

Introductory Article

An introduction to genes, genomes and disease

Peter A Hall,*¹ Jorge S Reis-Filho,² Ian PM Tomlinson³ and Richard Poulson⁴

¹Institute of Pathology, School of Medicine, Dentistry and Biomedical Science, Queen's University of Belfast, Belfast BT12 6BL, UK

²The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

³Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

⁴Histopathology Laboratory, Cancer Research UK, London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

*Correspondence to:

Peter A Hall, Institute of Pathology, School of Medicine, Dentistry and Biomedical Science, Queen's University of Belfast, Belfast BT12 6BL, UK.
E-mail: peter.hall@qub.ac.uk

No conflicts of interest were declared.

Abstract

The human and other genome projects and subsequent resequencing programmes have provided new perspectives on the nature of the gene and how genes function. Understanding the complexity of the eukaryotic nucleus and the diversity of genetic regulatory mechanisms, including the role of non-coding RNAs, translational control mechanisms and the extraordinary prevalence of splicing, will be central to understanding how genes function, as will the recognition of gene dosage issues. This introduction to the 2010 Annual Review Issue, Genes, Genomes and Disease, provides overviews of these areas and then considers their relevance to a range of human diseases, including cardiovascular and renal disease, neural tube defects and cancer. The *p53* gene is considered as an example of a massively regulated gene and the genetic perturbations in cancer are considered in a historical perspective. High-throughput genomic and transcriptomic methods have led to a paradigm shift in the way cancers are perceived and have changed the way translational research is performed. The progress in our understanding of chromosomal rearrangements in cancer, once believed to be incredibly rare events in epithelial malignancies, is discussed. The identification of low-penetrance cancer susceptibility genes through genome-wide association studies and their implications are reviewed. The contribution and limitations of expression profiling are discussed. In the last series of reviews, future challenges are addressed: the promise of synthetic lethality strategies in cancer therapy, a case for 'systems' approaches to genetic networks and the potential of single molecule genetic technologies. Finally, the question 'Does massively parallel DNA resequencing signify the end of histopathology as we know it?' is posed. Readers should find that the 2010 Annual Review Issue is an invaluable resource on contemporary genetics and its applications to understanding disease.

Copyright © 2009 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.

Received: 6 October 2009

Revised: 6 October 2009

Accepted: 6 October 2009

Keywords: gene; genome; disease; miRNA; ncRNA; chromosome territory; gene regulation; p53; technology; sequencing; high throughput; gene regulation; copy number variation

The impact of modern molecular biology on our understanding of disease and disease processes has been dramatic and inexorable. Barely a quarter of a century ago, the cloning and sequencing of a gene were perceived as monumental tasks, and the understanding of its regulation and the properties of its encoded polypeptide was a Herculean effort. DNA sequencing was in its infancy and was slow, labour-intensive and error-prone. While there were, even then, promises of genetic information having direct clinical impact, this was, in the 1980s, hardly a realized goal. Moreover, our concept of a gene was perhaps simplistic and naïve and our perception of genetic regulation and gene structure were, in the context of mammalian systems, no more than embryonic. Nevertheless, considerable progress was

made, even if it was patchy and disjointed. It did, of course, preface the era of the human genome project with its complex and troublesome conception and gestation, which has been reviewed by key participants [1,2]. The impact of the knowledge gained from the sequencing of the human and many other genomes has been astonishing and there have been dramatic consequences for our understanding of how genes and genomes work. Indeed, our view of the genome has changed [3,4], with a new perception of the gene as a complex highly regulated physiological unit, as opposed to being merely a region that encodes a polypeptide. We now recognize there to be widespread transcription from much of the genome [5] and that non-(protein) coding transcripts are extraordinarily abundant [6]. Indeed, it is also becoming clear that

there may often not be the colinearity between gene sequences and resultant polypeptides that dogma has long suggested [7].

Our developing understanding of the complexity of genes and their regulation and the complex nature of nuclear organization does, of course, have two important consequences. First, it provides an intellectual framework for the resolution of the gene number paradox, ie the relative paucity of genes in higher metazoan genomes as compared with (for example) yeast. There are perhaps no more than four times as many genes in mammalian genomes compared with *Saccharomyces cerevisiae* (~6300), with flies being intermediate with ~15 000 genes. Clearly the architectural and developmental complexity emanates from regulatory complexity, as argued cogently by Davidson [8]. The second and crucial consequence lies in our understanding of human disease, which can only be understood at a mechanistic level by considering our new and developing knowledge of genes and genomes. The focus of this Annual Review Issue (the 11th in the series) is to consider the biology of genes and genomes and how they can go wrong in disease. We were prompted to focus this issue on this area for many reasons, but were persuaded by the rapid development of technologies that are opening our eyes to the multilayered nature of genetic information, and that are truly leading medical research to a point where widespread clinical applications are becoming a reality, as opposed to unfulfilled promises.

The first part of our Annual Review Issue focuses on our developing understanding of basic regulatory mechanisms that control gene function and expression. The simple concept that the nucleus is merely a haphazard arrangement of DNA and histones was long ago dispelled and we now recognize that there is a well-defined architecture and structure to this organelle. Lever and Sheer [9] overview this topic and provide a clear review of the essential anatomy of the nucleus and how this modulates its physiology and pathophysiology. The unravelling of these large-scale factors is but one part of our new genetic insight into cellular physiology. The central role of non-coding RNAs (ncRNAs) in gene regulation was discovered more than a decade ago and is now seen to be relevant in the control of many cellular processes and can be deranged in diverse disease states. Compelling evidence for this comes from many areas and the role of ncRNAs in haematopoiesis and its reregulation in haematopoietic malignancy has been a major stimulus in the field [10]. In this issue, Taft, Pang, Mercer, Dinger and Mattick [11] provide a comprehensive overview of ncRNAs and their role in disease. This is timely, given the recent award of the first Jeremy Jass Prize for Research Excellence in Pathology to work in the epigenetic regulation of miRNA in breast cancer [12]. One area where ncRNA can have a clear role is in the context of regulation of protein synthesis, and work in many laboratories has opened our eyes to perturbation of this having

pathogenetic roles; LeQuesne, Spriggs, Bushell and Willis crystallize these ideas in their review [13]. Furthermore, analyses of ncRNAs in different types of human cancers have helped resolve classification dilemmas [14–16] and led to the identification of genes that may be involved in the metastatic process [17].

The use of alternate splicing to create diverse transcripts and hence protein isoforms has been recognized since the seminal observations of Sharp and Roberts, who received the Nobel Prize for their discovery [18]. While this is well-recognized to occur in some genes, the extraordinary prevalence of splicing in mammalian genomes has only recently become apparent [19]. Not only does this create diversity of potential protein isoforms, eg in p53 [20–22], it also creates yet more layers of potential control [22,23]. The pathobiology of splicing is here reviewed by Ward and Cooper [24].

The idea that gene dosage can have major implications for cellular function is not a new one but the observation of copy number variation (CNV) has reawakened our interest in this phenomenon [25]. The relevance to human disease is becoming apparent with reports of CNV in neuropsychiatric conditions [26], cardiovascular disease [27], neoplasia [28,29] and many other conditions. Methods for easy analysis of CNV are being reported [30,31] and open the door to a comprehensive analysis of this in clinical samples, and hence the potential for a deeper understanding of the contribution of CNV to disease and disease susceptibility. In fact, the systematic analysis of CNVs has called into question the concept of the 'normal', entirely diploid human genome, given that distinct CNVs can be found in the genomes of different somatic tissues from the same individual [32] and between identical twins [33]. Building on a recent review of genetic dominance and dominant negative mechanisms in disease [34], Veitia and Birchler explore the whole area of gene dosage (including haploinsufficiency and CNV) and its relevance to disease processes [35].

With this background of contemporary perspectives on some of the core mechanisms of genetic regulation (and dysregulation in disease), we then turn to the impact of such information in areas of clinical importance. We could not be comprehensive in coverage but rather have focused on four areas where advances are being turned into real clinical progress, namely cardiovascular [36] and renal disease [37], neural tube defects [38] and cancer [39–44]. Chico, Milo and Crossman discuss the impact of emerging technologies, and information from model systems such as zebrafish as well as clinical studies, on our understanding of cardiovascular disease [36], while McKnight, Currie and Maxwell comprehensively summarize our current knowledge of the role of genetic factors in chronic renal disease [37]. Neural tube defects (NTD) cause untold misery for affected children and their families, and in this issue Copp and Greene link our extensive knowledge of the 200 or so genes that cause

are many general points of importance in this comprehensive and rigorous analysis of this field. Another area where progress has been promised but has sometimes been slow is in the development of therapeutics based on our developing genetic and genomic insights. Ashworth's group has been at the forefront of using innovative high-throughput strategies to tailor therapies to specific genetic abnormalities, and in particular the use of synthetic lethal approaches to tumours with defective DNA repair mechanisms. Here Martin, Hewish and Ashworth review this important topic [44] and Erler and Linding develop the notion that considering genetics in a 'systems' context will provide new possibilities in terms of biomarkers and novel therapeutic strategies and enhance our understanding of the genetics of disease [45].

Ultimately, the possibility of our knowledge of genes and genetics having an impact on clinical care (prevention, diagnosis, prognostication or treatment) will depend on affordable, fast and easily applicable technologies. Whilst we have many powerful tools already, the potential for fast, cheap sequencing is central to future applications, and McCaughan and Dear [31] review for us the current state of play of single-molecule approaches to genomics, which have already had so much impact in genetic research.

We end this Annual Review Issue with some speculation. A decade ago, Aparicio, Ponder and Caldas speculated whether massively parallel transcriptome analysis might signify the end of cancer histopathology as we know it [46]. For many reasons (see eg [43]), the impact of expression array analysis in clinical practice was much less than had been anticipated. However, 10 years on we asked Aparicio and Huntsman to revisit this class of question [47] and ask whether massively parallel DNA resequencing might signify the end of histopathology as we know it. The next few years will surely see huge strides and the impact of the new technologies and our developing understanding of genes, genomes and disease. The days of 'fishing with chips' may be numbered [48] and single-molecule genomics promises cheap, fast, high-throughput genetic analysis [31,49]: this may be the way forward. Real clinical impact cannot be far away!

Teaching materials

A PowerPoint slide of the figure from this review is supplied as supporting information in the online version of this article.

References

- Sulston J, Ferry G. *The Common Thread — A Story of Science, Politics, Ethics and the Human Genome*. Bantam Press: London, 2002.
- Venter JC. *A Life Decoded: My Genome: My Life*. Penguin: London, 2008.
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res* 2007;**17**:669–681.
- Pearson H. Codes and enigmas. *Nature* 2006;**444**:259–261.
- Gingeras TR. Origin of phenotypes: genes and transcripts. *Genome Res* 2007;**17**:682–690.
- Wu JQ, Du J, Rozowsky J, Zhang Z, Urban AE, Euskirchen G, et al. Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. *Genome Biol* 2008;**9**:R3.
- Gingeras TR. Implications of chimaeric non-colinear transcripts. *Nature* 2009;**461**:206–211.
- Davidson E. *The Regulatory Genome. Gene Regulatory Networks in Development and Evolution*. Academic Press: Amsterdam, 2006.
- Lever E, Sheer D. The role of nuclear organization in cancer. *J Pathol* 2009;DOI:10.1002/path.2651
- Garzon R, Croce CM. MicroRNAs in normal and malignant hematopoiesis. *Curr Opin Hematol* 2008;**15**:352–358
- Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS. Noncoding RNAs: regulators of disease. *J Pathol* 2009;DOI:10.1002/path.2638
- Lehmann U, Hasemeier B, Christgen M, Müller M, Römermann D, Länger F, et al. Epigenetic inactivation of microRNA gene *hsa-mir-9-1* in human breast cancer. *J Pathol* 2008;**214**:17–24.
- Le Quesne JPC, Spriggs KA, Bushell M, Willis AE. Dysregulation of protein synthesis and disease. *J Pathol* 2009;DOI:10.1002/path.2627
- Gillis AJ, Stoop HJ, Hersmus R, Oosterhuis JW, Sun Y, Chen C, et al. High-throughput microRNAome analysis in human germ cell tumours. *J Pathol* 2007;**213**:319–328.
- Leucci E, Cocco M, Onnis A, De Falco G, van Cleef P, Bellan C, et al. MYC translocation-negative classical Burkitt lymphoma cases: an alternative pathogenetic mechanism involving miRNA deregulation. *J Pathol* 2008;**216**:440–450.
- Nakada C, Matsuura K, Tsukamoto Y, Tanigawa M, Yoshimoto T, Narimatsu T, et al. Genome-wide microRNA expression profiling in renal cell carcinoma: significant down-regulation of miR-141 and miR-200c. *J Pathol* 2008;**216**:418–427.
- Baffa R, Fassan M, Volinia S, O'Hara B, Liu CG, Palazzo JP, et al. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *J Pathol* 2009;**219**:214–21.
- http://nobelprize.org/nobel_prizes/medicine/laureates/1993/index.html (accessed 2 October 2009).
- Castle JC, Zhang C, Shah JK, Kulkarni AV, Kalsotra A, Cooper TA, et al. Expression of 24 426 human alternative splicing events and predicted *cis*-regulation in 48 tissues and cell lines. *Nat Genet* 2008;**40**:1416–1425.
- Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, et al. p53 isoforms can regulate p53 transcriptional activity. *Genes Dev* 2005;**19**:2122–2137.
- Fujita K, Mondal AM, Horikawa I, Nguyen GH, Kumamoto K, Sohn JJ, et al. p53 isoforms $\Delta 133p53$ and p53 β are endogenous regulators of replicative cellular senescence. *Nat Cell Biol* 2009;**11**:1135–1142.
- Hollstein M, Hainaut P. Massively regulated genes: the example of p53. *J Pathol* 2009;DOI:10.1002/path.2637
- McDade SS, Hall PA, Russell SE. Translational control of SEPT9 isoforms is perturbed in disease. *Hum Mol Genet* 2007;**16**:742–752.
- Ward AJ, Cooper TA. The pathobiology of splicing. *J Pathol* 2009;DOI:10.1002/path.2649
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;**453**:56–64.
- Tam GW, Redon R, Carter NP, Grant SG. The role of DNA copy number variation in schizophrenia. *Biol Psychiat* 2009;DOI:10.1016/j.biopsych.2009.07.027
- Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, et al. *De novo* copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet* 2009;**41**:931–935.

28. Diskin SJ, Hou C, Glessner JT, Attiyeh EF, Laudenslager M, Bosse K, *et al.* Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* 2009;**459**:987–991.
29. Forsshew T, Tatevossian RG, Lawson AR, Ma J, Neale G, Ogunkolade BW, *et al.* Activation of the ERK/MAPK pathway: a signature genetic defect in posterior fossa pilocytic astrocytomas. *J Pathol* 2009;**218**:172–181.
30. McCaughan F, Darai-Ramqvist E, Bankier AT, Konfortov BA, Foster N, George PJ, *et al.* Microdissection molecular copy-number counting (microMCC) — unlocking cancer archives with digital PCR. *J Pathol* 2008;**216**:307–316.
31. McCaughan F, Dear P. Single-molecule genomics. *J Pathol* 2009;DOI:10.1002/path.2647
32. Piotrowski A, Bruder CE, Andersson R, de Ståhl TD, Menzel U, Sandgren J, *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* 2008;**29**:1118–1124.
33. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, de Ståhl TD, *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet.* 2008;**82**:763–771.
34. Veitia RA. Dominant negative factors in health and disease. *J Pathol* 2009;**218**:409–418.
35. Veitia RA, Birchler JA. Dominance and gene dosage balance in health and disease: why levels matter! *J Pathol* 2009;DOI:10.1002/path.2623
36. Chico TJA, Milo M, Crossman DC. The genetics of cardiovascular disease: new insights from emerging approaches. *J Pathol* 2009;DOI:10.1002/path.2641
37. McKnight AJ, Currie D, Maxwell AP. Unravelling the genetic basis of renal diseases; from single gene to multifactorial disorders. *J Pathol* 2009;DOI:10.1002/path.2639
38. Copp AJ, Greene NDE. Genetics and development of neural tube defects. *J Pathol* 2009;DOI:10.1002/path.2643
39. Bell DW. Our changing view of the genomic landscape of cancer. *J Pathol* 2009;DOI:10.1002/path.2645
40. Edwards PAW. Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol* 2009;DOI:10.1002/path.2632
41. Cazier J-B, Tomlinson IPM. General lessons from large-scale studies to identify human cancer predisposition genes. *J Pathol* 2009;DOI:10.1002/path.2650
42. Zuckerman V, Wolyniec K, Sionov RV, Haupt S, Haupt Y. Tumour suppression by p53: the importance of apoptosis and cellular senescence. *J Pathol* 2009;**219**:3–15.
43. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J Pathol* 2009;DOI:10.1002/path.2648
44. Martin SA, Hewish M, Lord CJ, Ashworth A. Genomic instability and the selection of treatments for cancer. *J Pathol* 2009;DOI:10.1002/path.2631
45. Erler JT, Linding R. Network-based drugs and biomarkers. *J Pathol* 2009;DOI:10.1002/path.2646
46. Aparicio SA, Caldas C, Ponder B. Does massively parallel transcriptome analysis signify the end of cancer histopathology as we know it? *Genome Biol.* 2000;**1**:REVIEWS 1021.
47. Aparicio SA, Huntsman DG. Does massively parallel DNA resequencing signify the end of histopathology as we know it? *J Pathol* 2009;DOI:10.1002/path.2636
48. Ledford H. The death of microarrays? *Nature* 2008;**455**:847.
49. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**:57–63.