



11 Years Ago

The draft human genome

OUR GENOME UNVEILED

Unless the human genome contains a lot of genes that are opaque to our computers, it is clear that we do not gain our undoubted complexity over worms and plants by using many more genes. Understanding what does give us our complexity — our enormous behavioural repertoire, ability to produce conscious action, remarkable physical coordination (shared with other vertebrates), precisely tuned alterations in response to external variations of the environment, learning, memory ... need I go on? — remains a challenge for the future.

David Baltimore

From *Nature* 15 February 2001

GENOME SPEAK

With the draft in hand, researchers have a new tool for studying the regulatory regions and networks of genes. Comparisons with other genomes should reveal common regulatory elements, and the environments of genes shared with other species may offer insight into function and regulation beyond the level of individual genes. The draft is also a starting point for studies of the three-dimensional packing of the genome into a cell's nucleus. Such packing is likely to influence gene regulation ... The human genome lies before us, ready for interpretation.

Peer Bork and Richard Copley

From *Nature* 15 February 2001

more than 2 million putative enhancers without known targets, revealing the enormous expanse of the regulatory genome landscape that is yet to be explored. Chromosome-conformation-capture methods that detect long-range physical associations between distant DNA regions are attempting to bridge this gap. Indeed, Sanyal and colleagues⁷ applied these techniques to survey such associations across 1% of the genome.

The ENCODE data start to paint a picture of the logic and architecture of transcriptional networks, in which DNA binding of a few high-affinity transcription factors displaces nucleosomes and creates a DHS, which in turn facilitates the binding of further, lower-affinity factors. The results also support the idea that transcription-factor binding can block DNA methylation (a chemical modification of DNA that affects gene expression), rather than the other way around — which is highly relevant to the interpretation of disease-associated sites of altered DNA methylation¹¹.

The exquisite cell-type specificity of regulatory elements revealed by the ENCODE studies emphasizes the importance of having appropriate biological material on which to test hypotheses. The researchers have focused their efforts on a set of well-established cell lines, with selected assays extended to some freshly isolated cells. Challenges for the future include following the dynamic changes in the regulatory landscape during specific developmental pathways, and understanding chromatin structure in tissues containing heterogeneous cell populations.

Wendy A. Bickmore is in the Medical Research Council Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
e-mail: wendy.bickmore@igmm.ed.ac.uk

Non-coding but functional

INÉS BARROSO

The vast majority of the human genome does not code for proteins and, until now, did not seem to contain defined gene-regulatory elements. Why evolution would maintain large amounts of 'useless' DNA had remained a mystery, and seemed wasteful. It turns out, however, that there are good reasons to keep this DNA. Results from the ENCODE project²⁻⁸ show that most of these stretches of DNA harbour regions that bind proteins and RNA molecules, bringing these into positions from which they cooperate with each other to regulate the function and level of expression of protein-coding genes. In addition, it seems that widespread transcription from non-coding

DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs.

What are the implications of these results for genetic studies of complex human traits and disease? Genome-wide association studies (GWAS), which link variations in DNA sequence with specific traits and diseases, have in recent years become the workhorse of the field, and have identified thousands of DNA variants associated with hundreds of complex

“The results imply that sequencing studies focusing on protein-coding sequences risk missing crucial parts of the genome.”

traits (such as height) and diseases (such as diabetes). But association is not causality, and identifying those variants that are causally linked to a given disease or trait, and understanding how they exert such influence, has been difficult. Further-

more, most of these associated variants lie in non-coding regions, so their functional effects have remained undefined.

The ENCODE project provides a detailed map of additional functional non-coding units in the human genome, including some that have cell-type-specific activity. In fact, the catalogue contains many more functional non-coding regions than genes. These data show that results of GWAS are typically enriched for variants that lie within such non-coding functional units, sometimes in a cell-type-specific manner that is consistent with certain traits, suggesting that many of these regions could be causally linked to disease. Thus, the project demonstrates that non-coding regions must be considered when interpreting GWAS results, and it provides a strong motivation for reinterpreting previous GWAS findings. Furthermore, these results imply that sequencing studies focusing on protein-coding sequences (the 'exome') risk missing crucial parts of the genome and the ability to identify true causal variants.

However, although the ENCODE catalogues represent a remarkable tour de force, they contain only an initial exploration of the depths of our genome, because many more cell types must yet be investigated. Some of the remaining challenges for scientists searching for causal disease variants lie in: accessing data derived from cell types and tissues relevant to the disease under study; understanding how these functional units affect genes that may be distantly located⁷; and the ability to generalize such results to the entire organism.

Inés Barroso is at the Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK, and at the University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.
e-mail: ib1@sanger.ac.uk