

A microscopic image of cells, likely stained with a combination of red and blue dyes. The red staining highlights certain cellular structures, possibly nuclei or specific organelles, while the blue staining provides a general view of the cell morphology. The background is dark, making the stained cells stand out.

Benjamin Lewin

genes

VIII

Genes VII

Benjamin Lewin

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York

Athens Auckland Bangkok Bogotá Buenos Aires Calcutta
Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paulo Singapore Taipei Tokyo Toronto Warsaw
with associated companies in Berlin Ibadan

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Oxford University Press and Cell Press, 2000

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2000, reprinted 2000,2002

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose this same condition on any acquirer

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
(Data applied for)

ISBN 0-19-879277-8 (Pbk)

0-19-879276-X (Hbk)

Typeset by J&L Composition Ltd, Filey, North Yorkshire
Printed in The Italy by LegoPrint

Genes VII

Preface

This edition of GENES marks the most significant reorganization for several editions. It is time to acknowledge that the power of direct analysis of the genome is making a significant difference to our approach. This is now recognized in the first part of GENES, which starts with the concept of the gene as a segment of DNA coding for protein and then proceeds directly to the characterization of the genome in terms of its content of genes.

The rapidity of our advance in analyzing the genome has led to some confusion about means and ends. A genome is no more than a sum of the genes it contains, although the number and complexity of these genes gives us important insights into the nature of their connection into regulatory networks. The focus of this book, as its name suggests, therefore remains firmly on the gene.

The thesis of GENES is that only by understanding the structure and function of the gene itself will we be able in turn to understand the operation of the genome as a whole. Although the emphasis has shifted to the characterization of eukaryotic genes, and therefore to their analysis by the direct techniques of molecular biology rather than the subtlety of genetics, the classical approach remains intellectually penetrating. It remains an aim of this book to integrate both approaches in the context of a unified approach to prokaryotes and eukaryotes.

We consider the gene from all aspects: the basic forms that it takes, the numbers and relationships among genes in a genome, their packaging into chromosomes, the process of gene expression from transcription through translation, the reproduction and safeguarding of the structure of the gene, and finally some aspects of the overall circuitry through which genotype determines phenotype.

Almost two decades after the first edition of GENES, this latest edition is a celebration of the remarkable progress during that time, and, I hope, an indication of the future to come.

B. L.
Cambridge, Massachusetts
January 2000

Outline

Part 1 Genes	1	16 Retroviruses and retroposons	485
1 Genes are DNA	3	17 Rearrangement of DNA	507
2 From genes to genomes	37		
3 How many genes are there?	67		
4 Clusters and repeats	89		

Part 2 Proteins	117	Part 5 The nucleus	543
5 Messenger DNA	119	18 Chromosomes	545
6 Protein synthesis	139	19 Nucleosomes	567
7 Using the genetic code	167	20 Initiation of transcription	617
8 Protein localization	191	21 Regulation of transcription	649
		22 Nuclear splicing	685
		23 Catalytic RNA	719
		24 Immune diversity	741

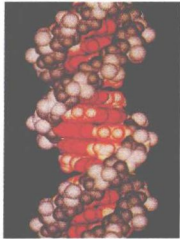
Part 3 mRNA	231	Part 6 Cells	773
9 Transcription	232	25 Protein trafficking	775
10 The operon	273	26 Signal transduction	801
11 Phage strategies	319	27 Cell cycle and growth regulation	835
		28 Oncogenes and cancer	875
		29 Gradients, cascades, and signaling pathways	913

Part 4 DNA	347		
12 The replicon	349		
13 DNA replication	385		
14 Recombination and repair	415		
15 Transposons	457		

Glossary	953
-----------------	------------

Index	973
--------------	------------

Contents



Part 1 Genes 1

1 Genes are DNA 3

DNA is the genetic material	4
DNA is a double helix	6
DNA replication is semiconservative	9
Nucleic acids hybridize by base pairing	12
Mutations change the sequence of DNA	14
Mutations are concentrated at hotspots	17
A cistron is a single stretch of DNA	19
The nature of multiple alleles	21
Recombination occurs by physical exchange of DNA	23
The genetic code is triplet	25
Bacterial genes and proteins are colinear	27
<i>cis</i> -acting sites and <i>trans</i> -acting molecules	30
Genetic information can be provided by DNA or RNA	31

2 From genes to genomes 37

Genes can be mapped by restriction cleavage	38
How variable are individual genomes?	41
Eukaryotic genes are often interrupted	44
Organization of interrupted genes may be conserved	46
Exon sequences are conserved but introns vary	48
Genes can be isolated by the conservation of exons	49
Genes show a wide distribution of sizes	53
Some DNA sequences code for more than one protein	55
How did interrupted genes evolve?	58
The scope of the paradigm	63

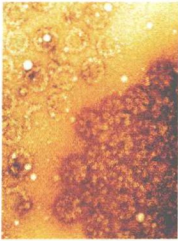
3 How many genes are there? 67

Why are genomes so large?	68
Eukaryotic genomes have several sequence components	69
Most structural genes lie in nonrepetitive DNA	73

Total gene number is known for several organisms	75
How many genes are essential?	76
How many genes are expressed?	78
Organelles have DNA	80
Organelle genomes are circular DNAs that code for organelle proteins	82
Mitochondrial DNA codes for few proteins	83
The chloroplast genome codes for ~100 proteins and RNAs	84

4 Clusters and repeats **89**

Gene clusters are formed by duplication and divergence	90
Sequence divergence is the basis for the evolutionary clock	92
Pseudogenes are dead ends of evolution	95
Unequal crossing-over rearranges gene clusters	97
Genes for rRNA form a repeated tandem unit	100
Crossover fixation could maintain identical repeats	103
Satellite DNAs often lie in heterochromatin	106
Arthropod satellites have very short identical repeats	108
Mammalian satellites consist of hierarchical repeats	108
Minisatellites are useful for genetic mapping	113



Part 2 Proteins **117**

5 Messenger RNA **119**

Transfer RNA is the adapter	120
Messenger RNA is translated by ribosomes	124
The life cycle of messenger RNA	126
Translation of eukaryotic mRNA	129
The 5' end of eukaryotic mRNA is capped	130
The 3' terminus is polyadenylated	131
Degradation pathways for mRNA	133

6 Protein synthesis **139**

The stages of protein synthesis	140
Initiation in bacteria needs 30S subunits and accessory factors	143
A special initiator tRNA starts the polypeptide chain	144
Initiation involves base pairing between mRNA and rRNA	147
Small subunits scan for initiation sites on eukaryotic mRNA	149
Elongation factor T loads aminoacyl-tRNA into the A site	152
Translocation moves the ribosome	154
Three codons terminate protein synthesis	157
Ribosomes have several active centers	159

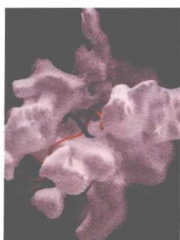
The role of ribosomal RNA in protein synthesis 162

7 Using the genetic code 167

Codon-anticodon recognition involves wobbling 169
 tRNA contains modified bases that influence its pairing properties 170
 The genetic code is altered in mitochondria 174
 tRNAs are charged with amino acids by individual synthetases 176
 Accuracy depends on proofreading 179
 Suppressor tRNAs have mutated anticodons that read new codons 182
 The accuracy of translation 185
 tRNA may influence the reading frame 187

8 Protein localization 191

Chaperones may be required for protein folding 194
 Post-translational membrane insertion depends on leader sequences 198
 A hierarchy of sequences determines location within organelles 201
 The translocation apparatus interacts with signal sequences 203
 How do proteins enter and leave membranes? 208
 Anchor signals are needed for membrane residence 212
 Bacteria use both co-translational and post-translational translocation 215
 Pores control nuclear ingress and egress 216
 Protein degradation by proteasomes 224



Part 3 mRNA 231

9 Transcription 232

Transcription is catalyzed by RNA polymerase 234
 RNA polymerase consists of multiple subunits 238
 Sigma factor controls binding to DNA 240
 Promoter recognition depends on consensus sequences 244
 RNA polymerase binds to one face of DNA 247
 Substitution of sigma factors may control initiation 250
 Sigma factors may be organized into cascades 253
 Bacterial RNA polymerase has two modes of termination 257
 How does rho factor work? 259
 Antitermination depends on specific sites 262
 More subunits for RNA polymerase 267

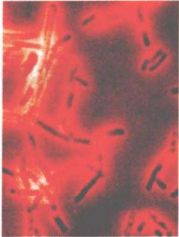
10 The operon 273

Structural gene clusters are coordinately controlled 275
 Repressor is controlled by a small molecule inducer 277

Mutations identify the operator and the regulator gene	280
Repressor protein binds to the operator and is released by inducer	283
The specificity of protein-DNA interactions	288
Repression can occur at multiple loci	291
Distinguishing positive and negative control	292
Catabolite repression involves positive regulation at the promoter	294
Adverse growth conditions provoke the stringent response	298
Autogenous control may occur at translation	301
Alternative secondary structures control attenuation	306
Small RNA molecules can regulate translation	312

11 Phage strategies 319

Lytic development is controlled by a cascade	321
Functional clustering in phages T7 and T4	324
The lambda lytic cascade relies on antitermination	325
Lysogeny is maintained by an autogenous circuit	330
The DNA-binding form of repressor is a dimer	333
Repressor binds cooperatively at each operator using a helix-turn-helix motif	334
How is repressor synthesis established?	339
A second repressor is needed for lytic infection	342
A delicate balance: lysogeny versus lysis	344



Part 4 DNA 347

12 The replicon 349

Origins can be mapped by autoradiography and electrophoresis	350
The bacterial genome is a single circular replicon	354
Each eukaryotic chromosome contains many replicons	355
Isolating the origins of yeast replicons	357
D loops maintain mitochondrial origins	358
The problem of linear replicons	361
Rolling circles produce multimers of a replicon	363
Single-stranded genomes are generated for bacterial conjugation	366
Connecting bacterial replication to the cell cycle	370
Cell division and chromosome segregation	371
Multiple systems ensure plasmid survival in bacterial populations	376
Plasmid incompatibility is connected with copy number	378

13 DNA replication 385

DNA polymerases: the enzymes that make DNA	386
DNA synthesis is semidiscontinuous and primed by RNA	390

Coordinating synthesis of the lagging and leading strands	393
The replication apparatus of phage T4	401
Creating the replication forks at an origin	402
Common events in priming replication at the origin	405
Does methylation at the origin regulate initiation?	406
Licensing factor controls eukaryotic rereplication	408

14 Recombination and repair **415**

Breakage and reunion involves heteroduplex DNA	418
Double-strand breaks initiate recombination	420
Double-strand breaks initiate synapsis	422
Bacterial recombination involves single-strand assimilation	426
Gene conversion accounts for interallelic recombination	431
Topological manipulation of DNA	432
Specialized recombination involves breakage and reunion at specific sites	437
Repair systems correct damage to DNA	441
Excision repair systems in <i>E. coli</i>	444
Controlling the direction of mismatch repair	446
Retrieval systems in <i>E. coli</i>	449
RecA triggers the SOS system	450
Eukaryotic repair systems	452

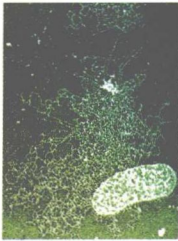
15 Transposons **457**

Insertion sequences are simple transposition modules	458
Composite transposons have IS modules	460
Transposition occurs by both replicative and nonreplicative mechanisms	462
Common intermediates for transposition	465
Replicative transposition proceeds through a cointegrate	467
Nonreplicative transposition proceeds by breakage and reunion	469
TnA transposition requires transposase and resolvase	470
Transposition of Tn10 has multiple controls	472
Controlling elements in maize cause breakage and rearrangements	473
Controlling elements in maize form families of transposons	476
<i>Spm</i> elements influence gene expression	478
The role of transposable elements in hybrid dysgenesis	479

16 Retroviruses and retroposons **485**

The retrovirus life cycle involves transposition-like events	486
Retroviruses may transduce cellular sequences	494
Yeast Ty elements resemble retroviruses	496
Many transposable elements reside in <i>D. melanogaster</i>	499
Retroposons fall into two classes	500

17 Rearrangement of DNA	507
The mating pathway is triggered by signal transduction	508
Yeast can switch silent and active loci for mating type	511
Silent cassettes at <i>HML</i> and <i>HMR</i> are repressed	515
Unidirectional transposition is initiated by the recipient <i>MAT</i> locus	516
Regulation of <i>HO</i> expression	518
Trypanosomes rearrange DNA to express new surface antigens	519
Interaction of Ti plasmid DNA with the plant genome	524
Selection of amplified genomic sequences	530
Exogenous sequences can be introduced into cells and animals by transfection	533



Part 5 The Nucleus 543

18 Chromosomes	545
Condensing viral genomes into their coats	546
The bacterial genome is a nucleoid with many supercoiled loops	549
Loops, domains, and scaffolds in eukaryotic DNA	551
The contrast between interphase chromatin and mitotic chromosomes	553
The extended state of lampbrush chromosomes	556
Transcription disrupts the structure of polytene chromosomes	557
The eukaryotic chromosome as a segregation device	559
Telomeres seal the ends of chromosomes	562
19 Nucleosomes	567
The nucleosome is the subunit of all chromatin	568
DNA is coiled in arrays of nucleosomes	571
DNA structure varies on the nucleosomal surface	574
Supercoiling and the periodicity of DNA	576
The path of nucleosomes in the chromatin fiber	578
Organization of the histone octamer	580
Reproduction of chromatin requires assembly of nucleosomes	583
Do nucleosomes lie at specific positions?	586
Are transcribed genes organized in nucleosomes?	589
DNAase hypersensitive sites change chromatin structure	593
Domains define regions that contain active genes	595
Heterochromatin depends on interactions with histones	597
Global changes in X chromosomes	601
Methylation is responsible for imprinting	603
Modes of epigenetic inheritance	606

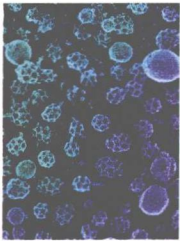
20 Initiation of transcription	617
Eukaryotic RNA polymerases consist of many subunits	619
Promoter elements are defined by mutations and footprinting	620
RNA polymerase I has a bipartite promoter	622
RNA polymerase III uses both downstream and upstream promoters	624
The basal apparatus consists of RNA polymerase II and general factors	627
A connection between transcription and repair	632
Promoters for RNA polymerase II have short sequence elements	634
Enhancers contain bidirectional elements that assist initiation	637
Independent domains bind DNA and activate transcription	641
Interaction of upstream factors with the basal apparatus	644

21 Regulation of transcription	649
Response elements identify genes under common regulation	650
There are many types of DNA-binding domains	652
A zinc finger motif is a DNA-binding domain	654
Steroid receptors have several independent domains	656
Homeodomains bind related targets in DNA	660
Helix-loop-helix proteins interact by combinatorial association	662
Leucine zippers are involved in dimer formation	664
Chromatin remodeling is an active process	666
Histone acetylation and deacetylation control chromatin activity	669
Polycomb and trithorax are antagonistic repressors and activators	672
Long range regulation and insulation of domains	674
Gene expression is associated with demethylation	678

22 Nuclear splicing	685
Nuclear splice junctions are interchangeable but are read in pairs	687
Nuclear splicing proceeds through a lariat	690
The spliceosome contains snRNAs	692
Group II introns autosplice via lariat formation	699
Alternative splicing involves differential use of splice junctions	702
<i>cis</i> -splicing and <i>trans</i> -splicing reactions	705
Yeast tRNA splicing involves cutting and rejoining	707
3' ends are generated by termination and by cleavage reactions	711

23 Catalytic RNA	719
Group I introns undertake self-splicing by transesterification	720
Group I introns form a characteristic secondary structure	723
Ribozymes have various catalytic activities	725
Some introns code for proteins that sponsor mobility	728
RNA can have ribonuclease activities	731

RNA editing utilizes information from several sources	733
<hr/>	
24 Immune diversity	741
Clonal selection amplifies lymphocytes that respond to individual antigens	743
Immunoglobulin genes are assembled from their parts in lymphocytes	745
The diversity of germline information	750
Recombination between V and C gene segments generates deletions and rearrangements	752
Allelic exclusion is triggered by productive rearrangement	757
DNA recombination causes class switching	758
Early heavy chain expression can be changed by RNA processing	760
Somatic mutation generates additional diversity	761
B cell development and memory	762
T-cell receptors are related to immunoglobulins	764
The major histocompatibility locus codes for many genes of the immune system	768
<hr/>	
Part 6 Cells	773
<hr/>	
25 Protein trafficking	775
Oligosaccharides are added to proteins in the ER and Golgi	778
Coated vesicles transport both exported and imported proteins	781
Budding and fusion reactions	786
Protein localization depends on further signals	791
Receptors recycle via endocytosis	794
<hr/>	
26 Signal transduction	801
Carriers and channels form water-soluble paths through the membrane	804
G proteins may activate or inhibit target proteins	809
Protein tyrosine kinases induce phosphorylation cascades	811
The Ras/MAPK pathway	816
Activating MAP kinase pathways	822
Cyclic AMP and activation of CREB	827
The JAK-STAT pathway	828
TGF β signals through Smads	830
<hr/>	
27 Cell cycle and growth regulation	835
Cycle progression depends on discrete control points	836
M phase kinase regulates entry into mitosis	840
Protein phosphorylation and dephosphorylation control the cell cycle	843
Cdc2 is the key regulator in yeasts	844
<i>CDC28</i> acts at both START and mitosis in <i>S. cerevisiae</i>	852
The animal cell cycle is controlled by many cdk-cyclin complexes	855



G0/G1 and G1/S transitions involve cdk inhibitors	858
Protein degradation is important in mitosis	861
Reorganization of the cell at mitosis	864
Apoptosis	866
<hr/>	
28 Oncogenes and cancer	875
Transforming viruses carry oncogenes	878
Retroviral oncogenes have cellular counterparts	881
Ras proto-oncogenes can be activated by mutation	883
Insertion, translocation, or amplification may activate proto-oncogenes	886
Oncogenes code for components of signal transduction cascades	890
Growth factor receptor kinases and cytoplasmic tyrosine kinases	892
Oncoproteins may regulate gene expression	896
RB is a tumor suppressor that controls the cell cycle	899
Tumor suppressor p53 suppresses growth or triggers apoptosis	901
Immortalization and transformation	906
<hr/>	
29 Gradients, cascades, and signaling pathways	913
Fly development uses a cascade of transcription factors	914
A gradient must be converted into discrete compartments	915
Maternal gene products establish gradients in early embryogenesis	917
Anterior-posterior development uses localized gene regulators	920
Dorsal-ventral development uses localized receptor-ligand interactions	923
TGF β /BMPs are diffusible morphogens	929
Cell fate is determined by compartments that form by the blastoderm stage	931
The wingless/wnt signaling pathway	938
Complex loci are extremely large and involved in regulation	940
The homeobox is a common coding motif in homeotic genes	946
<hr/>	
Glossary	953
<hr/>	
Index	973

Part 1

Genes

