

GENETIC VARIATION

A LABORATORY MANUAL



EDITED BY

M.P. Weiner • S.B. Gabriel • J.C. Stephens

IPR2024-01419

Illumina Inc., v. Molecular Loop Biosciences, Inc.

Molecular Loop's Exhibit 2022

EX2022, Page 1

Genetic Variation

A LABORATORY MANUAL

ALSO FROM COLD SPRING HARBOR LABORATORY PRESS

RELATED LABORATORY MANUALS

Basic Methods in Microscopy: Protocols and Concepts from Cells—A Laboratory Manual

Gene Transfer: Delivery and Expression of DNA and RNA—A Laboratory Manual

Molecular Cloning: A Laboratory Manual

The Condensed Protocols from Molecular Cloning: A Laboratory Manual

Protein-Protein Interactions: A Molecular Cloning Manual, Second Edition

WEB SITE

www.cshprotocols.org

OTHER RELATED TITLES

At the Bench: A Laboratory Navigator, Updated Edition

Binding and Kinetics for Molecular Biologists

Bioinformatics: Sequence and Genome Analysis, Second Edition

Epigenetics

Experimental Design for Biologists

Lab Math: A Handbook of Measurements, Calculations, and Other Quantitative Skills for Use at the Bench

Lab Ref, Volume 1: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench

Lab Ref, Volume 2: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench

The Dog and Its Genome

Genetic Variation

A LABORATORY MANUAL

EDITED BY

Michael P. Weiner

RainDance Technologies

Stacey B. Gabriel

The Broad Institute of MIT and Harvard

J. Claiborne Stephens

Motif BioSciences Inc.

<http://www.cshprotocols.org>



COLD SPRING HARBOR LABORATORY PRESS

Cold Spring Harbor, New York • <http://www.cshlpress.com>

All rights reserved
© 2007 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Printed in the United States of America

Publisher	John Inglis
Acquisition Editor and Managing Editor	Kaaren Janssen
Development, Marketing, & Sales Director	Jan Argentine
Developmental Editors	Irene Pech, Martin Winer, and Laurie Goodman
Project Coordinator	Inez Sialiano
Production Manager	Denise Weiss
Production Editor	Patricia Barker
Desktop Editor	Lauren Heller
Sales Account Manager	Elizabeth Powers
Cover Designer	Lewis Agrell

Front cover artwork: (*Top left*) Confocal imaging of cells in mitosis (tubulin stained green, chromosome stained orange), from P. Waadsworth, in *Live cell imaging: A laboratory manual* (ed. R.D. Goldman and D. Spector); (*top center*) diagram redrawn from Chapter 3 by D. Evans and S. Purcell; (*top right*) *Mus musculus*, Gary Henderson; (*middle left*) J.D. Watson lecture note #1 entitled "What is Life," courtesy James D. Watson Collection, Cold Spring Harbor Laboratory Archives; (*middle center*) double helix by Michael P. Weiner; (*middle right*) phylogeny, Michael P. Weiner; (*bottom left*) data from an experiment showing the expression of thousands of genes on a single GeneChip® probe array, image courtesy of Affymetrix; (*bottom center*) Bonobo chimpanzee ©2005 William H. Calvin, University of Washington; (*bottom right*) maize photographs, courtesy Cold Spring Harbor Laboratory Archives.

Library of Congress Cataloging-in-Publication Data

Genetic variation : a laboratory manual / edited by Michael P. Weiner,
Stacey Gabriel, J. Claiborne Stephens.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-87969-780-8 (pbk. : alk. paper) -- ISBN 978-0-87969-779-2
(printed hardcover : alk. paper)
1. Molecular genetics--Laboratory manuals. 2. Variation
(Biology)--Laboratory manuals. I. Weiner, Michael P., 1958- II. Gabriel,
Stacey. III. Stephens, J. Claiborne. IV. Title.

QH440.5.G455 2007
572.8078--dc22

2007023448

10 9 8 7 6 5 4 3 2 1

Students and researchers using the procedures in this manual do so at their own risk. Cold Spring Harbor Laboratory makes no representations or warranties with respect to the material set forth in this manual and has no liability in connection with the use of these materials. All registered trademarks, trade names, and brand names mentioned in this book are the property of the respective owners. Readers should please consult individual manufacturers and other resources for current and specific product information.

With the exception of those suppliers listed in the text with their addresses, all suppliers mentioned in this manual can be found on the BioSupplyNet Web site at <http://www.biosupplynet.com/>.

All World Wide Web addresses are accurate to the best of our knowledge at the time of printing.

Procedures for the humane treatment of animals must be observed at all times. Check with the local animal facility for guidelines.

Certain experimental procedures in this manual may be the subject of national or local legislation or agency restrictions. Users of this manual are responsible for obtaining the relevant permissions, certificates, or licenses in these cases. Neither the authors of this manual nor Cold Spring Harbor Laboratory assumes any responsibility for failure of a user to do so.

The polymerase chain reaction process is covered by certain patent and proprietary rights. Users of this manual are responsible for obtaining any licenses necessary to practice PCR or to commercialize the results of such use. COLD SPRING HARBOR LABORATORY MAKES NO REPRESENTATION THAT USE OF THE INFORMATION IN THIS MANUAL WILL NOT INFRINGE ANY PATENT OR OTHER PROPRIETARY RIGHT.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press, provided that the appropriate fee is paid directly to the Copyright Clearance Center (CCC). Write or call CCC at 222 Rosewood Drive, Danvers, MA 01923 (978-750-8400) for information about fees and regulations. Prior to photocopying items for educational classroom use, contact CCC at the above address. Additional information on CCC can be obtained at CCC Online at <http://www.copyright.com/>.

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory Press, 500 Sunnyside Blvd., Woodbury, New York 11797-2924. Phone: 1-800-843-4388 in the Continental U.S. and Canada. All other locations: (516) 422-4100. FAX: (516) 422-4097. E-mail: cshpress@cshl.edu. For a complete catalog of all Cold Spring Harbor Laboratory Press publications, visit our World Wide Web site <http://www.cshlpress.com/>.

*Our collective gratitude goes out to the investigators who strive
to alleviate human suffering through an increased
understanding of the biological principles of the world around us*

Contents

Foreword, *xiii*

Preface, *xv*

Contributors, *xvii*

- 1 Ethical Issues in Human Genetic Research: The Global Experience, 1
K. Arnold and J. van der Walt

SECTION 1: STUDY DESIGN

Section 1 covers issues that deal with the importance of study design in addressing the sheer volume of data to be evaluated. Chapter 2 highlights that, for human genetic analyses, certain populations may have more favorable characteristics for associating traits with genes or precise genomic regions (e.g., disease-gene mapping). The size of the sample from a given population is critical for ensuring that a study is adequately powered to find a genetic association when it exists. Chapter 3 considers sample size and all other relevant factors for determining the power of genetic studies. Chapter 4 reviews the analytical methodology available for human genetic studies. In particular, strengths and weaknesses of linkage mapping, where pedigrees play an essential role, are contrasted with genetic association analyses, in which study subjects are typically unrelated. Both types of studies have their place, and the choice of method is strongly dependent on the type of disease or trait being addressed. Chapter 5 gives detailed insights into one of the most important resources available: the dbSNP database, a resource that maintains the current inventory of known variable sites in the human genome. Chapter 5 is an introduction to accessing the multitude of data cataloged in dbSNP. Section 1 closes with Chapter 6, which provides a user's guide to the suite of tools for manipulating HapMap data. The HapMap data set is a unique resource in that it provides an enormous, consistent set of genotypic data (3.9 million SNPs) for 270 individuals from four world populations. These data give very high-resolution detail of the genetic architecture of these four populations and whet our appetite for more samples in these and additional populations!

- 2 Population Choice as a Consideration for Genetic Analysis Study Design, 11
J.C. Stephens and M. Bamshad
- 3 Power Calculations, 19
D.M. Evans and S. Purcell
- 4 Genetic Analysis: Moving between Linkage and Association, 31
A.V. Smith
- 5 NCBI dbSNP Database: Content and Searching, 41
M.L. Feolo and S.T. Sherry
- 6 Using the HapMap Web Site, 63
A.V. Smith

SECTION 2: LABORATORY PROTOCOLS

The raw material of genetic research is DNA; individual-to-individual variation is the prism through which it is analyzed. This section deals with two critical aspects in the study of DNA variation: sample preparation and methods of common variation analysis. DNA of sufficient quality and quantity for the current generation of large-scale experimentation is often limiting. To address this, methods of DNA extraction, amplification, and cell culture are described. Parts 2 and 3 of this section turn to the generation of genotype data—beginning with the simplest and most abundant form of sequence variation, single-nucleotide polymorphism, and conclude with more complex DNA variation and the emerging study of DNA copy number variation. The specific topics are not meant to be an exhaustive description of all options for sample preparation and genotyping. Rather, the intent is to sample the wide range of sample type (cell culture and amplified DNA) and experimental scale (small-scale “home brew” genotyping vs. genome-wide standard products) that contribute to current genetic studies.

PART 1: PREPARATION OF DNA AND RNA

- 7 Isolation of Plant DNA for Genotyping Analysis, 81
N.M. Springer
 - Protocol 1: Isolation of Plant DNA for PCR and Genotyping Using Organic Extraction and CTAB, 82
N.M. Springer
 - Protocol 2: DNA Extraction from Freeze-dried Plant Tissue with CTAB in a 96-Well Format, 84
A.-P. Hsia, H.D. Chen, K. Ohtsu, and P.S. Schnable
 - Protocol 3: Purification of *Arabidopsis* DNA in 96-Well Plates Using the PUREGENE DNA Purification Kit, 87
Y. Li
- 8 Preparing RNA from Plant Tissues, 91
A.-P. Hsia, H.D. Chen, K. Ohtsu, and P.S. Schnable
 - Protocol 1: Maize Tissue Preparation and Extraction of RNA from Target Cells, 93
 - Protocol 2: T7-based RNA Amplification from Maize SAM, 97
- 9 Preparing DNA from Mammalian Sources, 103
A. Sahota, A.I. Brooks, and J.A. Tischfield
 - Protocol 1: Preparing DNA from Cell Pellets, 107
A. Sahota, A.I. Brooks, and J.A. Tischfield
 - Protocol 2: Preparing DNA from Fixed Tissue: Extraction and Whole-Genome Amplification, 110
C. Brueck, C. Brown, S. Michalik, D. Vassar-Nieto, E. Mueller, and G. Davis
 - Protocol 3: DNA Isolation from Rat Tail or Ear, 114
E. Cuppen
 - Protocol 4: Preparing DNA from Buccal Cells, 116
I.B. King
 - Protocol 5: Preparing Genomic DNA from Whole Blood: Small- and Mid-Scale Extraction, 120
A. Sahota, A.I. Brooks, J.A. Tischfield, and I.B. King
 - Protocol 6: Preparing DNA from Blood: Large-Scale Extraction, 124
A. Sahota, A.I. Brooks, and J.A. Tischfield
 - Protocol 7: Preparing DNA from Saliva, 129
A. Sahota, A.I. Brooks, and J.A. Tischfield
 - Protocol 8: Whole-Genome Amplification of Genomic DNA Using PCR, 131
C. Brueck, C. Brown, S. Michalik, D. Vassar-Nieto, E. Mueller, and G. Davis
 - Protocol 9: Single-Cell Whole-Genome Amplification, 136
C. Brueck, C. Brown, S. Michalik, D. Vassar-Nieto, E. Mueller, and G. Davis

Protocol 10: Whole-Genome Amplification using Φ 29 DNA Polymerase, 140

N.P. Burt

Protocol 11: DNA Extraction from Forensic Samples Using Chelex, 146

J.M. Butler

Protocol 12: Estimation of DNA Concentration in Forensic Samples Prior to Multiplex PCR Amplification, 148

J.M. Butler

PART 2: SNP VARIATION ANALYSIS

10 Intermediate-Throughput Laboratory-Scale Genotyping Solutions, 155

S.J. Macdonald

11 Intermediate-Throughput Laboratory-Scale Genotyping Protocols, 171

E. Cuppen, S.J. Macdonald, C. Ha, P.-Y. Kwok, W.B. Barbazuk, A.-P. Hsia, H.D. Chen, Y. Fu, K. Ohtsu, and P.S. Schnable

Protocol 1: Genotyping by Allele-specific Amplification, 172

E. Cuppen

Protocol 2: Genotyping by Dideoxy Resequencing, 174

E. Cuppen

Protocol 3: Oligonucleotide Ligation Assay, 176

S.J. Macdonald

Protocol 4: Template-directed Dye-Incorporation Assay with Fluorescence Polarization Detection, 182

C. Ha and P.-Y. Kwok

Protocol 5: Temperature Gradient Capillary Electrophoresis Assay, 192

A.-P. Hsia, H.D. Chen, and P.S. Schnable

Protocol 6: SNP Mining from Maize 454 EST Sequences, 195

W.B. Barbazuk, S. Emrich, and P.S. Schnable

12 Molecular Inversion Probes and Universal Tag Arrays: Application to Highplex Targeted SNP Genotyping, 199

G. Karlin-Neumann, M. Sedova, R. Sapolsky, J. Forman, Y. Wang, M. Moorhead, and M. Faham

Protocol: Targeted SNP Genotyping with MIP and Universal Tag Arrays, 203

13 Whole-Genome Genotyping, 213

S.B. Gabriel and M.P. Weiner

PART 3: COPY NUMBER AND COMPLEX VARIATION ANALYSIS

14 Comparative Genomic Hybridization to Detect Variation in the Copy Number of Large DNA Segments, 221

I.N. Holcomb and B.J. Trask

15 Representational Oligonucleotide Microarray Analysis Detection of Genetic Variation, 233

R. Lucito

16 Whole-Genome Sampling Analysis to Detect Copy Number Changes in FFPE Samples, 241

S. Jacobs

17 Molecular Inversion Probe Targeted Genotyping: Application to Copy Number Determination, 249

G. Karlin-Neumann, M. Sedova, R. Sapolsky, S. Lin, Y. Wang, M. Moorhead, and M. Faham

18 Microsatellite Markers for Linkage and Association Studies, 253

J. Gulcher

SECTION 3: DATA ANALYSIS

Following on the generation of multiple types of genetic data, Section 3 covers multiple approaches to data analysis. With rich reference data sets available, such as the Human HapMap, there are continually evolving approaches for data analysis to extract maximal information from genome variation. These include strategies for SNP selection for maximizing the “return on investment” of SNP genotyping by pinpointing the most informative SNPs to study; approaches to data visualization to simplify complex genotype data sets; methods of analysis to study association of genotype with phenotype; the use of genetic data sets to study evolutionary phenomena such as selection. While the density and breadth of genetic data that are routinely at hand for the investigator continue to expand, the specific approaches described here represent a set of issues and challenges amenable to the current generation of genome-wide SNP data for use in genetic association analysis.

- 19 Considerations for SNP Selection, 263
C. Carlson
- 20 Selection and Evaluation of tag-SNPs Using Tagger and HapMap, 283
P.I.W. de Bakker
- 21 Haploview: Visualization and Analysis of SNP Genotype Data, 289
J.C. Barrett
- 22 Considerations for Copy Number Analysis of FFPE Samples, 295
S. Jacobs
- 23 Assessing Significance in Genetic Association Studies, 303
M.J. Daly
- 24 Assessing Human Variation Data for Signatures of Natural Selection, 309
M. Bamshad and J.C. Stephens

SECTION 4: VARIATION STUDIES IN MODEL ORGANISMS

Section 4 includes an overview of eight different model species: three plant and five animal. The plant chapters include *Arabidopsis*, a traditional model for plant geneticists and biochemists, and two important crop plants (corn and rice). The animal chapters include rodent (mouse and rat), canine (domestic dog), feline (domestic cat), and primate (chimpanzee) species. For each species, the authors have included a description of the genome sequence as it currently exists (at the time of this writing), historical origin of the species, phenotypic variations, association studies, and Internet resources. Several of the chapters include information about the phylogeny of related strains and related species.

- 25 *Arabidopsis*, 317
Y. Li and J.O. Borevitz
- 26 Maize, 325
W.B. Barbazuk, A.-P. Hsia, H.D. Chen, Y. Fu, K. Ohtsu, and P.S. Schnable
- 27 Rice, 335
H. Leung, K.L. McNally, and D. Mackill
- 28 The Mouse, 353
C.M. Wade and M.J. Daly
- 29 The Rat, 361
E. Cuppen, N. Hübner, H.J. Jacob, and A.E. Kwitek
- 30 The Cat, 381
M.J. Lipinski, N. Billings, and L.A. Lyons

- 31 The Dog, 395
K. Lindblad-Toh and E.A. Ostrander
- 32 The Chimpanzee, 409
T.S. Mikkelsen, M.C. Zody, and K. Lindblad-Toh

SECTION 5: INSIGHTS INTO HUMAN VARIATION

Section 5 contains two diverse applications of genomic research to human beings. Chapter 33 illustrates the utility of sex-specific markers, those from mtDNA and the Y chromosome, in the elucidation of human migrations through Polynesia. Chapter 34 gives a thorough overview of the requirements and application of DNA testing to forensics. The section concludes with a third chapter that looks ahead to the future of genome research in humans and other species and to possible societal implications.

- 33 Genealogical Markers: mtDNA and the Y Chromosome, 421
M. Stoneking and M. Kayser
- 34 Forensic DNA Testing, 437
J.M. Butler
- 35 The Human Genome: What Lies Ahead, 453
M.P. Weiner and J.C. Stephens

APPENDIX: Cautions, 457

Index, 461