Beta-defensins and Crohn's disease: confusion from counting copies.

Edward J. Hollox, PhD

Department of Genetics, University of Leicester, UK

Correspondence to:

Dr E.J.Hollox

Department of Genetics

Adrian Building

University of Leicester

University Road

Leicester LE1 7RH

UK

Crohn's disease (CD) is a debilitating chronic inflammatory disease, which most commonly affects the terminal ileum and colon, but can affect any part of the gastrointestinal tract. In the West, the disease is frequent, with a prevalence of around 150 per 100,000 (ref 1), and patients have an increased risk of other inflammatory/autoimmune diseases such as psoriasis, ankylosing spondylitis and multiple sclerosis. It is a disease where both environmental and genetic variation make an approximately equal contribution to disease risk; and the current consensus suggests the disease is a result of an inappropriate immune response against gut microbes in a genetically susceptible host. The role in CD of a family of antimicrobial proteins and inflammatory mediators called beta-defensins is of considerable interest, yet there is much controversy on how levels of these proteins, and variation of the genes that encode them, affects susceptibility to CD. In this issue, Bentley et al. (2) present data on the genomic copy number of beta-defensin genes that not only fails to support previous work (3), but completely contradicts it. What is going on?

In recent years, there has been considerable success in dissecting the individual environmental and genetic components to CD. Smoking has been shown to markedly affect the clinical course of CD (4), and has been the most significant environmental component yet identified. Genomewide scans have been particularly successful in identifying genetic loci involved, initially by using family data and linkage analysis, followed more recently by using large case-control cohorts and association analyses. These have both confirmed susceptibility loci and identified new loci, with NOD2/CARD15 confirmed as the locus of strongest effect, with an odds ratio of 4 (refs 5, 6). This locus alone contributes over 1% of the variance in total disease risk, but all 32 loci identified to date explain only 10% of the total disease risk variance, suggesting that the variation at these loci each contribute very subtly to the etiology

of the disease. This leaves four/fifths of the genetic variation in risk unexplained, and there have been several suggestions on where this variation is hidden, including rare sequence polymorphisms and complex copy number polymorphism (where the number of copies of a gene vary between people). Case-control sample sizes of tens of thousands combined with new technologies are needed to determine whether the hidden genetic risk really is in these classes of polymorphism.

With this background, any clues from models of the etiology of the CD on where to look for genetic variation in risk are very welcome. If, as commonly accepted, CD is the result of a loss of tolerance to gut flora and a subsequent inappropriate immune response leading to tissue damage, then factors involved in mediating tolerance to gut flora may have critical roles in controlling the progression to CD. One of these factors may be the beta-defensins, which are a family of proteins sharing a conserved six-cysteine motif that forms three disulphide bridges generating a characteristic tertiary protein structure. They have multiple functions, including direct antimicrobial activity and cytokine activity in inflammatory signalling pathways. Four human beta-defensins are expressed in the gut: human beta-defensin 1 (HBD-1) is expressed constitutively and human beta-defensins 2, 3 and 4 (HBD-2, -3, -4) are induced by bacterial and inflammatory stimuli, with HBD-2 induction mediated by the NOD2/CARD15 protein (7,8). In CD, there is no induction of HBD-2 despite a severe inflammatory response. This is in contrast to the inflammatory bowel disease ulcerative colitis, where high levels of HBD-2 are characteristic of the inflammation sites; and also of the inflammatory skin disease psoriasis, where expression of HBD-2 is a characteristic event and expression level associated with clinical severity (9). Perhaps this low level of beta-defensin induction is the key characteristic of CD, and restoring beta-defensin levels would ameliorate the disease?

Analysis of the beta-defensin genes revealed that they show copy number polymorphism (CNP); that is, different individuals have different numbers of the same gene in their genomes. The genes encoding the four beta-defensins expressed in the gut (genes *DEFB1*, *DEFB4*, *DEFB103* and *DEFB104* encoding proteins HBD-1, -2, -3 and -4 respectively) are clustered at chromosomal region 8p23.1, and *DEFB4*, *DEFB103* and *DEFB104* (but not *DEFB1*) show copy number polymorphism and vary together as a group. For most genes we would expect two copies per diploid genome, but for this cluster different people commonly have between two and seven copies, with four copies being the most frequent (39% in England). The high and low extreme copy numbers are also at appreciable frequencies in the population (4% for two copies, 1.5% for 7 copies, in England) (10, 11). This prompts the question of whether the low gene copy number of beta-defensins is linked to the absence of beta-defensin induction in CD, particularly *DEFB4* (HBD-2), and therefore whether low gene copy number of *DEFB4* itself predisposes an individual to CD.

Fellerman et al (3) tackled this question by analysing two CD case-control cohorts from Germany and the United States. They found that CD patients had slightly lower mean *DEFB4* copy number than unaffected controls from the same population, suggesting that low *DEFB4* copy number was a susceptibility factor for CD. In addition, there was a correlation between *DEFB4* copy number and expression levels of DEFB4 mRNA in mucosal biopsies from patients, supporting a direct causative link between low gene copy number, low expression and development of CD. The genetic association of low copy number and disease was significant only in those with colonic CD rather than ileal CD, and the authors suggested that this reflects a difference in the levels of HBD-2 between ileum and colon. However, in this issue of the AJG, Bentley and colleagues present data showing a completely opposite effect

of *DEFB4* gene copy number, suggesting that *high* copy number is associated with both ileal and colonic CD (4). Another report published recently adds more data into this maelstrom of mixed messages: by measuring HBD-2 protein levels in colonic biopsies in ex vivo culture, they find *no* association of *DEFB4* copy number with expression level in patients or controls (12).

What are we to make of these contradicting results? Findings from early genetic association studies on single nucleotide polymorphisms (SNPs) were often not reproduced in subsequent studies. An example is in Alzheimer's disease, where an association study examining the threonine-alanine polymorphism at amino acid position 13 of the *SERPINA3* gene (encoding alpha-1-antichymotrypsin) produced an *increased* risk associated with the threonine allele (13). A subsequent study, also using samples from the Japanese population, showed a significantly *decreased* risk due to that same threonine allele (14). A meta-analysis, where these two studies and many others were analysed together, produced strong evidence that this SNP had no effect at all on susceptibility to Alzheimer's disease (15). In this and other cases, these conflicting results are usually due to small sample size limiting the power to detect small effects and enhancing the possibility of detecting false-positive results. When one genetic locus is analysed, samples sizes of thousands needed for genomewide association studies (GWAS) are not necessarily required. This is because GWAS tests many thousands of loci for association with the disease of interest, and because of the power lost due to multiple testing, very large sample sizes are required to detect weaker effects. Nevertheless, sample sizes of several hundred, at least, are needed to minimise false-positive rates, and independent replication in another cohort of the same, or larger, sample size allows a high degree of confidence in the result. Bentley has a respectable sample size (466 patients and 329 controls) but no replication (2), and Fellermann has replication, but small sample sizes (85 patients, 20

controls and 149 patients, 169 controls, ref 3). With this cohort size, any observations are at best only suggestive, at worst merely anecdotal.

Accurately typing gene copy number is difficult using current technologies, certainly much more difficult than the robust and reproducible single nucleotide polymorphism (SNP) typing that is now commonplace. This is for two reasons, firstly the ubiquity and importance of SNPs has been realised for some time, allowing sufficient time for new technologies to be developed. Secondly, the inherent physical nature of SNPs allows more discrimination: there is a qualitative chemical difference between a cytosine and an adenine at a particular nucleotide position, whilst for CNPs there is only a qualitative difference - less or more of the same DNA sequence. Quantitative real-time PCR assays, such as the assay used by both Bentley and Fellermann, have been used extensively to robustly detect differences in mRNA expression levels, levels that typically differ by several orders of magnitude. Yet for gene copy number analysis using genomic DNA, these assays are required to distinguish four from five copies, a relative increase of 0.25, which is a considerable challenge. Bentley and colleagues produce lots of data using this approach, and tackle the issue of experimental error rate, but even in their hands repeating the assay five times still produces a 5% error rate (2).

These two aspects of copy number association studies can confound a third source of error which may lead to false positive associations. In an association study, we interpret a difference in allele frequency (for a SNP) or copy number distribution (for a CNP) between cases and controls as evidence of an influence of that particular polymorphism on disease. But there may be other reasons for that difference: population stratification or a difference in the physico-chemical properties of the DNA between cases and controls. An example of the latter was given in a case-control SNP association study where case and controls produced

biased genotype data, despite their being tested blind to disease status, and randomly assorting cases and controls among experiments (16).

The contribution of Bentley et al. is valuable because it shows that the case for low *DEFB4* gene copy number association with CD is doubtful. But we should be wary of concluding that the opposite is true; as the authors state: "A consensus view of the relationship of beta-defensin genomic copy number with CD will only be obtained through further association and expression studies in large phenotypically well-defined cohorts." (2) The cohorts should be as large as possible, and the methods used as reliable as possible in order to get the right answer, and to prevent time and money being spent on research with shaky foundations. Until the publication of the paper in this issue of AJG, the consensus was that the jury was out on the case of association of low copy number *DEFB4* with CD. Now, the jury is back in the courtroom, and the verdict is "not proven".

Word count excluding references 1763

1. Stone MA, Mayberry JF, Baker R. (2003) Prevalence and management of inflammatory bowel disease: a cross-sectional study from central England. Eur J Gastroenterol Hepatol. 15:1275-1280

2. Bentley RW, Pearson J, Gearry RB, Barclay ML, McKinney C, Merriman TR, Roberts RL (2009) Association of higher *DEFB4* genomic copy number with Crohn's disease. Am J Gastroenterol, *this issue*

3. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B, Stange EF (2006) A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet 79:439-448

4. Thomas GA, Rhodes J, Green JT, Richardson C (2000) Role of smoking in inflammatory bowel disease: implications for therapy. Postgrad Med J 76:273-279

5. Barrett JC, Hansoul S, Nicolae DL and 61 co-authors (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 40:955-962

6. Van Limbergen J, Wilson DC, Satsangi J. (2009) The genetics of Crohn's disease. Annu Rev Genomics Hum Genet 10:89-116

7. Voss E, Wehkamp J, Wehkamp K, Stange EF, Schröder JM, Harder J. (2006) NOD2/CARD15 mediates induction of the antimicrobial peptide human beta-defensin-2. J Biol Chem 281:2005-2011

8. Ramasundara M, Leach ST, Lemberg DA, Day AS. (2009) Defensins and inflammation: the role of defensins in inflammatory bowel disease. J Gastroenterol Hepatol 24:202-208.

9. Jansen PA, Rodijk-Olthuis D, Hollox EJ, Kamsteeg M, Tjabringa GS, de Jongh GJ, van Vlijmen-Willems IM, Bergboer JG, van Rossum MM, de Jong EM, den Heijer M, Evers

AW, Bergers M, Armour JA, Zeeuwen PL, Schalkwijk J. (2009) Beta-defensin-2 protein is a serum biomarker for disease activity in psoriasis and reaches biologically relevant concentrations in lesional skin. PLoS One 4:e4725

10. Hollox EJ, Armour JA, Barber JC (2003) Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. Am J Hum Genet. 73:591-600

11. Hollox EJ, Barber JC, Brookes AJ, Armour JA (2008) Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23.1. Genome Res 18:1686-1697

12. Aldhous MC, Noble CL, Satsangi J (2009) Dysregulation of human beta-defensin-2 protein in inflammatory bowel disease. PLoS One 4:e6285

13. Muramatsu T, Matsushita S, Arai H, Sasaki H, Higuchi S (1996) Alpha 1-antichymotrypsin gene polymorphism and risk for Alzheimer's disease. J Neural Transm 103:1205-1210

14. Yoshiiwa A, Kamino K, Yamamoto H, Kobayashi T, Imagawa M, Nonomura Y, Yoneda H, Sakai T, Nishiwaki Y, Sato N, Rakugi H, Miki T, Ogihara T (1997) Alpha 1-antichymotrypsin as a risk modifier for late-onset Alzheimer's disease in Japanese apolipoprotein E epsilon 4 allele carriers. Ann Neurol 42:115-117

15. Bertram L, McQueen M, Mullin K, Blacker D, Tanzi R. The AlzGene Database. Alzheimer Research Forum. Available at: http://www.alzgene.org. Accessed 01 September 2009.

16. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 37:1243-1246