone modification. *Cell* 1978;**14**(1):115–21.
119. Kruh J. Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Mol Cell Biochem* 1982;**42**(2):65–82.
120. Candido EP, Reeves R, Davie JR. Sodium butyrate inhibits histone deacetylation in cultured cells. *Cell* 1978;**14**(1):105–13.
121. Boffa LC, Vidali G, Mann RS, Allfrey VG. Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. *J Biol Chem* 1978;**253**(10):3364–6.
122. Lutter LC, Judis L, Paretti RF. Effects of histone acetylation on chromatin topology in vivo. *Mol Cell Biol* 1992;**12**(11):5004–14.
123. Almouzni G, Khochbin S, Dimitrov S, Wolffe AP. Histone acetylation influences both gene expression and development of Xenopus laevis. *Dev Biol* 1994;**165**(2):654–69.
124. Van Lint C, Emiliani S, Ott M, Verdin E. Transcriptional activation and chromatin remodeling of the HIV-1 promoter in response to histone acetylation. *EMBO J* 1996;**15**(5):1112–20.
125. Archer SY, Hodin RA. Histone acetylation and cancer. *Curr Opin Genet Dev* 1999;**9**(2):171–4.
126. Daroqui MC, Augenlicht LH. Transcriptional attenuation in colon carcinoma cells in response to butyrate. *Cancer Prev Res (Phila)* 2010;**3**(10):1292–302.
127. Paolicchi E, Gemignani F, Krstic-Demonacos M, Dedhar S, Mutti L, Landi S. Targeting hypoxic response for cancer therapy. *Oncotarget* 2016;**7**(12):13464–78.
128. Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol Cell* 2012;**48**(4):612–26.
129. Losson H, Schnekenburger M, Dicato M, Diederich M. Natural compound histone deacetylase inhibitors (HDACi): synergy with inflammatory signaling pathway modulators and clinical applications in cancer. *Molecules* 2016;**21**(11).
130. Lakshmaiah KC, Jacob LA, Aparna S, Lokanatha D, Saldanha SC. Epigenetic therapy of cancer with histone deacetylase inhibitors. *J Cancer Res Ther* 2014;**10**(3):469–78.
131. Pouillart PR. Role of butyric acid and its derivatives in the treatment of colorectal cancer and hemoglobinopathies. *Life Sci* 1998;**63**(20):1739–60.
132. Csordas A. Butyrate, aspirin and colorectal cancer. *Eur J Cancer Prev* 1996;**5**(4):221–31.
133. Jung M. Inhibitors of histone deacetylase as new anticancer agents. *Curr Med Chem* 2001;**8**(12):1505–11.
134. Mariadason JM, Corner GA, Augenlicht LH. Genetic reprogramming in pathways of colonic cell maturation induced by short chain fatty acids: comparison with trichostatin A, sulindac, and curcumin and implications for chemoprevention of colon cancer. *Cancer Res* 2000;**60**(16):4561–72.
135. Van Lint C, Emiliani S, Verdin E. The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation. *Gene Expr* 1996;**5**(4–5):245–53.
136. Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 2000;**19**(6):1176–9.
137. Li L, Davie JR. The role of Sp1 and Sp3 in normal and cancer cell biology. *Ann Anat* 2010;**192**(5):275–83.
138. Kim TY, Kim IS, Jong HS, Lee JW, Kim TY, Jung M, et al. Transcriptional induction of DLC-1 gene through Sp1 sites by histone deacetylase inhibitors in gastric cancer cells. *Exp Mol Med* 2008;**40**(6):639–46.
139. Yu DC, Waby JS, Chirakkal H, Staton CA, Corfe BM. Butyrate suppresses expression of neuropilin I in colorectal cell lines through inhibition of Sp1 transactivation. *Mol Cancer* 2010;**9**:276.
140. Mikkelsen IM, Huseby NE, Visvikis A, Moens U. Activation of the gamma-glutamyltransferase promoter 2 in the rat colon carcinoma

cell line CC531 by histone deacetylase inhibitors is mediated through the Sp1 binding motif. *Biochem Pharmacol* 2002;**64**(2):307–15.

141. Tsubaki J, Hwa V, Twigg SM, Rosenfeld RG. Differential activation of the IGF binding protein-3 promoter by butyrate in prostate cancer cells. *Endocrinology* 2002;**143**(5):1778–88.
142. Yang J, Kawai Y, Hanson RW, Arinze IJ. Sodium butyrate induces transcription from the G alpha(i2) gene promoter through multiple Sp1 sites in the promoter and by activating the MEK-ERK signal transduction pathway. *J Biol Chem* 2001;**276**(28):25742–52.
143. Lu Y, Lotan R. Transcriptional regulation by butyrate of mouse galectin-1 gene in embryonal carcinoma cells. *Biochim Biophys Acta* 1999;**1444**(1):85–91.
144. Taniura S, Kamitani H, Watanabe T, Eling TE. Transcriptional regulation of cyclooxygenase-1 by histone deacetylase inhibitors in normal human astrocyte cells. *J Biol Chem* 2002;**277**(19):16823–30.
145. Kim JH, Meng S, Shei A, Hodin RA. A novel Sp1-related cis element involved in intestinal alkaline phosphatase gene transcription. *Am J Phys* 1999;**276**(4 Pt 1):G800–7.
146. Sowa Y, Orita T, Hiranabe-Minamikawa S, Nakano K, Mizuno T, Nomura H, et al. Histone deacetylase inhibitor activates the p21/WAF1/Cip1 gene promoter through the Sp1 sites. *Ann N Y Acad Sci* 1999;**886**:195–9.
147. Xiao H, Hasegawa T, Isobe K. p300 collaborates with Sp1 and Sp3 in p21(waf1/cip1) promoter activation induced by histone deacetylase inhibitor. *J Biol Chem* 2000;**275**(2):1371–6.
148. Doetzlhofer A, Rotheneder H, Lagger G, Koranda M, Kurtev V, Brosch G, et al. Histone deacetylase 1 can repress transcription by binding to Sp1. *Mol Cell Biol* 1999;**19**(8):5504–11.
149. Maehara K, Uekawa N, Isobe K. Effects of histone acetylation on transcriptional regulation of manganese superoxide dismutase gene. *Biochem Biophys Res Commun* 2002;**295**(1):187–92.
150. Zhang Y, Dufau ML. Silencing of transcription of the human luteinizing hormone receptor gene by histone deacetylase-mSin3A complex. *J Biol Chem* 2002;**277**(36):33431–8.
151. Mottet D, Pirotte S, Lamour V, Hagedorn M, Javerzat S, Bikfalvi A, et al. HDAC4 represses p21(WAF1/Cip1) expression in human cancer cells through a Sp1-dependent, p53-independent mechanism. *Oncogene* 2009;**28**(2):243–56.
152. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 2000;**25**(3):338–42.
153. Zhu P, Martin E, Mengwasser J, Schlag P, Janssen KP, Gottlicher M. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. *Cancer Cell* 2004;**5**(5):455–63.
154. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* 2002;**296**(5570):1132–6.
155. Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet* 2000;**24**(1):88–91.
156. Kristensen LS, Nielsen HM, Hansen LL. Epigenetics and cancer treatment. *Eur J Pharmacol* 2009;**625**(1–3):131–42.
157. Ma H, Baumann CT, Li H, Strahl BD, Rice R, Jelinek MA, et al. Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr Biol* 2001;**11**(24):1981–5.
158. Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 2001;**293**(5539):2453–5.
159. Noma K, Allis CD, Grewal SI. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* 2001;**293**(5532):1150–5.
160. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 2001;**410**(6824):120–4.
161. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000;**406**(6796):593–9.
162. Yeates TO. Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell* 2002;**111**(1):5–7.
163. Lachner M, O'Sullivan RJ, Jenuwein T. An epigenetic road map for histone lysine methylation. *J Cell Sci* 2003;**116**(Pt 11):2117–24.
164. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 2001;**107**(3):323–37.
165. Ng SS, Yue WW, Oppermann U, Klose RJ. Dynamic protein methylation in chromatin biology. *Cell Mol Life Sci* 2009;**66**(3):407–22.
166. Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* 2006;**7**(9):715–27.
167. Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev* 2008;**22**(9):1115–40.
168. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, et al. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 2009;**139**(7):1303–14.
169. Herz HM, Shilatifard A. The JARID2-PRC2 duality. *Genes Dev* 2010;**24**(9):857–61.
170. Li G, Margueron R, Ku M, Chambon P, Bernstein BE, Reinberg D. Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev* 2010;**24**(4):368–80.
171. Landeira D, Sauer S, Poot R, Dvorkina M, Mazzarella L, Jorgensen HF, et al. Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat Cell Biol* 2010;**12**(6):618–24.
172. Wang Y, Wysocka J, Sayegh J, Lee YH, Perlin JR, Leonelli L, et al. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* 2004;**306**(5694):279–83.
173. Cuthbert GL, Daujat S, Snowden AW, Erdjument-Bromage H, Hagiwara T, Yamada M, et al. Histone deimination antagonizes arginine methylation. *Cell* 2004;**118**(5):545–53.
174. Daujat S, Bauer UM, Shah V, Turner B, Berger S, Kouzarides T. Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Curr Biol* 2002;**12**(24):2090–7.
175. Pal S, Yun R, Datta A, Lacomis L, Erdjument-Bromage H, Kumar J, et al. mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. *Mol Cell Biol* 2003;**23**(21):7475–87.
176. O'Brien KB, Alberich-Jorda M, Yadav N, Kocher O, Diruscio A, Ebralidze A, et al. CARM1 is required for proper control of proliferation and differentiation of pulmonary epithelial cells. *Development* 2010;**137**(13):2147–56.
177. Lee JS, Smith E, Shilatifard A. The language of histone crosstalk. *Cell* 2010;**142**(5):682–5.

178. Perez-Cadahia B, Drobic B, Khan P, Shivashankar CC, Davie JR. Current understanding and importance of histone phosphorylation in regulating chromatin biology. *Curr Opin Drug Discov Dev* 2010;**13**(5):613–22.
179. Labrador M, Corces VG. Phosphorylation of histone H3 during transcriptional activation depends on promoter structure. *Genes Dev* 2003;**17**(1):43–8.
180. Cheung P, Allis CD, Sassone-Corsi P. Signaling to chromatin through histone modifications. *Cell* 2000;**103**(2):263–71.
181. Mahadevan LC, Willis AC, Barratt MJ. Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. *Cell* 1991;**65**(5):775–83.
182. Thomson S, Clayton AL, Hazzalin CA, Rose S, Barratt MJ, Mahadevan LC. The nucleosomal response associated with immediate-early gene induction is mediated via alternative MAP kinase cascades: MSK1 as a potential histone H3/HMG-14 kinase. *EMBO J* 1999;**18**(17):4779–93.
183. Clayton AL, Mahadevan LC. MAP kinase-mediated phosphoacetylation of histone H3 and inducible gene regulation. *FEBS Lett* 2003;**546**(1):51–8.
184. Keen N, Taylor S. Aurora-kinase inhibitors as anticancer agents. *Nat Rev Cancer* 2004;**4**(12):927–36.
185. Goepfert TM, Brinkley BR. The centrosome-associated Aurora/Ipl-like kinase family. *Curr Top Dev Biol* 2000;**49**:331–42.
186. Yamamoto Y, Verma UN, Prajapati S, Kwak YT, Gaynor RB. Histone H3 phosphorylation by IKK-alpha is critical for cytokine-induced gene expression. *Nature* 2003;**423**(6940):655–9.
187. Soloaga A, Thomson S, Wiggin GR, Rampersaud N, Dyson MH, Hazzalin CA, et al. MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J* 2003;**22**(11):2788–97.
188. Ajiro K. Histone H2B phosphorylation in mammalian apoptotic cells. An association with DNA fragmentation. *J Biol Chem* 2000;**275**(1):439–43.
189. Cheung WL, Ajiro K, Samejima K, Kloc M, Cheung P, Mizzen CA, et al. Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* 2003;**113**(4):507–17.
190. Kawashima SA, Yamagishi Y, Honda T, Ishiguro K, Watanabe Y. Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science* 2010;**327**(5962):172–7.
191. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol Cell* 2000;**5**(6):905–15.
192. Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, et al. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol Cell* 2000;**5**(6):917–26.
193. Barth TK, Imhof A. Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* 2010;**35**(11):618–26.
194. Koppens MA, Bounova G, Gargiulo G, Tanger E, Janssen H, Cornelissen-Steijger P, et al. Deletion of polycomb repressive complex 2 from mouse intestine causes loss of stem cells. *Gastroenterology* 2016;**151**(4):684. 612.
195. Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 2002;**419**(6907):624–9.
196. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* 2003;**100**(20):11606–11.
197. Shaffer DR, Pandolfi PP. Breaking the rules of cancer. *Nat Med* 2006;**12**(1):14–5.
198. Nikoloski G, Langemeijer SM, Kuiper RP, Knops R, Massop M, Tonnissen ER, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* 2010;**42**(8):665–7.
199. Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* 2010;**42**(8):722–6.
200. Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science* 2008;**322**(5908):1695–9.
201. Kodach LL, Jacobs RJ, Heijmans J, van Noesel CJ, Langers AM, Verspaget HW, et al. The role of EZH2 and DNA methylation in the silencing of the tumour suppressor RUNX3 in colorectal cancer. *Carcinogenesis* 2010;**31**(9):1567–75.
202. Wang HJ, Ruan HJ, He XJ, Ma YY, Jiang XT, Xia YJ, et al. MicroRNA-101 is down-regulated in gastric cancer and involved in cell migration and invasion. *Eur J Cancer* 2010;**46**(12):2295–303.
203. Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 2010;**17**(5):443–54.
204. Iwama A, Oguro H, Negishi M, Kato Y, Morita Y, Tsukui H, et al. Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. *Immunity* 2004;**21**(6):843–51.
205. Lessard J, Sauvageau G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* 2003;**423**(6937):255–60.
206. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL, et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* 2003;**423**(6937):302–5.
207. Yanai H, Atsumi N, Tanaka T, Nakamura N, Komai Y, Omachi T, et al. Intestinal cancer stem cells marked by Bmi1 or Lgr5 expression contribute to tumor propagation via clonal expansion. *Sci Rep* 2017;**7**:41838.
208. Li N, Nakauka-Ddamba A, Tobias J, Jensen ST, Lengner CJ. Mouse label-retaining cells are molecularly and functionally distinct from reserve intestinal stem cells. *Gastroenterology* 2016;**151**(2):298–310. e7.
209. Srinivasan T, Than EB, Bu P, Tung KL, Chen KY, Augenlicht L, et al. Notch signalling regulates asymmetric division and interconversion between lgr5 and bmi1 expressing intestinal stem cells. *Sci Rep* 2016;**6**:26069.
210. Konuma T, Oguro H, Iwama A. Role of the polycomb group proteins in hematopoietic stem cells. *Develop Growth Differ* 2010;**52**(6):505–16.
211. Li Z, Cao R, Wang M, Myers MP, Zhang Y, Xu RM. Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J Biol Chem* 2006;**281**(29):20643–9.
212. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, et al. MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 2002;**10**(5):1107–17.
213. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ. MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. *Mol Cell Biol* 2001;**21**(7):2249–58.
214. Rozenblatt-Rosen O, Rozovskaia T, Burakov D, Sedkov Y, Tillib S, Blechman J, et al. The C-terminal SET domains of ALL-1 and

TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. *Proc Natl Acad Sci U S A* 1998;**95**(8):4152–7.
215. Kohler C, Hennig L. Regulation of cell identity by plant Polycomb and trithorax group proteins. *Curr Opin Genet Dev* 2010;**20**(5):541–7.
216. Gracanin A, Dreijerink KM, van der Luijt RB, Lips CJ, Hoppener JW. Tissue selectivity in multiple endocrine neoplasia type 1-associated tumorigenesis. *Cancer Res* 2009;**69**(16):6371–4.
217. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1). *Best Pract Res Clin Endocrinol Metab* 2010;**24**(3):355–70.
218. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* 2004;**6**(8):731–40.
219. Shukla A, Chaurasia P, Bhaumik SR. Histone methylation and ubiquitination with their cross-talk and roles in gene expression and stability. *Cell Mol Life Sci* 2009;**66**(8):1419–33.
220. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998;**19**(2):187–91.
221. Pedersen MT, Helin K. Histone demethylases in development and disease. *Trends Cell Biol* 2010;**20**(11):662–71.
222. Hansen RS, Gartler SM, Scott CR, Chen SH, Laird CD. Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Hum Mol Genet* 1992;**1**(8):571–8.
223. Murata T, Kurokawa R, Krones A, Tatsumi K, Ishii M, Taki T, et al. Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome. *Hum Mol Genet* 2001;**10**(10):1071–6.
224. Bhaumik SR, Smith E, Shilatifard A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* 2007;**14**(11):1008–16.
225. Nottke A, Colaiacovo MP, Shi Y. Developmental roles of the histone lysine demethylases. *Development* 2009;**136**(6):879–89.
226. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* 1997;**57**(5):808–11.
227. Ogawa O, Eccles MR, Szeto J, McNoe LA, Yun K, Maw MA, et al. Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* 1993;**362**(6422):749–51.
228. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, et al. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 1998;**93**(3):361–72.
229. Xu W, Edmondson DG, Evrard YA, Wakamiya M, Behringer RR, Roth SY. Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. *Nat Genet* 2000;**26**(2):229–32.
230. Gayther SA, Batley SJ, Linger L, Bannister A, Thorpe K, Chin SF, et al. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet* 2000;**24**(3):300–3.
231. Kouzarides T. Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev* 1999;**9**(1):40–8.
232. Borrow J, Stanton Jr. VP, Andresen JM, Becher R, Behm FG, Chaganti RS, et al. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet* 1996;**14**(1):33–41.
233. Carapeti M, Aguiar RC, Watmore AE, Goldman JM, Cross NC. Consistent fusion of MOZ and TIF2 in AML with inv(8)(p11q13). *Cancer Genet Cytogenet* 1999;**113**(1):70–2.
234. Redner RL, Wang J, Liu JM. Chromatin remodeling and leukemia: new therapeutic paradigms. *Blood* 1999;**94**(2):417–28.
235. Grignani F, De Matteis S, Nervi C, Tomassoni L, Gelmetti V, Cioce M, et al. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukaemia. *Nature* 1998;**391**(6669):815–8.
236. Jones LK, Saha V. Chromatin modification, leukaemia and implications for therapy. *Br J Haematol* 2002;**118**(3):714–27.
237. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009;**106**(28):11667–72.
238. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 2010;**142**(3):409–19.
239. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;**464**(7291):1071–6.
240. Wan Y, Chang HY. HOTAIR: flight of noncoding RNAs in cancer metastasis. *Cell Cycle* 2010;**9**(17):3391–2.
241. Barber BA, Rastegar M. Epigenetic control of Hox genes during neurogenesis, development, and disease. *Ann Anat* 2010;**192**(5):261–74.
242. Rajasekhar VK, Begemann M. Concise review: roles of polycomb group proteins in development and disease: a stem cell perspective. *Stem Cells* 2007;**25**(10):2498–510.
243. Yang X, Lay F, Han H, Jones PA. Targeting DNA methylation for epigenetic therapy. *Trends Pharmacol Sci* 2010;**31**(11):536–46.
244. Goodrich JA, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* 2010;**11**(8):549–58.
245. Carles CC, Fletcher JC. Missing links between histones and RNA Pol II arising from SAND? *Epigenetics* 2010;**5**(5):381–5.
246. Chavan AV, Somani RR. HDAC inhibitors—new generation of target specific treatment. *Mini-Rev Med Chem* 2010;**10**(13):1263–76.
247. Jazirehi AR. Regulation of apoptosis-associated genes by histone deacetylase inhibitors: implications in cancer therapy. *Anti-Cancer Drugs* 2010;**21**(9):805–13.
248. Pili R, Liu G, Chintala S, Verheul H, Rehman S, Attwood K, et al. Combination of the histone deacetylase inhibitor vorinostat with bevacizumab in patients with clear-cell renal cell carcinoma: a multicentre, single-arm phase I/II clinical trial. *Br J Cancer* 2017;**116**(7):874–83.
249. Schobert R, Biersack B. Multimodal HDAC inhibitors with improved anticancer activity. *Curr Cancer Drug Targets* 2017; https://doi.org/10.2174/1568009617666170206102613.
250. Johnsson AE, Wright AP. The role of specific HAT-HDAC interactions in transcriptional elongation. *Cell Cycle* 2010;**9**(3):467–71.
251. Marschalek R. Mixed lineage leukemia: roles in human malignancies and potential therapy. *FEBS J* 2010;**277**(8):1822–31.
252. Epstein R. In: Epstein R, editor. *Human molecular biology*. Cambridge: Cambridge University Press; 2003.
253. Baumann M, Pontiller J, Ernst W. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol Biotechnol* 2010;**45**(3):241–7.
254. Dynan WS, Tjian R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 1983;**35**:79–87.

255. Van Dyke MW, Roeder RG, Sawadogo M. Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* 1988;**241**:1335–8.
256. Hai T, Horikoshi M, Roeder RG, Green MR. Analysis of the role of the transcription factor ATF in the assembly of a functional preinitiation complex. *Cell* 1988;**54**:1043–51.
257. Dynan WS, Tjian R. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* 1985;**316**:774–8.
258. McKnight S, Tjian R. Transcriptional selectivity of viral genes in mammalian cells. *Cell* 1986;**46**:795–805.
259. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 1989;**245**:371–8.
260. Dynan WS. Modularity in promoters and enhancers. *Cell* 1989;**58**:1–4.
261. Brown DD. The role of stable complexes that repress and activate eucaryotic genes. *Cell* 1984;**37**:359–65.
262. Renkawitz R. Transcriptional repression in eukaryotes. *Trends Genet* 1990;**6**:192–7.
263. Jackson SM, Keech CA, Williamson DJ, Gutierrez-Hartmann A. Interaction of basal positive and negative transcription elements controls repression of the proximal rat prolactin promoter in nonpituitary cells. *Mol Cell Biol* 1992;**12**:2708–19.
264. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter—the gateway to transcription. *Curr Opin Cell Biol* 2008;**20**(3):253–9.
265. Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SH, Kadonaga JT. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev* 2017;**31**(1):6–11.
266. Struhl K. Yeast transcriptional regulatory mechanisms. *Annu Rev Genet* 1995;**29**:651–74.
267. Hahn S, Buratowski S, Sharp PA, Guarente L. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc Natl Acad Sci U S A* 1989;**86**:5718–22.
268. Pugh BF, Tjian R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 1990;**61**:1187–97.
269. Wefald FC, Devlin BH, Williams RS. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* 1990;**344**:260–2.
270. Sawadogo M, Roeder RG. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* 1985;**43**:165–75.
271. Horikoshi M, Hai T, Lin Y-S, Green MR, Roeder RG. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* 1988;**54**:1033–42.
272. Nakajima N, Horikoshi M, Roeder RG. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol Cell Biol* 1988;**8**:4028–40.
273. Smale ST, Baltimore D. The "initiator" as a transcription control element. *Cell* 1989;**57**:103–13.
274. Pugh BF, Tjian R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev* 1991;**5**:1935–45.
275. Weis L, Reinberg D. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. *FASEB J* 1992;**6**(14):3300–9.
276. Smale ST, Schmidt MC, Berk AJ, Baltimore D. Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proc Natl Acad Sci U S A* 1990;**87**:4509–13.
277. Wiborg O, Berglund L, Boel E, Norris F, Norris K, Rehfeld JF, et al. Structure of a human gastrin gene. *Proc Natl Acad Sci U S A* 1984;**81**:1067–9.
278. Andrisani OM, Dixon JE. Somatostatin gene regulation. *Annu Rev Physiol* 1990;**52**:793–806.
279. Haun RS, Dixon JE. A transcriptional enhancer essential for the expression of the rat cholecystokinin gene contains a sequence identical to the -296 element of the human c-fos gene. *J Biol Chem* 1990;**265**:15455–63.
280. Philippe J, Drucker DJ, Knepel W, Jepeal L, Misulovin Z, Habener JF. Alpha-cell-specific expression of the glucagon gene is conferred to the glucagon promoter element by the interactions of DNA-binding proteins. *Mol Cell Biol* 1988;**8**:4877–88.
281. Kopin AS, Wheeler MB, Nishitani J, McBride EW, Chang T-M, Chey WY, et al. The secretin gene: evolutionary history, alternative splicing, and developmental regulation. *Proc Natl Acad Sci U S A* 1991;**88**:5335–9.
282. Blasband AJ, Rogers KT, Chen X, Azizkhan JC, Lee DC. Characterization of the rat transforming growth factor a gene and identification of promoter sequences. *Mol Cell Biol* 1990;**10**:2111–21.
283. Revzin A. Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* 1989;**7**:346–55.
284. Mueller PR, Salser SJ, Wold B. Constitutive and metal-inducible protein: DNA interactions at the mouse metallothionein I promoter examined by in vivo and in vitro footprinting. *Genes Dev* 1988;**2**:412–27.
285. Landolfi NF, Yin X-M, Capra JD, Tucker PW. Protection analysis (or "footprinting") of specific protein-DNA complexes in crude nuclear extracts using methidiumpropyl-EDTA-iron (II). *Biotechniques* 1989;**7**:500–4.
286. Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res* 1981;**9**:3047–60.
287. Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* 1981;**9**:6505–25.
288. Williams M, Brys A, Weiner AM, Maizels N. A rapid method for determining the molecular weight of a protein bound to nucleic acid in a mobility shift assay. *Nucleic Acids Res* 1992;**20**:4935–6.
289. Herr W. Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA. *Proc Natl Acad Sci U S A* 1985;**82**:8009–13.
290. Billon N, Carlisi D, Datto MB, van Grunsven LA, Watt A, Wang XF, et al. Cooperation of Sp1 and p300 in the induction of the CDK inhibitor p21WAF1/CIP1 during NGF-mediated neuronal differentiation. *Oncogene* 1999;**18**(18):2872–82.
291. Silva CM, Tully DB, Petch LA, Jewell M, Cidlowski JA. Application of a protein-blotting procedure to the study of human glucocorticoid receptor interactions with DNA. *Proc Natl Acad Sci U S A* 1987;**84**:1744–8.
292. Miskimins WK, Roberts MP, McClelland A, Ruddle FH. Use of a protein-blotting procedure and a specific DNA probe to identify

nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc Natl Acad Sci U S A* 1985;**82**:6741–4.
293. Singh H, Clerc RG, LeBowitz JH. Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. *Biotechniques* 1989;**7**:252–61.
294. Das PM, Ramachandran K, vanWert J, Singal R. Chromatin immunoprecipitation assay. *Biotechniques* 2004;**37**(6):961–9.
295. Bernstein BE, Humphrey EL, Liu CL, Schreiber SL. The use of chromatin immunoprecipitation assays in genome-wide analyses of histone modifications. *Methods Enzymol* 2004;**376**:349–60.
296. Weinmann AS, Farnham PJ. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* 2002;**26**(1):37–47.
297. Wong E, Wei CL. ChIP'ing the mammalian genome: technical advances and insights into functional elements. *Genome Med* 2009;**1**(9):89.
298. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010;**11**(7):476–86.
299. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature* 2004;**431**(7006):343–9.
300. Mockler TC, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 2005;**85**(1):1–15.
301. Mantripragada KK, Buckley PG, de Stahl TD, Dumanski JP. Genomic microarrays in the spotlight. *Trends Genet* 2004;**20**(2):87–94.
302. Simon R, Mirlacher M, Sauter G. Tissue microarrays. *Biotechniques* 2004;**36**(1):98–105.
303. Cunliffe VT. Memory by modification: the influence of chromatin structure on gene expression during vertebrate development. *Gene* 2003;**305**(2):141–50.
304. Buchholz M, Gress TM. Application of DNA array analyses in the management of gastrointestinal cancer patients. *Dig Dis* 2003;**21**(4):309–14.
305. Cowell JK. High throughput determination of gains and losses of genetic material using high resolution BAC arrays and comparative genomic hybridization. *Comb Chem High Throughput Screen* 2004;**7**(6):587–96.
306. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008;**24**(3):133–41.
307. Horner DS, Pavesi G, Castrignano T, De Meo PD, Liuni S, Sammeth M, et al. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief Bioinform* 2010;**11**(2):181–97.
308. Ning K, Fermin D. SAW: a method to identify splicing events from RNA-Seq data based on splicing fingerprints. *PLoS One* 2010;**5**(8):e12047.
309. Morin RD, Zhao Y, Prabhu AL, Dhalla N, McDonald H, Pandoh P, et al. Preparation and analysis of microRNA libraries using the Illumina massively parallel sequencing technology. *Methods Mol Biol* 2010;**650**:173–99.
310. Huang RP. Protein arrays, an excellent tool in biomedical research. *Front Biosci* 2003;**8**:d559–76.
311. Ng JH, Ilag LL. Biomedical applications of protein chips. *J Cell Mol Med* 2002;**6**(3):329–40.
312. Cheung E, Kraus WL. Genomic analyses of hormone signaling and gene regulation. *Annu Rev Physiol* 2010;**72**:191–218.
313. Ding L, Wendl MC, Koboldt DC, Mardis ER. Analysis of next-generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet* 2010;**19**(R2):R188–96.
314. Sun Y, Cai Y, Mai V, Farmerie W, Yu F, Li J, et al. Advanced computational algorithms for microbial community analysis using massive 16S rRNA sequence data. *Nucleic Acids Res* 2010;**38**(22):e205.
315. Nagarajan N, Pop M. Sequencing and genome assembly using next-generation technologies. *Methods Mol Biol* 2010;**673**:1–17.
316. Nijkamp J, Winterbach W, van den Broek M, Daran JM, Reinders M, de Ridder D. Integrating genome assemblies with MAIA. *Bioinformatics* 2010;**26**(18):i433–9.
317. Zorio DA, Bentley DL. The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res* 2004;**296**(1):91–7.
318. Proudfoot N. Ending the message is not so simple. *Cell* 1996;**87**(5):779–81.
319. Bernstein P, Peltz SW, Ross J. The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* 1989;**9**:659–70.
320. Vassalli J-D, Huarte J, Belin D, Gubler P, Vassalli A, O'Connell ML, et al. Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes Dev* 1989;**3**:2163–71.
321. Gallie DR. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* 1991;**5**:2108–16.
322. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res* 2003;**31**(5):1375–86.
323. Takagaki Y, Ryner LC, Manley JL. Four factors are required for 3′-end cleavage of pre-mRNAs. *Genes Dev* 1989;**3**:1711–24.
324. Gilmartin GM, Nevins JR. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes Dev* 1989;**3**:2180–9.
325. Wahle E. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 1991;**66**:759–68.
326. Kirsh AL, Groudine M, Challoner PB. Polyadenylation and U7 snRNP-mediated cleavage: alternative modes of RNA 3′ processing in two avian histone H1 genes. *Genes Dev* 1993;**3**:2172–9.
327. Darnell J, Lodish H, Baltimore D. *Molecular cell biology*. New York: Scientific American Books, Inc.; 1990.
328. Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* 2003;**17**(11):1321–7.
329. Krainer AR, Maniatis T. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 necessary for pre-mRNA splicing in vitro. *Cell* 1985;**42**:725–36.
330. Krämer A, Frick M, Keller W. Separation of multiple components of HeLa cell nuclear extracts required for pre-messenger RNA splicing. *J Biol Chem* 1987;**262**:17630–40.
331. Maniatis T, Reed R. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* 1987;**325**:673–8.
332. Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 2003;**25**(12):1147–9.
333. Berget SM. Exon recognition in vertebrate splicing. *J Biol Chem* 1995;**270**(6):2411–4.
334. Reed R, Maniatis T. A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 1986;**46**:681–90.
335. Nelson KK, Green MR. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev* 1988;**2**:319–29.

336. Jacob M, Gallinaro H. The 5′ splice site: phylogenetic evolution and variable geometry of association with U1RNA. *Nucleic Acids Res* 1989;**17**:2159–80.
337. Krämer A. Presplicing complex formation requires two proteins and U2 snRNP. *Genes Dev* 1988;**2**:1155–67.
338. Nelson KK, Green MR. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev* 1989;**3**:1562–71.
339. Fabrizio P, McPheeters DS, Abelson J. In vitro assembly of yeast U6 snRNP: a functional assay. *Genes Dev* 1989;**3**:2137–50.
340. Reed R. The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev* 1989;**3**:2113–23.
341. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 1992;**71**:803–17.
342. Weiner AM. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* 1993;**72**:161–4.
343. Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. Multiple links between transcription and splicing. *RNA* 2004;**10**(10):1489–98.
344. Shin C, Manley JL. Cell signalling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol* 2004;**5**(9):727–38.
345. Darnell Jr. JE. The processing of RNA. *Sci Am* 1983;**249**:90–100.
346. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003;**72**:291–336.
347. Kashiwabara S, Zhuang T, Yamagata K, Noguchi J, Fukamizu A, Baba T. Identification of a novel isoform of poly(A) polymerase, TPAP, specifically present in the cytoplasm of spermatogenic cells. *Dev Biol* 2000;**228**(1):106–15.
348. Hoeffler JP, Meyer TE, Waeber G, Habener JF. Multiple adenosine 3′,5′-monophosphate response element DNA-binding proteins generated by gene diversification and alternative exon splicing. *Mol Endocrinol* 1990;**4**:920–30.
349. Helfman DM, Roscigno RF, Mulligan GJ, Finn LA, Weber KS. Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA. *Genes Dev* 1990;**4**:98–110.
350. Breitbart RE, Andreadis A, Nadal-Ginard B. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* 1987;**56**:467–95.
351. Klausner RD, Harford JB. Cis-trans models for post-transcriptional gene regulation. *Science* 1989;**246**:870–2.
352. Templeton DM, Liu Y. Genetic regulation of cell function in response to iron overload or chelation. *Biochim Biophys Acta* 2003;**1619**(2):113–24.
353. Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, et al. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 1987;**238**:1570–3.
354. Casey JL, Hentze MW, Koeller DM, Caughman SW, Rouault TA, Klausner RD, et al. Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science* 1988;**240**:924–8.
355. O'Hara SP, Mott JL, Splinter PL, Gores GJ, LaRusso NF. MicroRNAs: key modulators of posttranscriptional gene expression. *Gastroenterology* 2009;**136**(1):17–25.
356. Hand NJ, Master ZR, Eauclaire SF, Weinblatt DE, Matthews RP, Friedman JR. The microRNA-30 family is required for vertebrate hepatobiliary development. *Gastroenterology* 2009;**136**(3):1081–90.
357. Hammond SM. MicroRNAs as oncogenes. *Curr Opin Genet Dev* 2006;**16**(1):4–9.
358. Visone R, Petrocca F, Croce CM. Micro-RNAs in gastrointestinal and liver disease. *Gastroenterology* 2008;**135**(6):1866–9.
359. Zhang Z, Li Z, Gao C, Chen P, Chen J, Liu W, et al. miR-21 plays a pivotal role in gastric cancer pathogenesis and progression. *Lab Investig* 2008;**88**(12):1358–66.
360. Kim YK, Yu J, Han TS, Park SY, Namkoong B, Kim DH, et al. Functional links between clustered microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. *Nucleic Acids Res* 2009;**37**(5):1672–81.
361. Pemberton LF, Paschal BM. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 2005;**6**(3):187–98.
362. Strambio-De-Castillia C, Niepel M, Rout MP. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol* 2010;**11**(7):490–501.
363. Hoelz A, Blobel G. Cell biology: popping out of the nucleus. *Nature* 2004;**432**(7019):815–6.
364. Delaleau M, Borden KL. Multiple export mechanisms for mRNAs. *Cells* 2015;**4**(3):452–73.
365. Lusk CP, Makhnevych T, Wozniak RW. New ways to skin a kap: mechanisms for controlling nuclear transport. *Biochem Cell Biol* 2004;**82**(6):618–25.
366. Unwin PNT, Milligan RA. A large particle associated with the perimeter of the nuclear pore complex. *J Cell Biol* 1982;**93**:63–75.
367. Hinshaw JE, Carragher BO, Milligan RA. Architecture and design of the nuclear pore complex. *Cell* 1992;**69**:1133–41.
368. Rout MP, Aitchison JD, Magnasco MO, Chait BT. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol* 2003;**13**(12):622–8.
369. Lusk CP, King MC. The nucleus: keeping it together by keeping it apart. *Curr Opin Cell Biol* 2017;**44**:44–50.
370. Chi YH, Chen ZJ, Jeang KT. The nuclear envelopathies and human diseases. *J Biomed Sci* 2009;**16**:96.
371. Hernandez L, Roux KJ, Wong ES, Mounkes LC, Mutalif R, Navasankari R, et al. Functional coupling between the extracellular matrix and nuclear lamina by Wnt signaling in progeria. *Dev Cell* 2010;**19**(3):413–25.

Chapter 2

Gastrointestinal Hormones☆

Fiona M. Gribble, Frank Reimann, Geoffrey P. Roberts

2.1. INTRODUCTION

The intestinal tract secretes a number of hormones that coordinate local, peripheral, and central responses to food intake. Hormones produced in the stomach are regulated rapidly after food ingestion and are largely involved in the control of acid and enzyme secretion. As food reaches the small intestine, it triggers the secretion of a range of hormones that serve to match the release of digestive enzymes, electrolytes, and bile acids to the composition of the ingested food and to regulate the rate of delivery of nutrients into the duodenum. When nutrients are subsequently absorbed into the bloodstream, the parallel release of gut hormones reflects the rate of nutrient absorption and facilitates downstream hormonal responses such as insulin release, as well as sending signals to the brain to control appetitive behaviors.

Gut hormones are produced from specialized enteroendocrine cells (EECs) located in the epithelium of the gastrointestinal (GI) tract from the stomach through to the rectum. Like other cell types of the intestinal epithelium, EECs are continuously replaced by new cells formed from crypt stem cells. Approximately 1% of newly formed epithelial cells differentiate into EECs, and they share with neighboring enterocytes a similar life span of ~3–5 days in the small intestine, and up to a few weeks in the stomach and colon.[1] Many EECs have an apical surface facing into the intestinal lumen and a basolateral surface facing the interstitium, and are known as open-type cells because they make contact with luminal contents. The exception is the stomach, where except in the antrum, most EECs are closed type and do not have a surface opening into the lumen. Whereas open-type EECs are believed to respond primarily to nutritional stimuli arriving in the local vicinity after food ingestion, closed-type cells are regulated by paracrine, circulating, or neural signals, although nutrients might directly regulate these cells if concentrations rise in their vicinity postabsorption.

2.1.1 Production and Processing of Peptides by Enteroendocrine Cells

EECs have traditionally been classified and named according to the principal hormones they produce as determined by immunostaining (Table 2.1) with each hormone and cell type exhibiting a characteristic distribution along the length of the GI tract.[2] Gastric epithelium, for example, contains a large number of EECs-producing gastrin, somatostatin (SST), ghrelin, or histamine. Small intestine preferentially generates EECs-producing cholecystokinin (CCK), secretin, glucose-dependent insulinotropic polypeptide (GIP), glucagon-like peptides 1 and 2 (GLP-1, GLP-2), peptide YY (PYY), neurotensin (NT), and serotonin (5-HT). In the colon and rectum, EECs have been shown to secrete serotonin, GLP-1, GLP-2, PYY, NT, SST, and insulin-like peptide-5 (INSL5).

Recent molecular techniques examining EEC subpopulations, labeled with fluorescent reporters driven by hormone specific promoters in transgenic mice, have yielded transcriptomic data at odds with the simple EEC classification suggested by the early immunostaining studies that used only one or two antibodies at a time.[3,4]At the messenger RNA (mRNA) level, there is a high degree of overlap between different EEC types that were originally thought to be distinct, and many single EECs produce mRNA for a number of different gut hormones.[4,5] Coproduction of several different hormones in the same EECs has been confirmed by immunostaining.[6] It is now thought that intestinal EECs-producing CCK, GIP, secretin, GLP-1, PYY, and 5-HT form a continuum, with individual cells producing a mix of hormones dependent on their position along the GI tract. Within individual cells, there are conflicting views about whether coexpressed hormones are localized in the same or distinct vesicles, but no convincing evidence has yet been presented to show separate mobilization of different hormones from individual cells.[6,7]

Peptide hormones are biosynthesized as prepropeptides containing N-terminal signal sequences that direct the growing peptide chain into the lumen of the endoplasmic

Physiology of the Gastrointestinal Tract. https://doi.org/10.1016/B978-0-12-809954-4.00002-5

TABLE 2.1 Gut Hormones, Classic Cells of Origin, Principle Location Within the GI Tract and Functions

Peptide	Cell of Origin	Locations of GI Tract Secreted From	Function
Gastrin	G cells	Gastric antrum	Stimulates gastric acid secretion. Differentiation and integrity of gastric mucosa.
Somatostatin (SST)	D cells	Whole GI tract	Delays gastric emptying and gastrointestinal motility. Inhibits secretion of all other gastrointestinal hormones. Reduces colonic fluid secretion. Reduces bile flow and pancreatic exocrine secretion. Reduces splanchnic blood flow.
Ghrelin	X/A like cells	Stomach	Stimulates hunger. Protective during fasting induced hypoglycaemia.
Cholecystokinin (CCK)	I cells	Duodenum	Stimulates gallbladder contraction and pancreatic exocrine secretion. Inhibits gastric emptying and acid secretion. Signals satiety.
Secretin	S cells	Duodenum, jejunum	Stimulates pancreatic exocrine secretion. Inhibits gastric emptying and acid production.
Motilin	M cells	Duodenum, jejunum	Stimulates gastrointestinal motility.
Neurotensin (NT)	N cells	Ileum	Delays gastrointestinal motility. Stimulates pancreatic exocrine secretion.
Glucose-dependent insulinotropic polypeptide (GIP)	K cells	Duodenum, jejunum	Enhances glucose-stimulated insulin secretion (incretin effect). Promotes fat deposition. Reduces bone turnover.
Glucagon-like peptide 1 (GLP-1)	L cells	Jejunum, ileum, colon	Enhances glucose-stimulated insulin secretion (incretin effect), inhibits glucagon secretion. Delays gastric emptying. Signals satiety, reduces food intake.
Glucagon-like peptide 2 (GLP-2)	L cells	Jejunum, ileum, colon	Adaptation and recovery of intestinal mucosa in response to injury.
Oxyntomodulin	L cells	Jejunum, ileum, colon	Body weight homeostasis.
Peptide YY (PYY)	L cells	Jejunum, ileum, colon,	Signals satiety. Inhibits gastric emptying and acid secretion. Maintenance of salt/water homeostasis.
Insulin-like peptide 5 (INSL5)	L cells	Colon	Stimulates hunger.

reticulum during translation. Propeptides transit through the Golgi and are packaged into secretory vesicles where they are cleaved by prohormone convertases (PCs) and further posttranslationally modified by, for example, amidation, sulfation, or acylation. The predominant PC identified in most intestinal EECs is PC1/3, which cleaves propeptides at dibasic residues and is likely responsible for the majority of peptide hormone processing in the small intestine and colon.[8,9] By contrast, PC2 plays a more prominent role in the stomach.[10] EECs are often identified by immunolabeling with antibodies against chromogranins and secretogranins.[3,11] These large granin proteins are believed to play a functional role in vesicular packaging but are also subject to PC-mediated cleavage, resulting in the generation of smaller peptides that might themselves play signaling roles.[12]

2.1.2 How Do EECs Respond to Nutrition-Related Stimuli?

It has long been recognized that a wide variety of nutritional and nonnutritional signals trigger gut hormone secretion, with some stimuli preferentially linked to the release of certain hormones. Comparisons between plasma gut hormone concentrations following matched nutrient loads administered orally versus intravenously in humans have revealed that most gut hormones are preferentially released after oral nutrient ingestion.[13,14] Many additional studies have demonstrated that polymeric macronutrients must be digested into monomers (monosaccharides, free fatty acids, and monoacylglycerides or di/tripeptides and amino acids) before they are capable of triggering gut hormone release.[15]

Transcriptomic analysis and single-cell characterization of fluorescently tagged murine EECs have revealed that they express a range of receptors and transporters capable of detecting a wide variety of stimuli.[1] Even at the single cell level, individual EECs produce machinery capable of detecting multiple stimuli.[5] Unlike taste cells in the tongue, therefore, individual EECs seem to be multimodal rather than tuned to respond to single stimuli. There are two major molecular pathways by which EECs detect ingested nutrients—one involving nutrient transporters and a second involving G-protein-coupled receptors (GPCRs).

Enterocytes typically employ ion-coupled transporters to absorb nutrients across the brush border, using inwardly directed gradients for Na^+ or H^+ ions to drive the uphill absorption of nutrients. A large body of evidence supports the idea that many EECs have hijacked sodium-coupled glucose transporters (SGLT1) on the apical membrane to act as glucose sensors, as the coupled uptake of Na^+ ions with glucose molecules generates an inward current capable of triggering electrical activity, leading to Ca^{2+} entry through voltage-gated Ca^{2+} channels and activation of vesicular exocytotic pathways.[1] There is some, albeit weaker, evidence that certain amino acids and di/tripeptides might similarly trigger gut hormone release via their Na^+- and H^+-coupled uptake.[16,17]

Many small molecules are detected by members of the GPCR superfamily, which include receptors specifically responsive to small molecules including long- and short-chain fatty acids, monoacylglycerides, amino acids, bile acids, and bitter tastants. Nutrient and bile acid responsive GPCRs are highly and specifically expressed in EECs within the intestinal epithelium and likely underlie gut hormone responses to ingested fats and protein, as well as bile acids.[1] GPCRs linked to the stimulation of EECs are mostly G_s and G_q coupled, linked, respectively, to the elevation of cytoplasmic cAMP and Ca^{2+} concentrations. An increasing body of evidence suggests that coincident activation of different signaling pathways in EECs results in synergistic enhancement of gut hormone secretion.[18]

Rather than merely "tasting" the luminal contents, it is increasingly apparent that EECs respond to the local rates of nutrient absorption. In the case of glucose, the rate of SGLT1-mediated glucose uptake by EECs, and hence the degree of glucose-dependent membrane depolarization, will mirror rates of glucose influx by neighboring enterocytes, being determined by the local concentrations of glucose and Na^+ ions.[19] Results from perfused intestinal preparations and Ussing chambers have now shown that EEC receptors for long chain fatty acids and bile acids are functionally located on the basolateral rather than the apical surface of EECs, requiring local absorption across the epithelium prior to receptor activation.[20–22] Linking gut hormone secretion to local nutrient absorption might ensure that the circulating hormonal signal reflects the rate of nutrient entry into the bloodstream, rather than the mass of unabsorbed nutrients in the lumen that do not yet require the activation of a peripheral homeostatic response.

2.1.3 Pathophysiology Affecting Multiple Enteroendocrine Cell Subtypes

In the sections below, we will describe pathologies primarily affecting specific gut hormones, but there are a few conditions that have more generalized effects on the enteroendocrine system. There have been rare case reports of humans born with an almost complete lack of EECs due to mutations in the transcription factor NeuroG3, which is required for cell differentiation down the EEC pathway.[23,24] Affected neonates presented with severe malabsorptive diarrhea. Rare human cases have also been reported with homozygous loss of PC1/3 due to mutations in the PCSK1 gene, resulting in a variable presentation that can include malabsorptive diarrhea, impaired glucose homeostasis, and obesity, as well as other endocrinopathies attributable to the global deficiency of many active hormones and peptide neurotransmitters in the gut, pancreas, and central nervous system.[25] Secondary EEC deficiency associated with gastrointestinal symptoms has been described in the autoimmune-polyendocrine-candidiasis-ectodermal-dystrophy (APECED) syndrome, associated with a mutation in the AIRE gene.[25a]

Neuroendocrine tumors (NETs) of the GI tract can produce a range of unprocessed, partially processed, and fully processed peptide hormones, with the consequence that clinical presentations vary markedly between cases. In most cases, the exact pattern of active peptides produced by an individual tumor is not currently measurable, because of the lack of suitable methodology for the identification and quantification of partially processed peptides.

Some of the most dramatic gut hormone changes in humans have been observed after upper GI surgical procedures such as Roux-en-Y gastric bypass (RYGB) surgery, gastrectomy, or esophagectomy. RYGB and sleeve gastrectomy are performed routinely as a treatment for morbid obesity, but have dramatic metabolic consequences that result in the resolution of the majority of cases of type 2 diabetes.[26,27] As discussed in some of the sections below, dramatic postprandial elevations of gut hormones such as GLP-1 and PYY in these patients are likely caused by increased nutrient delivery to and absorption in the more distal small intestine, and almost certainly contribute to observed improvements in glucose tolerance and reduced appetite.[28,29] Similar hormonal changes have been observed in lean subjects, for example, following resection for gastric cancer, and may contribute to some of the symptoms encompassed under the umbrella of "dumping syndrome."[30]

2.1.4 Details of Specific Gut Hormones With Known Biological and Pathophysiological Roles

In the sections below, we provide details of the major identified gut hormones produced by EECs, focusing particularly on hormones that have known cognate receptors and

functional roles. Peptide sequence nomenclature is based on human sequences, as published in the Uniprot/Swissprot database. The list is not exhaustive and does not include the large number of additional signaling peptides produced by non-EECs types in the gut, such as enteric nerves [e.g., vasoactive intestinal peptide (VIP), gastrin-releasing peptide, galanin], Paneth cells (e.g., defensins), enterocytes (e.g., FGF 15/19), immune cells (e.g., interleukins), and as yet unidentified cell types (e.g., guanylin/uroguanylin).

2.2. GASTRIN

It was first observed in 1905 that mucosal extracts from the gastric antrum stimulated gastric acid secretion when injected intravenously in cats, but it was not until 1942 that this effect was demonstrated to be due to a peptide, gastrin, rather than contamination with histamine.[31,32] The main physiological actions of gastrin are regulation of gastric acid secretion and control of gastric epithelial cell growth and differentiation.[32,33]

2.2.1 Synthesis

Gastrin is primarily secreted from gastric antral G cells, but has also been identified in the pituitary gland, developing pancreas and sperm.[34–36] The gastrin gene encodes a 101 amino acid prepropeptide, containing a 21 amino acid N-terminal signal peptide and 80 amino acid progastrin peptide. All subsequent amino acid position nomenclature refers to the position in the preprogastrin peptide. Following cleavage of the signal peptide in the endoplasmic reticulum, progastrin is sulfated at tyrosine 86 and phosphorylated at serine 96. Further processing in the trans-golgi network and secretory vesicles results in the two mature, C-terminal amidated forms—Gastrin34 (G34) and Gastrin17 (G17). Progastrin is cleaved by PC 1/3 (PC1/3) and carboxypeptidase E (CPE) at amino acid positions 58-59 and 92-93 (the latter removing a C-terminal flanking peptide). The resulting 34 amino acid peptide (G34-Gly) is amidated by peptidyl-glycine α-amidating monooxygenase (PAM), with the glycine group acting as an amide donor. G34 is then cleaved by PC2 to G17, with the two forms present in human G cell vesicles at a G34:G17 ratio of 1:9 (Fig. 2.1).[37–39]

Gastrin shares a significant degree of sequence and structural homology with CCK, and G34 and G17 both act through the CCK2 receptor. Whereas G17 and G34 undergo regulated exocytosis, progastrin, and nonamidated forms of G17 and G34 are secreted via the constitutive pathway. They have no known receptor and have previously been regarded as inactive metabolites, although recent evidence suggests that they may play a role in colonic mucosal proliferation and have a complementary role to that of the amidated gastrins.[40,41]

2.2.2 Secretion

Gastrin secretion is regulated by neuronal-, hormonal-, and nutrient-responsive factors. Gastrin is secreted in response to luminal amino acids detected via apical calcium-sensing receptors (CaSRs), sympathetic and parasympathetic nervous activity, and gastrin-releasing peptide derived from local neurons.[42–50] Gastrin secretion is inhibited by SST, when the gastric luminal pH is below 3. Chronic use of proton pump inhibitors results in hypergastrinaemia.[51]

2.2.3 Function

Gastrin's key role in gastric acid secretion has been demonstrated through gastrin infusion and CCK2R antagonist experiments in man, immunoneutralization in dogs, and in gastrin gene knockout in mice.[52–55] Gastrin acts on enterochromaffin-like (ECL) cells to stimulate histamine secretion, which then acts in a paracrine fashion via H2 receptors on parietal cells to stimulate acid secretion.[56–58] Interestingly, in gastrin-deficient mice, the coinfusion of G17 and the nonamidated G17-Gly more potently restored gastric acid secretion than G17 alone.[40] In addition to stimulating acute histamine secretion, the gastrin upregulated the expression of histidine decarboxylase, the enzyme responsible for conversion of histidine to histamine, in ECL cells.[59] Although CCK2 receptors are also present on parietal cells, these appear to be only of limited role for gastrin

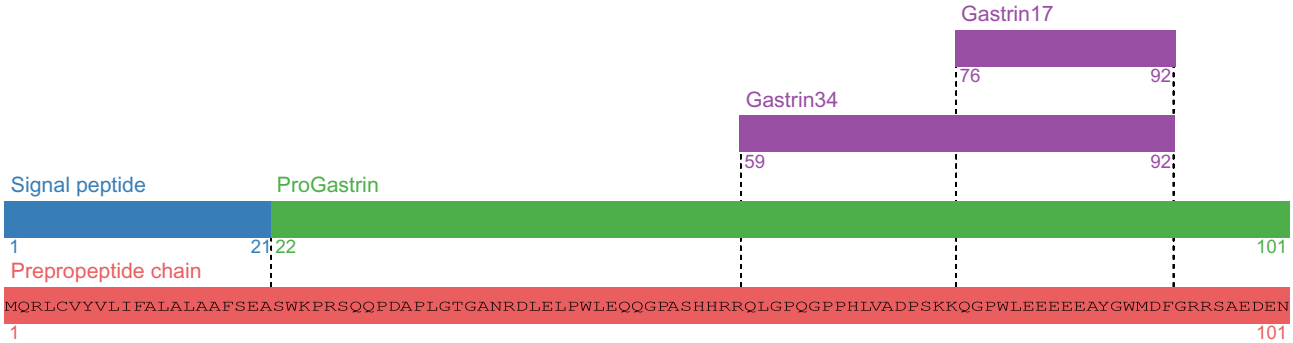


FIG. 2.1 Amino acid sequence of gastrin prepropeptide and secreted peptides, derived from Uniprot/Swissprot database.

stimulation of parietal cell acid secretion, with the majority of the effect of gastrin on acid secretion arising due to histamine from ECLs.[60]

Gastrin is not essential for the development and maintenance of the gastric mucosa, but gastrin gene knockout mice had reduced numbers of parietal and ECL cells, which could be restored by infusion of gastrin.[33,55] It therefore appears that gastrin plays a key role in the differentiation and integrity of the gastric mucosa, although the underlying pathways remain subject to ongoing investigation. One pathway of note involves the urokinase plasminogen activator (uPA) family, including uPA and plasminogen activator inhibitors 1 and 2 (PAI1, PAI2), which localize to gastric parietal and ECL cells.[61–63]

2.2.4 Clinical Aspects

Zollinger-Ellison syndrome, hypergastrinemia secondary to gastrin secreting NETs, is a cause of gastric acid hypersecretion, multiple peptic ulcers, and secretory diarrhea.[64] This is associated with multiple endocrine neoplasia type 1 (MEN1) in up to 20% of cases.[65]

The role of gastrin in gastric mucosal proliferation is of interest in the pathogenesis and treatment of gastric cancer.[66] Specifically, it has been demonstrated that gastrin stimulates the growth of gastric cancer cell lines in vitro by stimulation of CCK2 receptors, and nonendocrine gastric cancer cell lines can secrete gastrin, which may act in an autocrine fashion.[67,68] Despite this, any link between hypergastrinemia secondary to proton pump inhibitor therapy and an increased prevalence of gastric adenocarcinoma remains controversial.[69,70] However, there is a more established link between hypergastrinemia and ECL cell NETs of the stomach, evidence arising from potent H2 receptor blockade in rats using loxtidine and transgenic *Men1*/*Sst* knockout mice treated with omeprazole.[71,72] Gastric carcinoids in man can be associated with hypergastrinemia due to Zollinger-Ellison syndrome (principally in the presence of multiple endocrine neoplasia type 1 [MEN1]) or atrophic gastritis, but not PPI therapy.[73–76]

2.3. SST

SST was originally described in 1973 as a 14 amino acid peptide inhibitor of hypothalamic growth hormone secretion.[77] A 28 amino acid N-terminal extended form was subsequently identified from the GI tract, and the two SST forms are now considered together as a global counterregulatory hormone, with inhibitory effects in multiple target tissues.[78,79]

2.3.1 Synthesis

Both SST-14 and SST-28 are products of a single 116 amino acid prepropeptide translated from the SST gene. The prepropeptide consists of a 24 amino acid N-terminal signal peptide and a 92 amino acid propeptide, of which the terminal 14 and 28 amino acids correspond to the active SST peptides (Fig. 2.2).[80,81]

2.3.2 Secretion

Both 14 and 28 amino acid forms of SST are secreted from gastric and intestinal D cells and pancreatic δ cells, with SST-28 predominating in the small intestine, and SST-14 predominating in the rest of the GI tract and pancreas. Gastric D cells differ between the proximal and distal stomach, with oxyntic D cells exhibiting a closed-type morphology and those in the antrum an open-type morphology. Closed-type oxyntic D cells are inhibited by the vagus nerve soon after food ingestion, thereby reducing the tonic inhibitory control by SST of gastrin and histamine secretion that predominates between meals.[82,83] SST release from the distal antrum is stimulated by nutrient ingestion, reduced gastric pH, CCK, GIP, GLP-1, acetylcholine, VIP, CGRP, and secretin, resulting in a delayed feedback inhibition of gastric secretions that restores acid secretion to basal levels.[84–88]

2.3.3 Receptors

There are five G-protein-coupled SST receptors, labeled numerically from 1 to 5, with SSTR2 having two isoforms,

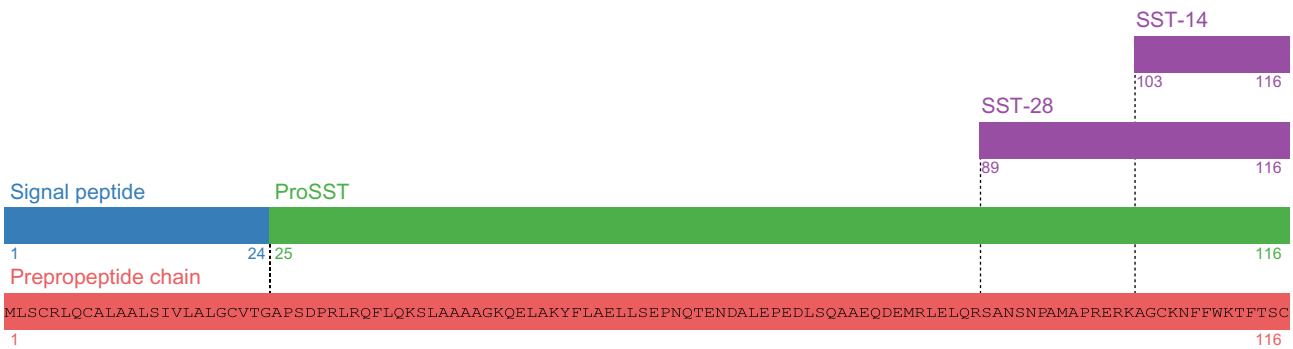


FIG. 2.2 Amino acid sequence of somatostatin prepropeptide and secreted peptides, derived from Uniprot/Swissprot database.

SSTR2A and SSTR2B.[89,90] All SSTRs act through pertussis toxin-sensitive pathways (G_i) to inhibit adenylate cyclase, activate inwardly rectifying potassium channels, and prevent cellular depolarization, calcium influx, and subsequent vesicle exocytosis.[90,91] SSTRs also activate other downstream pathways that reduce cellular proliferation through the action of protein tyrosine phosphatases on MAPKs.[89–91] SST-14 and SST-28 bind with equal affinity to SSTRs 1–4, but SST-28 has a 10–30 fold higher affinity for SSTR5 than other SSTRs, whereas SST-14 has reduced affinity at SSTR5.[90]

2.3.4 Function

2.3.4.1 Stomach

SST acts to inhibit gastrin-mediated acid secretion from gastric parietal cells, acting in a paracrine, endocrine, and neurocrine fashion. SST receptor knockout mouse experiments suggest this is mediated by SSTR2, although a detailed discussion of gastric acid secretion is the topic of a further chapter of this book.[92,93]

2.3.4.2 Gastrointestinal Motility

SST delays intestinal transit by slowing gastric emptying and prolonging migrating motor complexes (MMCs), as well as inhibiting the relaxation of the lower oesophageal sphincter.[94–97] It however remains a topic of some debate as to whether these are global effects, or if SST has differential effects on stomach, small intestine, and colon.[98,99] Experiments to elucidate the underlying mechanisms by which SST has this effect have focused on ex vivo intestines or intestinal smooth muscle. SST has been shown to inhibit VIP-induced relaxation or acetylcholine- and CCK-induced contraction independent of the intestinal section and species investigated; in isolated human colonic smooth muscle cells, removing thereby indirect effects through the modulation of the release of myenteric plexus-derived transmitters, a combination of SSTR1 and SSTR2 activity relaxed smooth muscle cells directly, although high concentrations in the absence of other contracting agents resulted in SST-induced contraction.[99] In rodent small intestine examined ex vivo, SST prolonged MMCs in a SSTR2 and nitric oxide-dependent fashion.[100]

2.3.4.3 Intestinal and Pancreatic Endocrine and Exocrine Secretion

In keeping with its global counterregulatory role, SST inhibits the secretion of multiple gut peptides, including gastrin, GLP-1, motilin, secretin, ghrelin, PYY, 5-HT, and GIP.[101–109]

SST, acting directly on colonocytes, reduces colonic fluid secretion.[110,111]

A series of in vivo and in vitro experiments in dogs, rodents, and humans have used gastroduodenal perfusion and sampling, bile duct ligation, and endoscopic sphincter of Oddi cannulation to examine the role of SST in biliopancreatic secretion. SST has been demonstrated to reduce bile flow by inhibiting secretion and enhancing resorption of fluid by cholangiocytes.[112,113] SST appears to inhibit secretin-mediated pancreatic bicarbonate secretion, but had limited effects on basal pancreatic secretion, with the net result of reduced sphincter of Oddi flow in human infusion experiments, albeit with conflicting evidence on whether it induces sphincter contraction.[107]

2.3.4.4 Splanchnic Circulation

Exogenous administration of SST or its analogues has been shown to reduce splanchnic blood flow and pressure in dogs and man, although there is little information on the underlying mechanism of action.[118–120] It has been proposed as a treatment for bleeding oesophageal varices, although there is evidence that its pressure lowering effects are less potent in the cirrhotic patient and a recent Cochrane review concluded that it had no mortality benefit and only a modest reduction in transfusion requirements.[121,122]

2.3.5 Clinical Aspects

SST analogues are of considerable utility in the diagnosis and treatment of gastroentero-pancreatic NETs. As many moderately and well-differentiated NETs express receptors to SST, radio-nucleotide labeled SST analogues can be used in the diagnosis and staging of disease and for targeted radiotherapy. Palliative treatment with SST analogues, in the presence of symptomatic NETs, can control hormone-mediated symptoms including diarrhea, tachycardia, and flushing and has recently been shown to delay tumor progression.[123–126]

Other GI uses of SST analogues are based on limited case series or expert opinion and utilize their counterregulatory and antisecretory effects. The evidence is at present equivocal on the benefits of SST analogues in the prevention of postpancreatectomy cutaneous fistula, or the treatment of enterocutaneous fistula.[127–130] Long- and short-acting SST analogues have also been used for the management of congenital hyperinsulinemia and reactive hypoglycemia and accelerated intestinal transit after upper GI surgery, with mixed success.[131–137]

2.4. GHRELIN

Ghrelin was first identified in 1999 as the endogenous ligand for the growth hormone secretagogue receptor (GHS-R).[138] While primarily described as an orexigenic hormone through its hypothalamic actions, it has also diverse roles including as a GHS, promoter of adipogenesis, and suppressor of pancreatic insulin secretion (Fig. 2.3).[139–141]

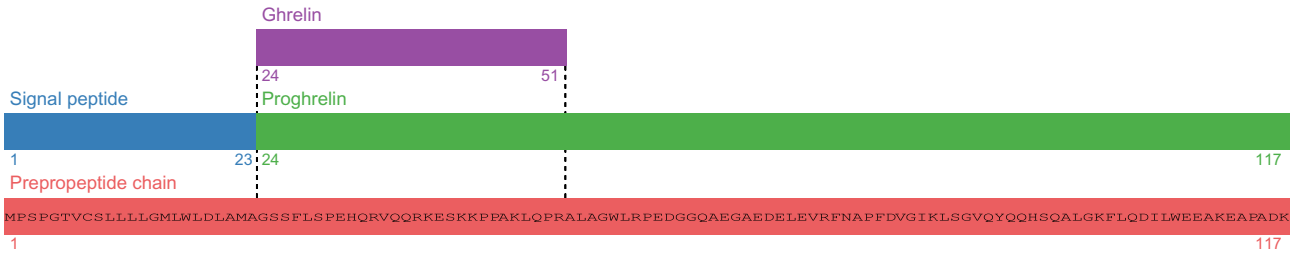


FIG. 2.3 Amino acid sequence of ghrelin prepropeptide and secreted peptide, derived from Uniprot/Swissprot database.

2.4.1 Synthesis

Ghrelin is primarily secreted by X/A-like cells of the stomach, but has also been identified in other tissues including duodenum, pancreas, lymphocytes, and the central nervous system.[142–145] Following total gastrectomy, circulating total and acyl ghrelin concentrations are undetectable, suggesting that the extra-gastric sources do not contribute significantly to circulating levels. It is encoded by the GHRL gene, located on chromosome 3p25-26. Translation of *Ghrl* mRNA produces a 117 amino acid preprohormone (preproghrelin), which is cleaved to the active 28 amino acid ghrelin by PC1/3. Ghrelin is modified by the addition of an octanoyl moiety to the hydroxyl group of the serine at position 3 of proghrelin catalyzed by ghrelin *O*-acyltransferase (GOAT, also membrane bound *O*-acyltransferase, MBOAT4), although it is unclear whether this step precedes or follows cleavage of proghrelin to ghrelin.[146–148] Acylation of serine 3 is essential for activity at the GHSR1a receptor, and acyl-ghrelin has historically been regarded as the active, and des-acyl-ghrelin the inactive, form of the peptide, although independent functions have been considered for the latter.[149] The fatty acid chain used for ghrelin acylation appears to derive from the diet.[150]

2.4.2 Secretion

Circulating concentrations of ghrelin are highest in the fasting state, with secretion suppressed by glucose and fat ingestion, and exercise, but less so by protein intake or gastric distension.[151–154] In vitro experimental evidence exists for direct sensing of fatty acids, glucose, and glutamate by X/A-like cells, for suppression of ghrelin secretion by insulin, leptin, and GLP-1 and for stimulation of ghrelin secretion by glucagon.[155–160] However, X/A cells are predominantly closed-type EECs making no contact with the gastric lumen, so they are likely regulated primarily by internal factors. Pharmacological experiments in rats demonstrated increased ghrelin secretion in response to muscarinic and beta-adrenergic activity and decreased secretion in response to alpha-adrenergic activity.[161] Plasma ghrelin concentrations increased in healthy humans in response to a cholinergic agonist and were suppressed by a muscarinic antagonist.[162] Vagotomy initially suppressed ghrelin secretion in rats, but seven days postvagotomy plasma ghrelin concentrations were elevated.[161] GLP-1 and PYY, independently and synergistically, suppressed ghrelin secretion in a study of 25 overweight men.[163] Investigation of FACS purified ghrelin secreting cells from mice demonstrated G-protein-coupled actions of α-CGRP, long- and short-chain fatty acids, lactate, SST, GIP, and α-MSH, but interestingly not PYY or GLP-1 on ghrelin secretion.[164]

2.4.3 Function

The majority of experimental studies on ghrelin have focused on the actions of its acyl form on the cognate G-protein-coupled receptor GHSR1a. Des-acyl-ghrelin is presumed to be an inactive metabolite, as it has no identified receptor or major physiological effects.[149,165]

Ghrelin plays a significant role in appetite regulation, stimulating neuropeptide Y/Agouti-related peptide (NPY/AgRP) neurons within the arcuate nucleus of the hypothalamus, in a counterregulatory fashion to leptin, to stimulate hunger and initiate feeding.[166–169] Multiple animal and human studies disagree on the relative importance of vagal-mediated ghrelin signaling versus direct central nervous system action of circulating ghrelin. In vivo animal models, and infusion experiments involving human participants who had undergone vagotomy, suggest that an intact vagus nerve is essential for the meal initiation effects of ghrelin.[170,171] This is consistent with the finding that ghrelin increased the sensitivity of gastric vagal mechanosensory nerves to stretch.[172] However, it has been demonstrated that central nervous system and intraperitoneal administration of ghrelin in rats resulted in similar feeding behavior, and a randomized controlled trial of ghrelin administration in patients after total gastrectomy (i.e., with minimal circulating ghrelin and a truncal vagotomy) improved food intake and reduced body weight loss.[166,173]

Ghrelin stimulates pituitary growth hormone secretion through its direct actions on GHSR1a.[174,175] In particular, it seems that ghrelin is a key stimulus of growth hormone-mediated gluconeogenesis in the fasting state.[176] In *GOAT* or ghrelin knockout mice, prolonged fasting resulted in profound hypoglycaemia, associated with a reduced growth hormone response, which was reversed on infusion of acyl-ghrelin.[177,178]

Ghrelin also influences glucose homeostasis in a growth hormone-independent fashion. Acyl-ghrelin indirectly inhibited glucose-mediated insulin and glucagon secretion in mouse models by stimulation of SST secretion from pancreatic islet delta cells.[140,179] Although a population of pancreatic islet ghrelin secreting cells has been described, supporting the possibility that ghrelin might also act in a paracrine fashion within pancreatic islets, the physiological importance of this finding remains controversial.[180] Examination of peripheral insulin sensitivity using an euglycaemic hyperinsulinemic clamp with and without exogenous ghrelin, in hypopituitary patients on stable doses of exogenous growth hormone, demonstrated that ghrelin increased peripheral insulin resistance independent of growth hormone.[181,182] Both acylated and nonacylated forms of ghrelin have been demonstrated to stimulate fat accumulation in human visceral adipocytes, through enhanced PPARgamma and SREBP1 signaling.[141]

The more direct roles of ghrelin in GI function include stimulation of gastric motility and increasing gastric acid secretion in a vagus and 5-HT-dependent fashion.[183,184]

2.4.4 Clinical Aspects

The diverse and as yet not completely understood roles of ghrelin in appetite regulation, glucose and energy homeostasis, GI motility, and higher-order cognitive functioning make it a fertile area for ongoing research into disease pathology and pharmacotherapy.

The orexigenic actions of ghrelin, an inverse correlation between fasting ghrelin concentrations and BMI and observations of reduced postprandial suppression in obese humans make it an attractive target for the treatment of obesity.[185–189] Plasma ghrelin concentrations are reduced in parallel with weight loss after sleeve gastrectomy, but variably increase after Roux-en-Y gastric bypass, suggesting it is not the primary regulator of appetite and body mass in humans.[167,190] Animal knockout and pharmacological models targeting ghrelin, GOAT, and GHSR1 have resulted in conflicting results regarding metabolic homeostasis and protection from diet-induced obesity, with the most convincing results showing a reduced incidence of diet-induced obesity in *Ghsr1*$^{-/-}$ mice.[191–195] One phase I/IIa trial of an antighrelin vaccine was halted due to lack of efficacy.[196]

Conversely, human studies of pharmacological augmentation of the ghrelin axis have proved more fruitful. Partial and total gastrectomy and esophagectomy result in reduced plasma ghrelin concentrations and simultaneous severing of afferent vagal fibers. This is often associated with reduced appetite, weight loss, and impaired quality of life.[197] One small trial of synthetic ghrelin in postgastrectomy patients yielded positive results on food intake.[173] GHSR1a agonists are also in late-phase clinical trials for cancer-related cachexia, with promising early results.[198] It has been proposed that the hyperphagia of Prader-Willi syndrome may be mediated by hyperghrelinemia, although recent evidence suggests the onset of elevated plasma ghrelin concentrations significantly predates hyperphagia and may be unrelated to the phenotype.[199]

The prokinetic effects of ghrelin offer a potential drug target for GI motility disorders. One small study has shown improved gastric emptying following the administration of a GHSR agonist in patients with diabetic gastroparesis, and the drug has entered phase 3 trials.[200]

2.5. CCK

Cholecystokinin (CCK) is widely distributed in the central and peripheral nervous systems as a neurotransmitter, and in I cells of the duodenal mucosa from which it is secreted into the bloodstream.[201,202] There are multiple posttranslational products of the CCK gene (including CCK-83, -58, -33, -22, -12, -8, and -5) that vary in length but share a common amidated C-terminus.[203,204] The multiple CCK peptides have a diverse range of functions, including stimulation of gallbladder contraction and pancreatic exocrine secretion, inhibition of gastric emptying and acid secretion, and signaling of satiety (Fig. 2.4).

2.5.1 Synthesis

The discovery of CCK dates back to the suggestion of a hormonal mechanism for gallbladder contraction in 1928, supported by an experiment wherein intestinal mucosal extracts were infused into dogs, cats, and guinea pigs and resulted in gallbladder contraction.[205] The peptide sequence of CCK was first described in 1968, with the C-terminal pentapeptide Gly-Trp-Met-Thr-Phe conserved across all CCK and gastrin peptides.[202]

The CCK gene is located at chromosome 3p22.1.[206] PreproCCK is a 115 amino acid protein, with an N-terminal signal sequence and spacer sequence followed by the bioactive domain.[207] Posttranslational processing involves cleavage of the signal peptide and addition of a C-terminal amide group, followed by cleavage of the 83 amino acid peptides at basic amino acid residues, most likely by PC1/3, although PC2 and PC5 have been implicated in the processing of CCK in rat brain.[207–209] Addition of a sulfate group to the tyrosine residue seven amino acids from the C-terminus confers activity at the CCK 1 receptor (CCK1R), whereas sulfated and nonsulfated peptides are equally active at the CCK 2 receptor (CCK2R), which also acts as the gastrin receptor.[210]

Measurement of CCK concentrations in plasma is challenging, with acknowledged discrepancies in the sensitivity of immunoassays to different CCK peptides, and no gold standard test.[211] High-pressure liquid chromatography extraction of human intestinal lysates, and of plasma after a

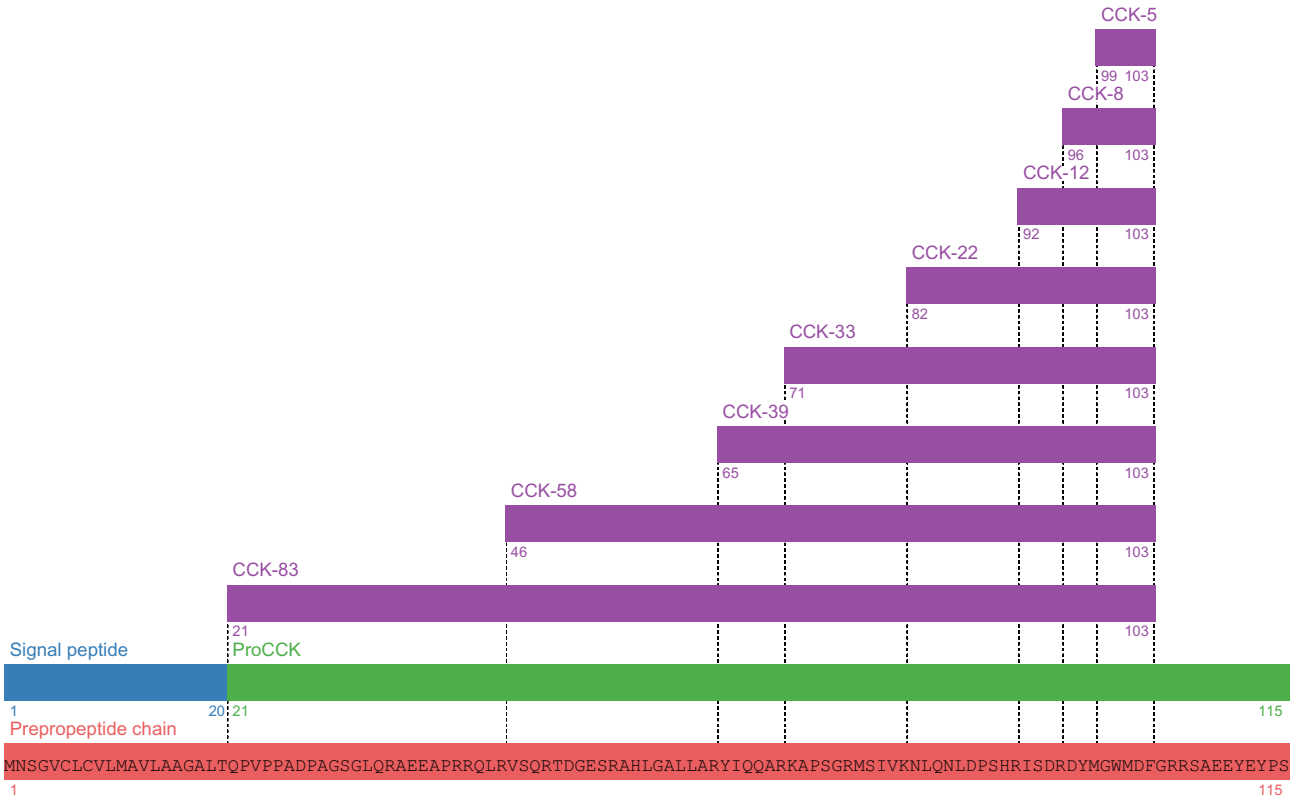


FIG. 2.4 Amino acid sequence of cholecystokinin prepropeptide and secreted peptides, derived from Uniprot/Swissprot database.

meal test, revealed the 58 amino acid peptide to be the most abundant in man, with minor amounts of the 39 and 8 amino acid peptides.[211,212] These results must, however, be interpreted in light of the finding that the in vivo half-life and hepatic clearance of CCK-8 is markedly faster than that of CCK-58.[213,214] It is probably reasonable to regard current CCK assays as satisfactory for the examination of relative concentrations of plasma total CCK, but given the variability in the rate of degradation and assay sensitivity between peptides, caution must be exercised when interpreting data attempting to assess the relative concentrations of different length CCK peptides.

It is acknowledged that CCK is widely distributed within the brain, with CCK-8 and CCK-5 being the primary neurotransmitter CCK peptides.[215] In depth discussion of the role of CCK as a central neurotransmitter is beyond the remit of this chapter.

2.5.2 Secretion

CCK is secreted into the circulation from open-type I cells in the duodenal and jejunal mucosa. Plasma concentrations of total CCK rise approximately three- to sevenfold in response to a mixed meal.[212,216] While it is clear that CCK is secreted in a nutrient specific fashion in man, our knowledge of the underlying receptors is heavily reliant on limited pharmacological experiments and animal and in vitro data. Lipids are the most potent stimulus of CCK secretion, followed by proteins, with only small effects triggered by intraduodenal carbohydrates.[217–220] Intraduodenal lipid stimulation of CCK secretion is dependent upon medium and long-chain fatty acids acting via FFA1 (GPR40), FFA4 (GPR120), and possibly CD36, with limited effects of short-chain fatty acids on plasma CCK concentration.[221–225] The mechanisms by which digested proteins stimulate CCK secretion include activation of PEPT1 and the calcium-sensing receptor (CaSR).[226–229] It appears that carbohydrates play a more limited role in CCK secretion.[230,231] In one small human study, intraduodenal acidification in the absence of nutrients did not stimulate CCK secretion.[219]

It remains unclear to what extent vagal tone influences CCK secretion, if at all. Studies in vagotomized humans have either failed to take account of altered gastric transit, or found conflicting results, and one limited study in rats suggested that vagal activation could induce CCK secretion, but did not explore this in a physiological fashion.[232–234]

2.5.3 Function

CCK acts via the G-protein-coupled receptors CCK1R and CCK2R (previously CCKAR and CCKBR). CCK1R has a 500–1000-fold greater affinity for sulfated than nonsulfated