

## Illuminating the Genome's Dark Matter

### *The Deeper Genome: Why There Is More to the Human Genome Than Meets the Eye*

Author: John Parrington

New York, NY, USA: Oxford University Press (2015).

272 pp. \$29.95

John Parrington's book *The Deeper Genome* provides us with a closer look at the enigma of junk DNA. Often referred to as the dark matter of the genome, junk DNA is an important but often overlooked part of the human genome. (Perhaps, in light of recent discoveries, this term might be somewhat problematic; however, it is the term that Parrington uses and we will follow that here.) The dark matter metaphor is particularly apt: akin to the great expanses of dark matter within our universe, junk DNA makes up the vast majority of the genome. And just like dark matter, whose properties can only be inferred from its effects on observable matter, much of our junk DNA can only currently be characterized by its regulatory effects on the smaller fraction of coding DNA. Further, just like dark matter is presumed to account for discrepancies in our understanding of the universe—between what's theoretically predicted versus actually observed—junk DNA is thought to account for the majority of associations with genetic diseases found by genome-wide association studies (GWAS). Insofar as this is the case, junk DNA is medically relevant and contains actionable information. Finally, just like dark matter contains historical evidence of events from the primordial universe, junk DNA, lacking the evolutionary pressures on protein-coding genes, contains vestigial information relating to the molecular history of our species.

*The Deeper Genome* is unique in that it provides an entertaining tale of personalities with lots of useful technical knowledge. We highly recommend the book as a supplement for classroom teaching as it covers fundamental concepts in an easily readable format.

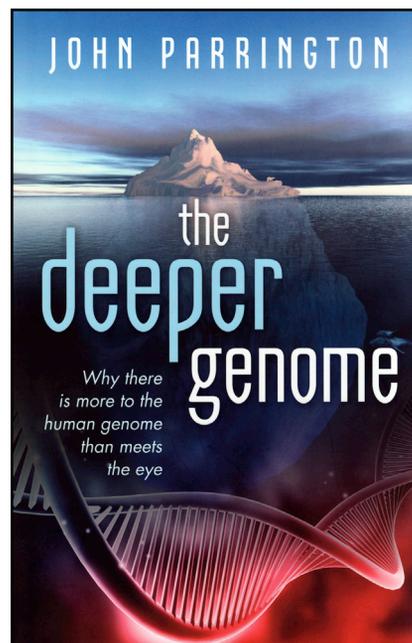
Students will likely find it even more interesting than many of the more standard textbooks. For example, the book

goes through the classic story of the double helix, starting all the way from Aristotle and eventually working its way through Darwin, Wallace, Mendel, and Morgan to more contemporary times, with Gilbert and Sanger. To wit, Parrington ably describes how early abstractions related to genetic inheritance were concretized over time and culminates in quite readable descriptions of current scientific research and experimentation. He provides lucid textual descriptions of the canonical theory of Darwinian evolution and its synthesis with the work of Mendel by Fisher and others. He even covers a somewhat forgotten alternative theory: the pre-Darwinian Lamarckian evolutionary theory and its recent resurgence—e.g., in Eugene Koonin's two-stroke process, with an initial Lamarckian epigenetics phase and a subsequent Darwinian selection phase. Finally, the

book contains a great description of the actual structure of DNA, excellently describing in prose the intricate 3D structure of the sugar-phosphate backbone. It also has a number of spare but incisive figures to illustrate key concepts, without detracting from the primarily textual description.

In one of its most interesting parts, the book expertly provides both the scientific and historical progression of how repressors were first abstractly described as fundamental principles of gene regulation by Nobel laureates Jacob and Monod through to the recent in-depth molecular understanding of these as transcription factors. Here, in elucidating aspects of scientific culture—like he does in a number of other places in the book—the author quotes James Watson in describing the relationship between Walter Gilbert and Mark Ptashne, who isolated and characterized the first transcription factors: “take young researchers, put them together in virtual seclusion, give them an unprecedented degree of freedom, and turn up the pressure by fostering competitiveness.” This is what makes this book particularly enjoyable—Parrington's ample use of tangential anecdotes and sidebars relating to the very-human aspects of the scientific process.

The book also provides an extensive discussion on how genes are linked to disease, with a particular focus on classic examples such as Huntington's, Phenylketonuria (PKU), and cystic fibrosis. Parrington helps the reader understand an important distinction in the nascent but quickly growing field of genetic testing: he provides clear examples that distinguish across the spectrum of informative to actionable genetic tests. For some diseases, such as Huntington's, we can test for the statistical likelihood of developing the late-onset disease but otherwise cannot provide medical intervention to prevent its onset. For other diseases like cystic fibrosis, which has become part of the standard prenatal genetic testing panel, parents can be forewarned and incorporate genomic technologies to prevent having children with the disease. And for yet a third group, like PKU, genetic testing can lead to early medical intervention such as dietary modifications that will greatly alleviate the disease effects. In



combining his research training and journalistic credentials, Parrington seamlessly delves into original research papers in the scientific literature and also describes their interpretation in the popular press. In particular, he contrasts the excitement surrounding the initial sequencing of the human genome in 2000 with the current sobering reflection that the expected breakthroughs have yet to arrive. In light of Parrington's time as a journalist actually covering the ENCODE rollout, the book uses the post-2012 ENCODE debate to frame his analysis and theories for junk DNA. While this is an effective method of piquing interest in the subject matter, Parrington sometimes cherry picks quotes to exaggerate the storminess of the debate. Parrington, whose coverage in this area was focused on the British protagonists, sometimes also provided less information regarding the global perspective on the ENCODE effort.

The book concludes with a discussion of human origin and neurogenomics. There is a nice description of how RNA sequencing studies comparing human and chimp brains give some insight into the developmental time of synapse formation and what the comparisons with the Neanderthal genome actually mean.

Finally, the book tempts the reader to think about how the genome project relates to its progenies: efforts to catalog all proteins (the proteome), RNA molecules (the transcriptome), metabolites (the metabolome), interactions (the interactome), and the foreshadowed neuroscience project relating to the connectome.

As this book is nominally devoted to discussing the genome, it would have been informative to discuss some other prominent genomics projects in greater detail—particularly those relevant to constraint in non-coding regions, such as the HapMap, 1000 Genomes, and GTEx Projects. Parrington could have also included more about related innovations, such as next-generation DNA sequencing, that make many of these large projects possible. It would have also been helpful to have a more technical discussion related to selection—i.e., exactly what is the absence of constraint (i.e., the state of being “junk”) and how does one measure this with well-known statistics such as the enrichment of rare alleles and the ratio of synonymous to non-synonymous changes.

Additionally, given the book's focus on junk DNA, it seems strange that there isn't any mention of the most common

constituents of junk DNA: the LINE and Alu elements. Likewise, the discussion of pseudogenes could have delved further and included, for example, recent research that suggests that transcribed pseudogenes actually function as regulatory non-coding RNAs, rather than as translated proteins.

Finally, it would have been interesting for Parrington to have drawn out conclusions and actionable repercussions of interest to the lay reader. In particular, issues of privacy vis-a-vis the genome loom large for the public, yet they are given little attention in the book. For example, an issue of notable significance to the lay public has been the use of non-coding junk DNA sequences by the police and others in DNA fingerprinting. In light of ENCODE, the public's concerns have been compounded—that databases of heretofore identifying but not descriptive DNA may now be more descriptive than previously thought, raising, among other concerns, significant privacy issues. That said, *The Deeper Genome* is a great read that definitely imparts knowledge in an entertaining fashion and connects the almost 99 percent of the genome that is not protein coding to all sorts of interesting questions. We highly recommend it.

**Dov Greenbaum<sup>1,2,\*</sup>  
and Mark Gerstein<sup>2,\*</sup>**

<sup>1</sup>Interdisciplinary Center Herzliya, Herzliya, 4610101, Israel

<sup>2</sup>Yale University, New Haven, CT 06520, USA

\*Correspondence: [dov.greenbaum@yale.edu](mailto:dov.greenbaum@yale.edu) (D.G.), [mark@gersteinlab.org](mailto:mark@gersteinlab.org) (M.G.)

<http://dx.doi.org/10.1016/j.cell.2015.10.073>