

INVITED SPEAKER PRESENTATION

Open Access

Most of the 6.5% - 10% of human DNA bases that are functional now will soon be turned over

Chris P Ponting^{1*}, Stephen J Meader¹, Gerton Lunter²

From Beyond the Genome: The true gene count, human evolution and disease genomics
Boston, MA, USA. 11-13 October 2010

Background

Despite the availability of an increasing number of mammalian genome sequences, and the considerable effort devoted to their analysis, two key questions still provoke much debate. (i) What fraction of a genome confers biological function, as opposed to the remaining proportion that has had no biological effect and thus has not been subject to selection? By careful scrutiny of protein-coding gene models it has become clear that approximately 1.06% of the human genome encodes (functional) protein-coding sequence. An even larger fraction of the genome has been inferred to contain functional sequence but estimates of this fraction's size have proved particularly contentious. (ii) Do genomes of different species contain different amounts of functional sequence, and is this measure related to organismal complexity? Similar numbers of protein-coding genes among diverse species suggests the possibility that our naive notion of complexity is fundamentally incorrect, and that many species are comparably complex, in a sense yet to be defined. Alternatively, it may be that much of the apparent differences in complexity between species are reflected by varying amounts of functional non-coding sequence.

Results

By applying Lunter's Neutral Indel Model to genomes drawn from pairs of diverse metazoans, we have been able to estimate that between 200 and 300 Mb (~6.5 - 10%) of the human genome is under functional constraint; this includes 5-8 times as many constrained non-coding bases than bases that encode proteins. By contrast, in *Drosophila melanogaster* only 56-66 Mb appear to be constrained, implying a ratio of non-coding

to coding constrained bases of ~ 2. This suggests that, rather than genome size or protein-coding gene complement, it is the number of functional bases that might best mirror our naïve preconceptions of organismal complexity. Furthermore, we observe that as the divergence between mammalian species increases, the predicted amount of pairwise shared functional sequence drops off dramatically, approximately halving in 90 million years of eutherian evolution.

Conclusion

These results provide strong evidence for the existence of substantial amounts of functional and mostly non-coding nucleotides that are specific to sub-clades of the mammalian phylogeny. Furthermore, mammalian genomes are predicted to contain greater amounts of putative functional bases than genomes of fish and fruit flies.

Author details

¹MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, OX1 3QX, UK. ²The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK.

Published: 11 October 2010

doi:10.1186/gb-2010-11-S1-110

Cite this article as: Ponting *et al.*: Most of the 6.5% - 10% of human DNA bases that are functional now will soon be turned over. *Genome Biology* 2010 11(Suppl 1):110.

¹MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, OX1 3QX, UK
Full list of author information is available at the end of the article