PRIMATOLOGY MONOGRAPHS

H. Hirai · H. Imai · Y. Go *Editors*

Post-Genome Biology of Primates

Springer

Primatology Monographs

Series Editors

Tetsuro Matsuzawa
Inuyama, Japan

Juichi Yamagiwa
Kyoto, Japan

For further volumes:
http://www.springer.com/series/8796

Hirohisa Hirai · Hiroo Imai · Yasuhiro Go
Editors

Post-Genome Biology of Primates

Springer

Editors
Hirohisa Hirai
Professor
Primate Research Institute
Kyoto University
41-2 Kanrin, Inuyama
Aichi 484-8506, Japan
hhirai@pri.kyoto-u.ac.jp

Yasuhiro Go
Assistant Professor
Primate Research Institute
Kyoto University
41-2 Kanrin, Inuyama
Aichi 484-8506, Japan
yago@pri.kyoto-u.ac.jp

Hiroo Imai
Associate Professor
Primate Research Institute
Kyoto University
41-2 Kanrin, Inuyama
Aichi 484-8506, Japan
imai@pri.kyoto-u.ac.jp

ISSN 2190-5967 e-ISSN 2190-5975
ISBN 978-4-431-54010-6 e-ISBN 978-4-431-54011-3
DOI 10.1007/978-4-431-54011-3
Springer Tokyo Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011943308

Cover illustration:
Front cover: *Top*: An adult male white-handed gibbon (*Hylobates lar*) at the Nakhon Ratchasima (Khorat) Zoo, Thailand. Photo by Hirohisa Hirai. *Center left*: Hybridization of a human bacterial artificial chromosome shows split signals (*red*) on two chromosomes of the white-cheeked gibbon, revealing an evolutionary breakpoint. Photo by Roscoe Stanyon. *Center middle*: A proboscis monkey metaphase counterstained in *blue* and hybridized by human chromosome paints: 1 in *green*, 3 in *yellow*, and 19 in *red*. Photo by Roscoe Stanyon. *Center right*: Screenshot of an alignment of short nucleotide reads produced by next-generation sequencing (NGS). Photo by Yasuhiro Go.
Back cover: Grooming by Japanese macaques (*Macaca fuscata fuscata*) at the Primate Research Institute, Kyoto University, Japan. Photo by Hirohisa Hirai.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

It is a great pleasure and honor to be asked to write a foreword to this volume, which addresses the *Post-Genome Biology of Primates*. It is hard to believe that just over 10 years ago there was a raging debate as to which primate genome should be selected for sequencing next, after the human, mouse, and rat genomes had been completed. The chimpanzee eventually won out, based on feedback from the academic community, but a strong minority believed that a better studied and more experimentally tractable animal such as the rhesus monkey or baboon should have had priority. In hindsight, all these arguments turned out to be meaningless, as the pace of genome sequencing increased so rapidly and the costs fell so dramatically that many primate genomes have been partially or completely sequenced within the past decade.

There is still a long way to go before one can say that we have covered all genomes that would be worth sequencing (one even could argue that all of them are), and population-level genomic information is still very limited for most primate genomes. But I think it is safe to say we are now indeed in an era where the genomic sequences that are already available can be used to explicate the genetic and genomic contributions toward primate evolution and phenotype. Indeed, we are now in a situation in which it is the phenotypic information has become rate limiting. In this volume, the editors have brought together an excellent collection of papers covering a wide variety of topics relevant to primate genomes, including evolution, genome structure, chromosome genomics, bioinformatics, and functions. Although it is impossible to do justice to all possible topics in this huge area of research, this book covers many that should be of interest, not only to those who study primate and primate genomes, but also for those wishing to understand human origins (“anthropogeny”) and the remarkable phenotypic diversity of primates. Also included are somewhat more theoretical papers about issues of interest to other readers.

This valuable resource will undoubtedly catalyze further sequencing of primate genomes as well as studies of primate phenotypes. Thus, although we are in a "Post-Genome Era," we will also continue to be in the "Genome Era" for some time yet. Meanwhile, please enjoy reading this timely and informative volume.

Ajit Varki
Distinguished Professor of Medicine and Cellular & Molecular Medicine
Co-Director, Center for Academic Research and Training in Anthropogeny (CARTA)
Co-Director, Glycobiology Research and Training Center
University of California, San Diego, La Jolla, CA, USA

Contents

Contributors

Kaylin A. Adipietro (*Chapter 5*)
Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, USA

Nicoletta Archidiacono (*Chapter 13*)
Department of Genetics and Microbiology, University of Bari, Bari, Italy

Filippo Aureli (*Chapter 7*)
Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool, UK
and
Instituto de Neuroetologia, Universidad Veracruzana, Xalapa, Mexico

Linda M. Fedigan (*Chapter 7*)
Department of Anthropology, University of Calgary, Calgary, AB, Canada

Yasuhiro Go (*Chapter 1*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Masami Hasegawa (*Chapter 16*)
School of Life Sciences, Fudan University, Shanghai, China
and
The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

Toshiyuki Hayakawa (*Chapter 8*)
Center for Human Evolution Modeling Research, Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Hirohisa Hirai (*Chapter 1, 15*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Chihiro Hiramatsu (*Chapter 7*)
Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan
and
Department of Psychology, Graduate School of Letters, Kyoto University, Kyoto, Japan

Hiroo Imai (*Chapter 1, 6*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Shoji Kawamura (*Chapter 7*)
Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

Philipp Khaitovich (*Chapter 3*)
Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
and
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Kenneth K. Kidd (*Chapter 9*)
Department of Genetics, Yale University School of Medicine, New Haven, CT, USA

Takashi Kitano (*Chapter 12*)
Department of Biomolecular Functional Engineering, College of Engineering, Ibaraki University, Hitachi, Japan

Akihiko Koga (*Chapter 11*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Stefan Müller (*Chapter 14*)
Institut für Humangenetik, Klinikum der Ludwig-Maximilians-Universität, Munich, Germany

Atsushi Matsui (*Chapter 16*)
Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Hiroaki Matsunami (*Chapter 5*)
Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, USA
and
Department of Neurobiology, Duke University Medical Center, Durham, NC, USA

Amanda D. Melin (*Chapter 7*)
Department of Anthropology, University of Calgary, Calgary, AB, Canada
and
Department of Anthropology, Dartmouth College, Hanover, NH, USA

Yoshihito Niimura (*Chapter 4*)
Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Hiroki Oota (*Chapter 9*)
Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan
and
Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan

Naoki Osada (*Chapter 2*)
Department of Population Genetics, National Institute of Genetics, Shizuoka, Japan

Mariano Rocchi (*Chapter 13*)
Department of Genetics and Microbiology, University of Bari, Bari, Italy

Yoko Satta (*Chapter 10*)
Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies, Sokendai Hayama, Hayama, Kanagawa, Japan

Colleen M. Schaffner (*Chapter 7*)
Psychology Department, University of Chester, Chester, UK
and
Instituto de Neuroetologia, Universidad Veracruzana, Xalapa, Mexico

Mehmet Somel (*Chapter 3*)
Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
and
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Roscoe Stanyon (*Chapter 13*)
Laboratory of Anthropology, Department of Evolutionary Biology, University of Florence, Florence, Italy

Tohru Sugawara (*Chapter 6*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan
and
Department of Reproductive Biology, National Research Institute for Child Health and Development, Tokyo, Japan

Masanaru Takai (*Chapter 17*)
Primate Research Institute, Kyoto University, Inuyama, Aichi, Japan

Lin Tang (*Chapter 3*)
Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Ajit Varki (*Chapter 8*)
Center for Academic Research and Training in Anthropogeny, Glycobiology Research and Training Center, Departments of Medicine and Cellular & Molecular Medicine, University of California, San Diego, La Jolla, CA, USA

Johannes Wienberg (*Chapter 14*)
Anthropology and Human Genetics, Department of Biology II, Ludwig-Maximilians-Universität, Munich, Germany
and
Chrombios GmbH, Raubling, Germany

Hanyi Zhuang (*Chapter 5*)
Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of National Ministry of Education, Shanghai Jiaotong University School of Medicine, Shanghai, China
and
Institute of Health Sciences, Shanghai Institutes for Biological Sciences of Chinese Academy of Sciences and Shanghai Jiaotong University School of Medicine, Shanghai, China

Chapter 1
Introduction

Yasuhiro Go, Hiroo Imai, and Hirohisa Hirai

1.1 Introduction

A decade ago, the first reports of the human draft genome were simultaneously published in *Nature* from the international Human Genome Project (International Human Genome Sequencing Consortium 2001) and in *Science* from the company Celera Genomics (Venter et al. 2001). Since the milestone of the human genome, genome projects of many organisms have been proposed, undertaken, and achieved in the past decade. These organisms include the mouse (Mouse Genome Sequencing Consortium 2002), rat (Rat Genome Sequencing Project Consortium 2004), dog (Lindblad-Toh et al. 2005), chimpanzee (Chimpanzee Sequencing and Analysis Consortium 2005), rhesus macaque (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), marsupial (Mikkelsen et al. 2007), and, more recently, the Neanderthal (Green et al. 2010). As for primates, besides the chimpanzee and rhesus macaque, many other primate genomes have been sequenced, such as gorilla, orangutan, gibbon, baboon, marmoset, tarsier, galago, and lemur. New insights are thus required to think about how we should use the vast information of genome sequences for post-genome investigations. Now is the best time to establish standpoints for genomic primatology in these early days in several areas of genomic research. Here we introduce the angles from which we investigate primates with the aim of understanding what makes us human.

This book consists of four sections: each has two to six chapters, as listed in the Table of Contents.

The first section is "Post-Genomic Approaches Toward Phenotype," in which we introduce approaches to uncover phenotypic changes, including physiological and behavioral changes, based on the genomic and transcriptome level, especially focusing

Y. Go • H. Imai • H. Hirai (✉)
Primate Research Institute, Kyoto University,
41-2 Kanrin, Inuyama, Aichi 484-8506, Japan
e-mail: yago@pri.kyoto-u.ac.jp; imai@pri.kyoto-u.ac.jp; hhirai@pri.kyoto-u.ac.jp

H. Hirai et al. (eds.), *Post-Genome Biology of Primates*, Primatology Monographs,
DOI 10.1007/978-4-431-54011-3_1,

on the sensory functions. The sensory functions that perceive various external signals, such as light, smell, and taste, play major roles in sensing physical or chemical environmental changes and in taking such information inside the organisms. The necessity for such sensory functions for each organism could be variable and heavily dependent on the environment to which each organism has adapted. This feature thus results in producing functional diversity of the sensory functions from organism to organism, and this diversity has conferred various species-specific phenotypic characters. The first chapter of this section, Chap. 2, is written by Naoki Osada from National Institute of Genetics. He highlights the importance of the transcriptome, a first outcome of the genome and a key component linking the genotype and phenotype of an organism, and introduces the recent advance of transcriptome studies in nonhuman primates and the quantification methodology for the transcriptome. The next chapter, Chap. 3, is a review by Mehmet Somel et al. from the Chinese Academy of Sciences and German Max Planck Society. They draw attention to one of the most distinguishing features that characterizes humans as distinct from the other primates, so-called neoteny. Neoteny is a form of heterochrony that is defined as a developmental change in the timing of events, leading to changes in size and shape. The authors examine whether human-specific changes can be seen and what kind of genes are involved in the molecular basis of neoteny using brain transcriptome results of humans, chimpanzees, and rhesus macaques covering almost all ages. In the third chapter, Chap. 4, Yoshihito Niimura from Tokyo Medical and Dental University reviews chemosensory receptor gene evolution in primates and mammals. Among chemosensory receptor genes, olfactory receptor genes are the largest multigene family in the mammalian genome, and the number of genes differs greatly among species (~1,000 genes in rodents, but fewer than 400 in primates). Niimura argues for a dynamic change of the repertoire of the olfactory receptor genes in the context of trade-off between vision and the olfaction system. The fourth chapter of this section, Chap. 5, is a review by Kaylin Adipietro et al. from Duke University. Although they also examine the evolution of odorant (olfactory) receptors in primates, as in the previous chapter, their studies are based on a more functional or system point of view. Using in vitro functional assays of the ligand sensitivity of odorant receptors and behavioral evaluation of responses to a set of smells, they found different responses to a set of ligands or smells among very closely related species and even between sexes in humans. This discovery in general implies that the response to some smells might be modulated at the transcriptional, metabolic, or epigenetic level. In the fifth chapter, Chap. 6, Tohru Sugawara and Hiroo Imai from Kyoto University highlight another important chemosensory receptor involved in taste perception. Of five taste modalities (sweet, bitter, sour, salty, and umami), the sense of bitter taste is known to be highly polymorphic. For instance, although people can taste some bitter compounds, such as phenylthiocarbamide (PTC), there are people who cannot sense the same bitter compound (called "non-tasters"). The genetic basis of this polymorphism has recently been attributed to one of the bitter taste receptor genes (*T2R38*). It is also known that such behavioral polymorphisms are observed not only in humans but also in some other primates. The authors then examined the genetic basis of such polymorphisms in chimpanzees

and uncovered the evolutionary origin and significance of such "non-tasters" in humans and chimpanzees. They also give us a nice review of the function and evolution of other taste-related genes. The final chapter of this section, Chap. 7, is a review by Shoji Kawamura et al. from the University of Tokyo on the evolution of the color vision system in primates. Color vision is a crucial cue for object detection, food identification, mate choice, and predator avoidance. Kawamura et al. show an evolutionary significance of the color vision (opsins) system in primates and examine precisely to what extent three-color vision (trichromacy) has an advantage over two-color vision (dichromacy) in the environment of free-ranging living primates. As New World monkeys are unique with respect to the existence of dichromatic and trichromatic individuals in a species or even in a group, they examine the genotypes of opsin genes in free-ranging New World monkeys and record each individual's behavioral data. Connecting the genetic and behavioral data in the wild monkeys, they uncover the advantages and disadvantages of trichromacy in the environmental context. This study gives us one of the best examples of how we can incorporate multidisciplinary approaches in the post-genome era of biology.

The second section, including chapters on "Genome Structure and Its Applications," explores the impact of genomic structural changes on human evolution. The first chapter, Chap. 8, is on the evolution of sialic acids, one of the important components of sugar chains. Toshiyuki Hayakawa from Kyoto University and Ajit Varki from the University of California, San Diego, give an overview of the evolution of sialic acid biology in primates and highlight sialic acid-related human-specific changes and their possible impact on human evolution. The discovery of such human-specific genetic changes is one of the hallmarks of human uniqueness and can be a good clue for thinking about a longstanding question: What makes us human? The second chapter of this section, Chap. 9, describes the evolution of the genes involved in alcohol metabolism in primates. Hiroki Oota from Kitasato University and Kenneth Kidd from Yale University show copy number variation of the alcohol dehydrogenase (*ADH*) gene in detail and disclose the independent origin of each *ADH* gene between apes and Old World monkeys. Based on the findings, they hypothesize that frugivorous feeding behavior facilitates the maintenance of taxon-specific duplicated genes because of the necessity of digesting ethanol generated by the fermentation of fruit sugar. In the third chapter, Chap. 10, Yoko Satta from The Graduate University for Advanced Studies (Sokendai) gives us a review of genome structure evolution especially focused on sex chromosomes. Because the genome sequencing projects covered a wide range of organisms, many types of structural changes, such as segmental duplications, copy number variations, and insertions and deletions, have been discerned and quantified. Among the genome (chromosomes), the Y chromosome is exceptional because it exists as a hemizygous chromosome in the genome and most of the mutations that could be deleterious are not then eliminated as a result of the arrest of recombination with the X chromosome. She shows the discontinuous structure of the human Y chromosome with respect to the evolutionary relationships of gametologous (homologous relationships between sex chromosomes) genes on the X chromosome and discusses the evolutionary origin and biological significance of sex chromosomes in the light of

human evolution. Akihiko Koga from Kyoto University reviews the impact of DNA-based transposed elements (DTEs) on the genome and their evolution in the fourth chapter, Chap. 11. Although most DTEs are thought to be dead in mammals, DTEs can trigger chromosomal rearrangements such as inversions, deletions, duplications, and translocations because of their repetitive nature. Koga discusses the potential contribution of DTEs to mammalian genome evolution. The last chapter of this section, Chap. 12, is a review by Takashi Kitano from Ibaraki University. He argues the possibility and extensibility of the phylogenetic network, an extended framework of the phylogenetic tree. Phylogenetic network methods have advantages of describing the genes with complex evolutionary genealogies resulting from processes such as recombination, hybridization, and gene conversion. He also introduces some practical applications using the phylogenetic network method.

The third section, "Chromosome Genomics," concerns molecular cytogenetics and chromosome evolution in primates. Classical comparative cytogenetics has a long history dating back to the 1950s, and since then it has used the information of the number of chromosomes (karyotypes), and chromosome banding such as Q-banding and G-banding. During the past 20 years, the introduction of molecular methods has made it possible to examine precise chromosome rearrangements among species, as revealed by the fluorescent in situ hybridization (FISH) method. Moreover, cytogenetic studies have also uncovered the mechanisms and biological meaning of the essential components of chromosomes, such as the centromere and telomere. Even now, these components are difficult to sequence by the ordinary genomic approaches as a consequence of the highly repetitive nature of their sequences. The first chapter of this section, Chap. 13, is a review by Roscoe Stanyon et al. from University of Florence. Focusing on the cytogenetic level of primate genome organization as shown by the chromosome painting method, they reveal the complex chromosome rearrangements that occurred during primate evolution and reconstruct the ancestral genome organizations. Moreover, they show intriguing phenomena of neocentromeres, which are newly formed in ectopic chromosomal regions and are even heritable in some cases, and the meaning of such neocentromeres from an evolutionary point of view. They also give a perspective of the future and possibilities of cytogenetics in the high-throughput genomic era. In the second chapter, Chap. 14, Stefan Müller and Johannes Wienberg from The Ludwig Maximilian University of Munich highlight chromosomal evolution in gibbons, one of the organisms with the highest rate of chromosome rearrangements in mammals. Although gibbons are classified as lesser or smaller apes and are phylogenetically close relatives, they diverged from humans and great apes 15–20 million years ago, and the rates of rearrangements in gibbons are 10–20 times higher than the mammalian default rate. They examine the evolutionary relationship of the highly differentiated chromosomal organizations in the four genera of gibbons (chromosome numbers from $2n=38$ to $2n=52$) based on chromosome painting methods such as those covered in the previous chapter (Chap. 13). In addition, they summarize the recent progress of elucidating the cause of the higher rate of rearrangements

with respect to the epigenomic changes occurring in this group. The final chapter, Chap. 15, is by Hirohisa Hirai from Kyoto University, who discusses the evolution and biological meaning of so-called genomic wastelands, mainly constructed from repetitive sequences in heterochromatic regions. Although humans and chimpanzees are reported to share approximately 99% of their genome sequences, this was calculated from alignable sequences of the genome. When one considers unalignable regions such as insertions and deletions, the difference of the genome between humans and chimpanzees is estimated to be about 3%. One such genomic component is heterochromatic regions that are usually enriched in centromere, telomere, and subtelomeric regions. He then argues how such heterochromatic regions (genomic wastelands) contribute to make us humans.

The fourth and last part of this book, "Evolution of Humans and Non-Human Primates," addresses the topic of primate evolution from the molecular and fossil points of view. In the first chapter, Chap. 16, Atsushi Matsui from Kyoto University and Masami Hasegawa from Fudan University summarize the recent advances of molecular primate phylogeny and point out remaining unsolved problems of molecular phylogenetic studies. Although the phylogenetic relationships of living primates are relatively well established, the divergence times among them are still controversial. They comprehensively examine the divergence time in each taxon and discuss the evolutionary scenario of primate evolution with reference to the geographic and fossil records. The last chapter, Chap. 17, is a review of primate fossil studies contributed by Masanaru Takai from Kyoto University. He especially focuses on the early time of primate evolution and draws conclusions about the place of primate origin. The North America origin hypothesis has long been accepted and is widespread among primatologists and paleontologists because of the rich fossil records of early primates. However, the author advocates that this view should be reconsidered and can be replaced by the southern continent origin hypothesis involving the Indian Continent or East Asia, based on incorporating the results from geographic evidence and recent molecular phylogenetic studies.

It is clear that this book does not completely cover the comprehensive fields of genome and post-genome biology in primates. Instead, we intended to organize the contents of the book to show front-line research for broadening one's insights and extending one's research interests incorporating various methods, technologies, and knowledge, as shown in this volume.

We are truly grateful to all the authors of this book for devoting their time to write the chapters and contributing to several refinements of the book. Thanks to all the effort, we are proud of publishing this book. We would also like to thank the series editors, Tetsuro Matsuzawa and Juichi Yamagiwa, for their special leadership and continuous support, and we thank Aiko Hiraguchi and Kaoru Hashimoro of Springer Japan for their dedicated assistance with the editing of this book. Finally, we give special thanks to numerous colleagues, postdoctoral researchers, and students. Without them, this book could never have been accomplished.

References

Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature (Lond) 437:69–87

Green RE, Krause J, Briggs AW et al (2010) A draft sequence of the Neandertal genome. Science 328:710–722

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature (Lond) 409:860–921

Lindblad-Toh K, Wade CM, Mikkelsen TS et al (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature (Lond) 438:803–819

Mikkelsen TS, Mikkelsen TS, Wakefield MJ et al (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature (Lond) 447:167–177

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature (Lond) 420:520–562

Rat Genome Sequencing Project Consortium (2004) Genome sequence of the brown Norway rat yields insights into mammalian evolution. Nature (Lond) 428:493–521

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234

Venter JC, Venter JC, Adams MD et al (2001) The sequence of the human genome. Science 291:1304–1351

Part I
Post-Genomic Approaches Toward Phenotype

Chapter 2
An Overview of Transcriptome Studies in Non-Human Primates

Naoki Osada

Abbreviations

cDNA	Complementary DNA
EST	Expressed sequence tag
SAGE	Serial analysis of gene expression

2.1 What Is a Transcriptome?

The word *transcriptome* is a combination of *transcript* and *genome*, which refers to the whole set of transcripts expressed in a cell or tissue. The word *genome* itself is a blend of *gene* and *chromosome*, which refers to the whole set of genes in an organism. Now the ending *-ome* has been applied somewhat excessively to represent any kind of massive biological dataset, for example, proteome, metabolome, phenome, interactome, and phylome. In a classical view, transcripts are equivalent to messenger RNAs (mRNAs) that encode functional proteins. However, recent progress in transcriptome analysis has demonstrated that a large number of noncoding sequences are transcribed to RNA, more than previously estimated. For example, a very deep sequencing of RNA expressed in mice revealed that more than 70% of the genome is actually transcribed to RNA, if introns are included (Carninci et al. 2005). The functions of some transcribed RNAs, such as micro-RNA and small interfering (si)RNA, have been extensively studied, whereas those of most of the noncoding RNAs remain unclear.

N. Osada (✉)
Department of Population Genetics, National Institute of Genetics,
1111 Yata, Mishima, Shizuoka 411-8540, Japan
e-mail: nosada@lab.nig.ac.jp

H. Hirai et al. (eds.), *Post-Genome Biology of Primates*, Primatology Monographs,
DOI 10.1007/978-4-431-54011-3_2, © Springer 2012

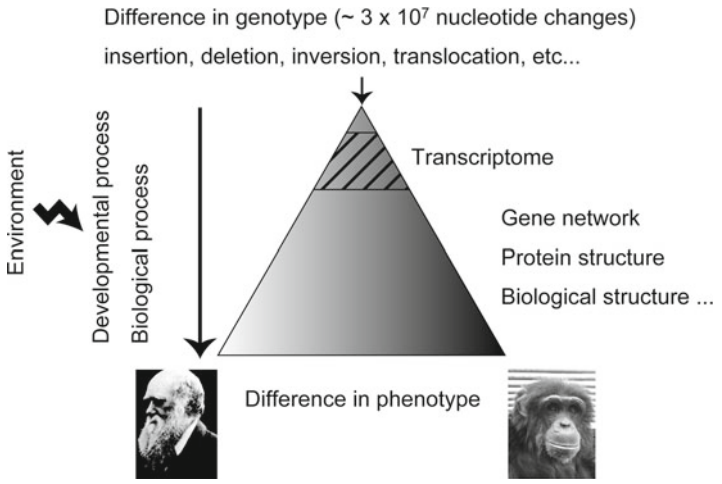

Fig. 2.1 A conceptual picture of the genotypic and phenotypic differences between humans and chimpanzees

Transcriptome data enable us to link the genotype to the phenotype of an organism. Figure 2.1 represents a conceptual picture of the genotypic and phenotypic differences between humans and chimpanzees. The human and chimpanzee genomes differ by about 1% at the nucleotide substitution level, and more at other levels including insertion, deletion, inversion, and translocation. Although not all these changes have been thoroughly cataloged, they are finite and countable features. These small genotypic differences increase according to their developmental and biological processes and by interactions with the environment of the organism, which leads to large phenotypic differences such as morphology and behavior that can be easily recognized. Compared to genotypic differences, phenotypic differences are far more complex. One can measure and compare some phenotypes, but the number of measurable phenotypes is very large (probably an infinite number). In addition, there are many cryptic phenotypes in organisms, which make the comparison of whole sets of phenotypes (phenomes) almost impossible. Transcriptome studies can be more complex than genomic studies, because a transcriptome may be a quantitative measurement and may vary in space and time (see also Chap. 3). However, compared to a phenotypic measurement, a transcriptome measurement could be a more neutral measurement of the feature of organisms. The transcriptome is sometimes referred to as an endophenotype. In other words, the transcriptome represents the very beginning of genotypic development into a complex of phenotypes.

For a long time, transcriptome studies were restricted to well-known model organisms. However, recent advances in molecular genetic techniques enable us to extend transcriptome studies to many non-model organisms, including non-human primates. Transcriptome studies in non-human primates have been strongly promoted for two major purposes: for conducting evolutionary and biomedical studies. In this review, the methodologies used in transcriptome studies and some results obtained from previous studies have been summarized.

2.2 Different Methods for Analyzing Transcriptome

2.2.1 Qualitative/Quantitative Studies

According to the type of data analyzed, transcriptome studies are classified into qualitative and quantitative studies. Qualitative studies analyze the sequence and structural differences among genes or the repertoire of expressed genes in samples, whereas quantitative studies evaluate the level of gene expression of many genes. Two major methods are widely used to analyze these different aspects of transcriptome studies. The underlying technologies, advantages, and disadvantages of these methods are briefly discussed in the following sections.

2.2.2 cDNA Sequencing

The first method is classical complementary DNA (cDNA) sequencing. mRNA in a sample is reverse transcribed to cDNA and cloned into plasmid vectors. A set of hosts (usually *Escherichia coli*) containing the vectors are called cDNA libraries. Because gene expression patterns differ by species, individual, sex, tissue, time course, and experimental treatment, many different types of cDNA libraries can be established.

In a practical transcriptome study by cDNA sequencing, hundreds to thousands of clones are randomly obtained from cDNA libraries, and their sequences are determined with a DNA sequencer to catalog many genes expressed in a target tissue. Usually, cDNA sequences at either the 5′-end or 3′-end are determined. These one-pass sequences are called expressed sequence tag (EST) sequences. EST sequences provide relatively limited information because the currently used Sanger sequencing method reads less than 1,000 bp and the average length of primate genes is greater than 2,000 bp. After EST sequencing has been performed, one may determine a full insert of clones by further sequencing with the primer-walking or shotgun methods. With standard cDNA library construction methods, obtained EST sequences are highly redundant, which signifies that many EST sequences represent the same gene. Therefore, the normalization method, which subtracts highly redundant mRNAs from a sample, is sometimes performed to construct cDNA libraries. Because most protein-coding genes in primates are highly conserved, one can investigate whether the obtained EST sequences are homologous to known human protein-coding genes by a homology search of public databases using a program such as BLAST.

There is a modified form of the conventional EST sequencing method, called serial analysis of gene expression (SAGE), in which short-sequence fragments cleaved by restriction enzymes (~15–30 bp) from the 5′- or 3′-end of cDNAs are concatenated and analyzed. The concatenated tags are sequenced with a DNA sequencer. Because a single one-pass sequence can simultaneously identify several tags, this method can determine sequence tags more efficiently than the conventional

EST sequencing method. Because the number of tags in SAGE analysis outnumbers that in ESTs, the tag count can be treated as a quantitative measure of the gene expression level. Unfortunately, this method is not as useful for studies involving non-human primate samples. Because the sequence tags of the SAGE method are much shorter than the EST sequencing, changes in even a few nucleotides may cause the misidentification of tags. The link between tags and genes is mainly based on the human genome data. Therefore, SAGE is not efficient in non-human primate studies unless there is prior information about which tag represents which gene.

Several cDNA libraries derived from non-human primates have been constructed and analyzed; for example, from chimpanzees (*Pan troglodytes*) (Hellmann et al. 2003; Sakate et al. 2003), orangutans (*Pongo pygmaeus*) (Mewes et al. 2004), rhesus macaques (*Macaca mulatta*) (Magness et al. 2005; Spindel et al. 2005), cynomolgus macaques (*Macaca fascicularis*) (Hida et al. 2000; Magness et al. 2005; Chen et al. 2006; Osada et al. 2008, 2009), pigtail macaques (*Macaca nemestrina*) (Magness et al. 2005), and common marmosets (*Callithrix jacchus*) (Datson et al. 2007). The EST sequence data of cynomolgus monkeys obtained by our research group is one of the largest non-human primate transcriptome datasets, containing 112,587 EST sequences (Osada et al. 2009). These EST sequences or cDNA clones can be used for the generation of DNA microarrays. Unfortunately, the EST data are not suitable for quantitative measurement of the gene expression level because a limited number of tags are counted. These one-pass sequences were error prone, but can be used for comparative studies of primate genes if many sequences per gene are obtained. Before the advent of non-human primate genome sequences, cDNA sequence comparison was the only method to compare many genes between primate species. Some interesting results of comparative genomics using transcriptome data are discussed later.

2.2.3 DNA Microarray

The second method uses DNA microarrays. In DNA microarray experiments, mRNAs are reverse transcribed to cDNAs, labeled, and hybridized to cDNAs or synthesized oligo-DNAs that are arrayed densely on a glass surface. The number of mRNA molecules in a sample can be measured from the relative signal intensity of each labeled probe.

cDNA microarrays are constructed by spotting cDNA clones, whose sequences are known by EST or full-length sequencing. On the other hand, oligo-DNA microarrays are constructed by spotting or synthesizing short oligonucleotides on a glass slide. Typical oligonucleotide probes for microarrays are 30–120-mers that are complementary to transcript sequences. In oligo-DNA microarrays, recent techniques can design multiple probes for a single transcript. If different probes are designed for different exons, then the microarray can detect a different expression level among alternatively spliced transcripts. For non-human primate microarrays, sequences from EST data are frequently used to design oligo-DNA microarrays.

In addition, there is a special type of DNA microarray called a tiling array. Oligo-DNA probes on tiling arrays are designed to cover a wide range of the whole genome sequence. Indeed, genome sequences of target regions must be determined to design a tiling array. Although a microarray based on transcript sequences cannot detect the expression of unknown genes, a tiling array can detect the expression of unknown transcripts in the genome, including that of noncoding RNAs and microRNAs. Studies using tiling arrays have established a very complex expression pattern in the human transcriptome (Bertone et al. 2004; Kampa et al. 2004).

Although a DNA microarray is a powerful tool to quantitatively analyze the transcriptome, there are some drawbacks of DNA microarray experiments in non-human primates. A species-specific DNA microarray should be designed in advance, because the DNA–DNA hybridization accuracy depends on sequence similarity between sample species and species used to design microarrays. This effect may cause serious problems, particularly cross-species expression comparison, because it is difficult to identify whether the observed changes in signal intensity are caused by sequence mismatches or changes in actual gene expression. If sequence divergence between species is relatively small, as between humans and chimpanzees, then its effect may be negligible. However, the effect becomes more pronounced when the target species are distantly related to microarray species. Because the effect of hybridization mismatches is much stronger in shorter probes, cDNA microarrays are supposed to be more robust to sequence mismatches than oligo-DNA microarrays (Walker et al. 2006; Jacquelin et al. 2007). Before the draft genome sequence of the rhesus macaque was determined, many studies used commercially available or custom-made human microarrays. Therefore, early studies of gene expression in non-human primates were based on human-specific microarrays (Zou et al. 2002; Marvanova et al. 2003; Sui et al. 2003; Vahey et al. 2003; Baskin et al. 2004; Rubins et al. 2004; Dillman and Phillips 2005; Ylostalo et al. 2005; Kothapalli et al. 2007; Nijland et al. 2007; Djavani et al. 2009). Although these studies obtained satisfactory results at some level, microarrays designed specifically for the particular non-human primate species would produce a much more accurate estimation of the gene expression level (Gilad et al. 2005). Up to the present, several microarrays specific to non-human primates have been developed (Osada et al. 2002; Gilad et al. 2005; Spindel et al. 2005; Datson et al. 2007; Kobasa et al. 2007; Wallace et al. 2007; Osada et al. 2008). Most of them are intended for biomedical research in non-human primates. At present, several DNA microarrays for humans and macaques are commercially available, and bioinformatics methods that mitigate the effect of sequence mismatches in oligo-DNA microarrays have been developed (Wang et al. 2004; Royce et al. 2007; Lin et al. 2009; Lu et al. 2009).

2.2.4 Next-Generation Sequencer

Recently, new DNA sequencing technologies that can identify a large number of short DNA fragments have been developed. Three different platforms are currently

available: FLX (Roche), Solexa GA (Illumina), and AB SOLiD (Life Sciences). The technologies used by these platforms differ, but each can identify millions to billions of transcript fragments in a single run. Because the number of detectable fragments is enormous, the counted fragments would correlate with the level of gene expression, as in SAGE analysis. Many studies have suggested that these new methods are capable of quantifying very low gene expression levels. For example, mRNAs that are expressed at less than one copy per cell were detectable (Hashimoto et al. 2009). Initial versions of GA and SOLiD produced only 25-bp-long sequences, and thus were not as efficient as the SAGE method for analyzing non-human primate data. The sequencing length, however, has been increased to 50–100 bp, which enables us to overcome the sequence mismatch problem. Therefore, these methods can identify a non-human primate transcriptome both qualitatively and quantitatively. If we have known genome sequences of target organisms, fragments can be easily mapped on the genome sequences; otherwise, the assembly of fragments would be more challenging. Such an assembly is designated as de novo transcriptome sequencing. Although only a limited number of non-human primate research studies have been performed using these methods, the technique has a great potential for investigating the transcriptome of non-human primates.

2.3 Subjects of Transcriptome Studies

2.3.1 Application to Biomedical Research

Many non-human primates are used as a model for humans in biomedical research. Biomedical studies include studies on infectious diseases, tissue transplantation, neurology, toxicology, and many other human diseases. In particular, pharmaceutical studies using genome-wide gene expression data are referred to as toxicogenomics and are of interest to many pharmaceutical researchers. The most popular non-human primates for biomedical research are the Old World monkeys such as macaques and baboons. Among New World monkeys, the marmoset is the most popular animal because it has a small body size and grows relatively fast. Biomedical research using invasive treatments in apes is strongly restricted because of ethical reasons. Even in other non-human primates, the investigational use of monkeys has been a debatable issue for a long time from the point of view of animal rights (Editorial 2008), but that question is not discussed in this review.

In typical biomedical studies, differences in gene expression after certain treatments are measured using DNA microarrays to identify the gene relevant to the biological response. A sampling point may be a time course for measuring temporal changes in gene expression. Because most biomedical studies try to detect differences in gene expression patterns between experimental and control samples, the sequencing mismatch problem, which was described in the previous section, is less problematic. Sequence mismatches may reduce the probe detection efficiency, but will have a lesser

effect on reproducibility in intraspecies comparison. Therefore, in many biomedical studies, non-human primate cDNA was hybridized to human-specific DNA microarrays (Zou et al. 2002; Marvanova et al. 2003; Sui et al. 2003; Vahey et al. 2003; Baskin et al. 2004; Rubins et al. 2004; Dillman and Phillips 2005; Ylostalo et al. 2005; Kothapalli et al. 2007; Nijland et al. 2007; Djavani et al. 2009).

2.3.2 Comparative Studies

Besides biomedical research, research comparing humans and non-human primates has attracted much attention of evolutionary biologists. It is not clear as to what kind of genetic components make humans phenotypically distinct from other non-human primates, especially in their high cognitive ability. Understanding humanity from a genomic perspective is a challenging but tantalizing issue in human evolutionary biology studies. In 1967, Sarich and Wilson used immunological reactions to investigate the similarity in protein structures of albumin (Sarich and Wilson 1967). They used these data to date the divergence between gorillas, chimpanzees, and humans, and concluded that their divergence was much more recent (~5 Mya) than was previously thought from morphological and fossil evidence (~15 Mya) (see also Chap. 16). At present, the very close relationship between humans and chimpanzees is supported by enormous amounts of DNA sequence data. It is now known that only approximately 1% of their genomes differ at the DNA sequence level (The Chimpanzee Sequencing and Analysis Consortium 2005).

2.3.2.1 Molecular Evolution Rate of the Primate Transcriptome

Using a comparative transcriptome analysis, which contrasts whole sets of genes among genomes, genes responsible for human-specific traits could be identified. A comparison between human and chimpanzee genome sequences revealed a difference of about 40,000 amino acids in their protein sequences (The Chimpanzee Sequencing and Analysis Consortium 2005). Many of these differences are probably neutral, that is, have no phenotypic effect, but some of them may have been affected by positive or negative selection. Here, positive and negative selection means natural selection on beneficial and deleterious mutations, respectively. It is intuitive that mutations causing beneficial phenotypic changes quickly spread within populations, whereas bad mutations are easily removed from populations. Mutations that are neither good nor bad are assumed to be selectively neutral.

To estimate the mode of protein evolution, the relative rate of protein evolution was measured by observing synonymous and nonsynonymous changes between species. Synonymous substitutions are nucleotide changes that do not affect encoded protein sequences, whereas nonsynonymous substitutions are nucleotide changes that alter encoded proteins. The rate of nucleotide substitution per site for

synonymous (K_S) and nonsynonymous (K_A) substitutions can be estimated using several statistical methods. It can be assumed that synonymous substitutions are mostly selectively neutral, although it is known that weak negative selection caused by translational efficiency may act on synonymous substitutions (Akashi 1994). K_A and K_S are also referred as d_N and d_S, respectively (see also Chap. 4).

If synonymous substitutions are selectively neutral, K_S equals the mutation rate under neutral theory of molecular evolution (Kimura 1968). Neutral theory of molecular evolution also predicts that if the changes in the protein are largely deleterious, K_A becomes much smaller than K_S. Comparison between human and chimpanzee genomes revealed that the average K_A/K_S was around 0.20–0.25, indicating that about 70–80% of amino acid changes in humans and chimpanzees are deleterious. However, the intensity of natural selection (estimated by K_A/K_S) is different for different genes. If some gene is biologically important and the negative selection intensity of the gene is very severe, the K_A/K_S value of the gene becomes extremely small. For example, humans and mice share identical amino acid sequences for one of the histone proteins, H4, which is a fundamental protein constituting chromosome structure. On the other hand, if many new mutations in the protein are beneficial to the organism, those mutations spread rapidly and become fixed in the species. Then, it is assumed that K_A exceeds K_S ($K_A/K_S>1$) for such a gene. Thus, comparison of human and chimpanzee transcriptome data at a nucleotide level may be useful to identify which genes have the greatest impact on the human–chimpanzee divergence.

Before the genomic era, when no non-human primate genome sequences had been determined, the only way to catalog a large number of genes in non-human primate genomes was by the analysis of cDNA libraries. Hellmann et al. (2003) constructed cDNA libraries derived from chimpanzee brain and testis, sequenced about 5,000 EST sequences from the libraries, and compared the sequences with those of humans. In conjunction with the human polymorphism data, they estimated the level of negative selection on genes after the human–chimpanzee divergence. Similarly, Sakate et al. (2003) constructed cDNA libraries derived from chimpanzee brain, skin, and liver, and estimated the molecular evolution rate of hundreds of genes. These two studies were the first large-scale comparisons of human–chimp genes.

Using the Old World monkeys, Osada et al. (2002) attempted to discover genes that have evolved rapidly after the human–macaque divergence, with about 10,000 EST sequences of cynomolgus macaques. They identified eight candidate genes that showed $K_A/K_S>1$. Interestingly, four of these are nuclear genes that encode mitochondrial components. By analyzing many other mitochondrial genes, Goodman and colleagues hypothesized that the rapid evolution of mitochondrial component genes may be responsible for the development of brains in the ape lineage, which consume much more energy than other organs (Grossman et al. 2004).

2.3.2.2 Finding Rapidly Evolving Genes Between Humans and Chimpanzees

After the chimpanzee and macaque genome sequences were published (The Chimpanzee Sequencing and Analysis Consortium 2005; Gibbs et al. 2007),

genome-wide comparison of transcript sequences became easier and much more popular. Combined with human polymorphism data, recent studies have established that several genes related to brain function evolved rapidly under positive selection after the human–chimpanzee divergence. These genes include the forkhead box P2 (*FOXP2*) (Enard et al. 2002b), abnormal spindle homologue, microcephaly associated (*ASPM*) (Zhang 2003; Mekel-Bobrov et al. 2005), and microcephalin 1 (*MCPH1*) (Evans et al. 2005). *FOXP2* is known to be involved in human genetic diseases related to vocalization. The other two genes are involved in brain development. Laboratory experiments in mice carrying genetically modified *FOXP2* revealed that disruption or mutations in *FOXP2* changed the ultrasonic vocalization and behavior of mice (Shu et al. 2005; Fujita et al. 2008; Groszer et al. 2008). The functions of these genes and their molecular evolution pattern suggest that they are good candidates for contributing to human-specific cognitive abilities.

These studies demonstrated that the function of some genes in the brain may have rapidly evolved under positive selection in humans. Then, what about the other brain-expressed genes? Dorus et al. (2004) reported that hundreds of genes related to the nervous system have evolved more rapidly in the human lineage than in macaques. This finding suggests that not a few neuronal genes have evolved under positive selection. On analyzing brain-expressed genes more extensively, however, researchers observed an opposite trend and concluded that brain-expressed genes may have evolved more slowly in the human lineage than in chimpanzees (Wang et al. 2007). The nearly neutral theory of molecular evolution predicts that any set of genes could have evolved more rapidly in the human lineage than in chimpanzees simply because of the smaller effective size of the human population, in which the elimination of slightly deleterious mutations is less effective (Ohta 1973). Therefore, to find a relative evolution rate in brain-expressed genes, it is necessary to calibrate rate with respect to the genome-wide average. The results of Wang et al. (2007) showed that although some functionally important genes have evolved rapidly because of positive selection, most brain-expressed genes have evolved rather slowly in the human lineage. This idea has been explained by the hypothesis that in tissues that have a more complex gene network, functional constraints on genes are stronger than in tissues which have a simpler gene network. As a result, protein sequences of brain-expressed genes in humans have evolved at an extremely slow rate. A similar trend has been reported by Shi et al. (2006). These findings, in turn, suggest that the changes in gene expression may be more important than those in protein sequences for the evolution of human brain. The pattern of gene expression evolution in the human brain is summarized later.

Although the brain is one of the most fascinating organs in humans, another class of genes – the immune genes – have been found to evolve rapidly according to many genome-wide comparisons between human and chimpanzee genomes (Clark et al. 2003; The Chimpanzee Sequencing and Analysis Consortium 2005; Nielsen et al. 2005). Pathogens evolve to adapt to the host immune system, and the host immune systems evolve to defend against pathogens. This process is sometimes referred to as an arms race. Indeed, many pathogens are known to specifically infect humans and not chimpanzees. Some pathogens, such as human immunodeficiency virus