METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Population Epigenetics

Methods and Protocols

Edited by

Paul Haggarty

Rowett Institute of Nutrition and Health
University of Aberdeen
Aberdeen, Scotland, UK

Kristina Harrison

Rowett Institute of Nutrition and Health
University of Aberdeen
Aberdeen, Scotland, UK

Humana Press

Editors
Paul Haggarty
Rowett Institute of Nutrition and Health
University of Aberdeen
Aberdeen, Scotland, UK

Kristina Harrison
Rowett Institute of Nutrition and Health
University of Aberdeen
Aberdeen, Scotland, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-6901-2 ISBN 978-1-4939-6903-6 (eBook)
DOI 10.1007/978-1-4939-6903-6

Library of Congress Control Number: 2017933297

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Population epigenetics is an emerging field that seeks to exploit the latest insights in epigenetics to improve our understanding of the factors that influence health and longevity. Epigenetics is at the heart of a series of feedback loops that allow crosstalk between the genome and its environment. Epigenetic status is influenced by a range of environmental exposures including diet and nutrition, lifestyle, social status, infertility and its treatment, and even the emotional environment. Early life has been highlighted as a period of heightened sensitivity when the environment can have long-lasting epigenetic effects. Epigenetic status is also influenced by genotype at the level of both the local DNA sequence being epigenetically marked and the genes coding for the factors controlling epigenetic processes.

The promise of epigenetics is that, unlike the genetic determinants of health, it is modifiable and potentially reversible. The field of population epigenetics is of increasing interest to policy makers searching for explanations for complex epidemiological observations and conceptual models on which to base interventions. In order to fully exploit the potential of this exciting new field, we need to better understand the environmental and genetic programming of epigenetic states, the persistence of these marks in time, and their effect on biological function and health in current and future generations. This volume describes laboratory methodologies that can help researchers achieve these goals.

The most commonly studied epigenetic phenomenon in the field of population epigenetics is DNA methylation. Because of this, and the ready availability of methods to measure it, DNA methylation is probably the mechanism most amenable to study in population epigenetics in the near future. DNA methylation can be investigated at the level of individual methylation sites, specific genes, regions of the genome, or functional groups (e.g., promoters). An increasing number of human studies use array-based technologies to measure a great many methylation sites in a single sample. The trend is toward larger arrays measuring more and more methylation sites, but these tend to focus on the coding regions of the human genome. A significant component of the global methylation signature (average level of methylation across the entire genome) is accounted for by repeat elements. There are a number of classes of transposons and these include the long interspersed nuclear elements (LINE1), short interspersed transposable nuclear elements (SINE), and the *Alu* family of *SINE* elements. Approximately 45% of the human genome is made up of repeat elements, some of which are able to move around the genome and have the potential to cause abnormal function and disease if inserted into areas of the genome where the sequence is important for function. These are often heavily methylated, and this has the effect of repressing transposition and protecting the early embryo in particular from potentially damaging genome rearrangement during critical periods of development. Transposable elements are frequently found in or near genes, and the chromatin conformation at retrotransposons may spread and influence the transcription of nearby genes. There are particular problems in measuring this class of epigenetic regulators, and *Ha et al.* present a targeted high-throughput sequencing protocol for determination of the location of mobile elements within the genome. *Hoad and Harrison* consider the design and optimization of DNA methylation pyrosequencing assays targeting region-specific repeat elements. *Hay et al.* also focus on the noncoding genome where they describe online data mining of existing

databases to identify functional regions of the genome affected by epigenetic modification and how these modifications might interact with polymorphic variation.

Chromatin is organized into accessible regions of euchromatin and poorly accessible regions of heterochromatin, and epigenetic control is fundamental to the transition between these states. Initiatives such as the ENCODE project have highlighted the importance of long-range epigenetic interactions to the function and regulation of the genome, and there is increasing interest in studying the large-scale epigenetic regulation of the genome in population studies. The chromosome conformation capture technique provides a way of assessing chromatin states in population studies. *Rudan and colleagues* describe the use of Hi-C while *Ea et al.* set out a quantitative 3C (3C-qPCR) protocol for improved quantitative analyses of intrachromosomal contacts. These authors also describe an algorithm for data normalization which allows more accurate comparisons between contact profiles.

The methylation state of the genome is a function of DNA methylation and demethylation, and much more is known about the former than the latter but that is beginning to change with our emerging understanding of the role of the 10–11 translocation (TET) proteins. *Thomson et al.* consider the potential functional role of 5-hydroxymethylcytosine (5hmC) and describe approaches to map this important modification.

One of the most important practical problems in population epigenetics results from tissue differences in epigenetic states. In many human cohort studies typically only peripheral blood or buccal cell DNA may be available but it cannot be assumed that epigenetic status in DNA from these sources reflects that in other tissues. The rationale for blood and buccal cell sampling is that epigenetic status within these cells is either indicative of key epigenetic events in the tissues and organs of interest or that it is simply a useful biomarker. However, this may not always be valid and heterogeneity of cell types, even within a blood sample, has the potential to confound research findings in population epigenetic studies. *Jones et al.* describe the use of a regression method to adjust for cell-type composition in DNA methylation data generated by methylation arrays, pyrosequencing or genome-wide bisulfite sequencing data. *Zou* describes a computational method (FaST-LMM-EWASher) which automatically corrects for cell-type composition without needing explicit prior knowledge of this.

In population studies there may be a limitation on the type and amount of material available for epigenetic analysis. *Butcher and Beck* describe nano-MeDIP-seq, a technique which allows methylome analysis using nanogram quantities of starting material. Most epigenetic studies are carried out in DNA derived from cells, but there is increasing interest in the potential for measurement of cell-free DNA in blood and other body fluids. *Jung et al.* describe methods for DNA methylation analysis of cell-free circulating DNA. Formalin-fixed, paraffin-embedded (FFPE) tissue is often studied in clinical research, but such samples are increasingly used in epidemiological study designs. *Jung et al.* also describe methods for epigenetic analysis of FFPE tissues and protocols for the preparation, bisulfite conversion, and DNA clean-up, for a wide range of tissue types.

The process of imprinting is particularly relevant to life course studies and the long-term effects on health of early environmental exposures. Imprinted genes are epigenetically regulated by methylation according to parental origin. The imprints are established early in development and, once set, the imprint persists in multiple tissue types over decades. There is evidence that some imprinting methylation in humans may be influenced by the early life environment. The characteristics of the imprinted genes—sensitivity to early life environment, stability in multiple tissues once set—make them particularly relevant to the study of early epigenetic programming of later health. *Skaar and Jirtle* describe methods for

examining epigenetic regulation within regulatory DNA sequences with allele-specific methylation and monoallelic expression of opposite alleles in a parent-of-origin-specific manner.

Population epigenetics produces particular bioinformatic and statistical challenges when carrying out analysis of epigenetic data. *Horgan and Chua* describe methods for checking and cleaning data, the importance of batch effects, correction for multiple comparisons and false discovery rates, and the use of multivariate methods such as principal component analysis. In population epigenetics a further challenge lies in relating epigenetic data to phenotypic and exposure data in individuals and groups. Depending on the study design, epigenetic states can be considered as either an outcome or an explanatory variable and these authors describe how to match the statistical modeling approaches to the experimental question.

Our hope is that the methods presented in this volume will allow population researchers to exploit the latest insights into epigenetics to improve our understanding of the factors that influence human health and longevity.

Aberdeen, Scotland, UK

Paul Haggarty
Kristina Harrison

Contents

Contributors

Stephan Beck • *UCL Cancer Institute, University College London, London, UK*

Lee M. Butcher • *UCL Cancer Institute, University College London, London, UK*

Sok-Peng Chua • *Biomathematics and Statistics, University of Aberdeen, Aberdeen, UK*

Franck Court • *Institut de Génétique Moléculaire de Montpellier, UMR5535, CNRS, Université de Montpellier, Montpellier, Cedex 5, France; Inserm UMR1103, CNRS UMR6293, F-63001 Clermont-Ferrand, France and Clermont Universite, Université d'Auvergne, Laboratoire GReD, Clermont-Ferrand, France*

Philip Cowie • *Institute of Medical Sciences, School of Medical Sciences, University of Aberdeen, Aberdeen, UK*

Dimo Dietrich • *Institute of Pathology, University Hospital Bonn (UKB), Bonn, Germany*

Vuthy Ea • *Institut de Génétique Moléculaire de Montpellier, UMR5535, CNRS, Université de Montpellier, Montpellier, Cedex 5, France*

Rachel D. Edgar • *Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada*

Thierry Forne • *Institut de Génétique Moléculaire de Montpellier, UMR5535, CNRS, Université de Montpellier, Montpellier, Cedex 5, France*

Hongseok Ha • *Department of Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ, USA; Human Genetic Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*

Suzana Hadjur • *Research Department of Cancer Biology, Cancer Institute, University College London, London, UK*

Kristina Harrison • *Natural Products Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, Scotland, UK*

Elizabeth A. Hay • *Institute of Medical Sciences, School of Medical Sciences, University of Aberdeen, Aberdeen, UK*

Gwen Hoad • *Lifelong Health Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, Scotland, UK*

Graham W. Horgan • *Biomathematics and Statistics, University of Aberdeen, Aberdeen, UK*

Sumaiya A. Islam • *Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada*

Randy L. Jirtle • *Department of Oncology, McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, WI, USA; Department of Sport and Exercise Sciences, Institute of Sport and Physical Activity Research (ISPAR), University of Bedfordshire, Bedford, Bedfordshire, UK*

Meaghan J. Jones • *Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada*

Maria Jung • *Institute of Pathology, University Hospital Bonn (UKB), Bonn, Germany*

Michael S. Kobor • *Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada*
Glen Kristiansen • *Institute of Pathology, University Hospital Bonn (UKB), Bonn, Germany*
Alasdair MacKenzie • *Institute of Medical Sciences, School of Medical Sciences, University of Aberdeen, Aberdeen, UK*
Richard R. Meehan • *MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, UK*
Colm E. Nestor • *The Centre for Individualized Medication, Linköping University Hospital, Linköping University, Linköping, Sweden*
Matteo Vietri Rudan • *Research Department of Cancer Biology, Cancer Institute, University College London, London, UK*
Tom Sexton • *Institute of Genetics and Molecular and Cellular Biology, CNRS UMR7104/ INSERM U964, Illkirch, France; University of Strasbourg, Illkirch, France*
David A. Skaar • *Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA*
John P. Thomson • *MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Edinburgh, UK*
Barbara Uhl • *Institute of Pathology, University Hospital Bonn (UKB), Bonn, Germany*
Nan Wang • *Department of Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ, USA; Human Genetic Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*
Jinchuan Xing • *Department of Genetics, Rutgers, the State University of New Jersey, Piscataway, NJ, USA; Human Genetic Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*
James Y. Zou • *School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA*