

upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* 2004; 74: 1102–1110.

14 Olds LC, Sibley E: Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Hum Mol Genet* 2003; 12: 2333–2340.

15 Troelsen JT *et al*: An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* 2003; 125: 1686–1694.

## Statistical Genetics

# Usual suspects in complex disease

CP Ponting and L Goodstadt

*European Journal of Human Genetics* (2005) 13, 269–270.

doi:10.1038/sj.ejhg.5201354

Published online 15 December 2004

Surprising observations are made by a recent study scrutinising the aetiology of complex diseases and the genetic mutations that they spring from.

Finding the common DNA variants that contribute significantly to genetic risk for common diseases is a key goal for medical science.<sup>1</sup> However, while distinguishing mutations that cause disease from harmless single nucleotide polymorphisms (SNPs) is difficult enough for the relatively few genetic diseases that are inherited in a simple Mendelian manner, it is a daunting task for complex traits where pathology arises from the interactions of multiple genes with the environment. Point mutations exert a broad spectrum of effects on human health. The most fearsome are those that disrupt development: they cause embryonic morbidity and are seldom observed in postnatal disease; at the other end of the scale are base substitutions under few constraints, such as most common SNPs, and in between fall most of the 1000 or so<sup>2,3</sup> deleterious mutations that are carried by the average person. In a recent article in *Proc. Natl. Acad. Sci. USA*, Thomas and Kejariwal<sup>4</sup> show what types of coding mutations we should expect in complex diseases. They find that these amino-acid changes mostly fall outside of conserved regions and cannot readily be distinguished from the coding sequence variation seen between healthy individuals.

These results are surprising because most amino-acid substitutions associated with Mendelian diseases are of conserved, and thus presumably essential, amino acids.<sup>2,3,5,6</sup> Hitherto, there has been little

evidence that this would be any different for complex diseases. Thomas and Kejariwal give three possible explanations for their findings.

First, they suggest that coding mutations in complex disease might cause subtle and almost imperceptible alterations to molecular function. If this were true, researchers proposing molecular dysfunction in complex diseases would not be able to use sequence evolutionary information to prosecute their case. Mutations that are mildly deleterious are difficult enough to substantiate experimentally for Mendelian diseases; for complex traits that are multifactorial, it may well be impossible to detect the knockon effects of such subtle alterations.

Thomas and Kejariwal also entertain the possibility that some of the 37 cases of coding SNPs they examined might not directly contribute to complex diseases. They felt that this explanation was unlikely because their findings remain significant even for a reduced number of cases, those about which they were most confident. Nevertheless, it cannot be discounted that many of these coding SNPs, instead of directly contributing to the complex disease, may merely be closely linked to the real disease-causing polymorphisms. These may lie, for example, in adjacent noncoding sequence that regulates transcription or translation. This suggestion should find favour among those advocating hunts for causative SNPs within regulatory regions, for example, King and Wilson<sup>7</sup> and Prokunina *et al.*<sup>8</sup>

The authors' final explanation is that lack of conservation may not rule out the functional importance of their disease-associated SNPs if these functions have been acquired only recently in primate evolution. Comparisons with more distantly related mammals might not show conservation if SNP sites have been evolving rapidly under adaptive pressures in our lineage. Adaptive evolution can be detected if  $K_A/K_S$  ratios<sup>9</sup> between mouse and human genes are greater than one. However, though the  $K_A/K_S$  ratios for complex disease genes are elevated relative to randomly selected genes, the ratios are still much less than one. In any case, the issue here is whether, out of all the many codons in a gene, a particular complex disease-associated SNP has been changing adaptively. This cannot be determined simply by comparing the entirety of one gene from one species with that from another. At best, the jury is still out on this question.

A re-examination of the same  $K_A/K_S$  data, however, indicates a functional bias in complex disease genes that was not noted in the original article. Secreted proteins and transmembrane molecules, such as receptors, are greatly over-represented among those encoded by the 32 complex disease genes that Thomas and Kejariwal analysed. These number 8 (25%) and 15 (47%) out of 32, respectively, far more than just the six of each type expected from their frequency (~20%) in the human genome.<sup>10</sup> Transmembrane and secreted proteins evolve more rapidly than average.<sup>11</sup> This alone would explain much of the elevation in  $K_A/K_S$  ratios seen for complex disease genes. Transmembrane and secreted proteins would be prime suspects in complex disease as they tend to have restricted expression profiles: the dysfunction of disease genes commonly afflicts few organs or tissues, rather than being systemic.<sup>12</sup>

Sites mutated in complex disease, according to Thomas and Kejariwal, are distinguished by their very ordinariness, being neither essential over long evolutionary time periods, nor different from

healthy variation. If such data are representative of all complex diseases, it will be difficult in the future to associate single amino-acid changes with protein dysfunction and complex pathology. Fortunately, it seems that it may be much easier to construct a photofit picture of what an average complex disease gene might look like: it would encode a transmembrane or a secreted protein, and would be expressed in few tissues. Perhaps, like its Mendelian disease counterparts, it would also suffer more germline mutations than other genes.<sup>13,14</sup> Investigators hunting the culprits of complex disease might be satisfied even with these few clues when narrowing down their lists of suspect genes ■

CP Ponting and L Goodstadt, MRC  
Functional Genetics Unit, University of  
Oxford, Department of Human Anatomy  
and Genetics, South Parks Road, Oxford  
OX1 3QX, UK.  
E-mail: [chris.ponting@anat.ox.ac.uk](mailto:chris.ponting@anat.ox.ac.uk)

## References

- 1 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
- 2 Chasman D, Adams RM: Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 2001; **307**: 683–706.
- 3 Sunyaev S, Ramensky V, Koch I, Lathe III W, Kondrashov AS, Bork P: Prediction of deleterious human alleles. *Hum Mol Genet* 2001; **10**: 591–597.
- 4 Thomas PD, Kejariwal A: Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci USA* 2004; **101**: 15398–15403.
- 5 Cargill M, Altshuler D, Ireland J *et al*: Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; **22**: 231–238.
- 6 Ng PC, Henikoff S: Predicting deleterious amino acid substitutions. *Genome Res* 2001; **11**: 863–874.
- 7 King MC, Wilson AC: Evolution at two levels in humans and chimpanzees. *Science* 1975; **188**: 107–116.
- 8 Prokunina L, Castillejo Lopez C, Oberg F *et al*: A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* 2002; **32**: 666–669.
- 9 Hurst LD: The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet TIG* 2002; **18**: 486.
- 10 International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
- 11 Mouse Genome Sequencing Consortium: Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; **420**: 520–562.
- 12 Winter EE, Goodstadt L, Ponting CP: Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res* 2004; **14**: 54–61.
- 13 Smith NG, Eyre-Walker A: Human disease genes: patterns and predictions. *Gene* 2003; **318**: 169–175.
- 14 Huang H, Winter EE, Wang H *et al*: Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol* 2004; **5**: R47.