

## From association to causality: the new frontier for complex traits

Nicholas Katsanis

Address: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.  
Email: katsanis@jhmi.edu

Published: 25 February 2009

*Genome Medicine* 2009, **1**:23 (doi:10.1186/gm23)

The electronic version of this article is the complete one and can be found online at <http://genomemedicine.com/content/1/2/23>

© 2009 BioMed Central Ltd

### Abstract

---

Technological and analytical advances have led to an unprecedented catalog of genomic regions associated with a broad range of clinically relevant phenotypes in humans. However, some examples notwithstanding, the causes of the overwhelming majority of genetic diseases remain obscure. More importantly, an emerging lesson from genome-wide association studies is that, in most instances, the resolution necessary for identifying actual genes that underlie the phenotype is limited, as is our ability to develop mechanistic, testable disease models from such studies. These new realities will probably necessitate a paradigm shift in our approach to complex traits, for which the combinatorial application of genomic and functional studies will be necessary to understand the mechanism and pathology of genetic disease. Here I will discuss these issues and highlight how additional sequencing and genotyping of ever-increasing cohort sizes without functional interpretation is unlikely to improve our ability to dissect the genetic basis of complex traits.

---

The sequencing of several genomes and the HapMap project, coupled to high-density genotyping technologies and new statistical tools, have made possible the systematic interrogation of the genetic basis of complex traits in humans [1,2]. Largely consistent with the community's expectations, the past two years have seen a logarithmically expanding number of genomic regions associated with a broad range of phenotypes (see [3] for an updated list). We now know of at least 16 potential susceptibility loci for type 2 diabetes in the human genome [4], and there are similar or better numbers for other diseases, including type 1 diabetes (24+ loci) and Crohn's disease (30+ loci; see [5,6] for examples). Similar progress has been made in the identification of genomic regions that influence quantitative traits, such as height [7], metabolite abundance [8] and drug response [9].

These discoveries infuse us with optimism about the potential for understanding the key homeostatic pathways that are causally related to these disorders, which could lead to new drug development and better patient management. At the same time, these studies have presented us with some

surprising findings. Most importantly, lost in the collective euphoria of apparently rapid progress is the fact that, some notable exceptions notwithstanding, we are now faced with three major challenges. Firstly, the total sum of disease risk attributed to common variants detected by genome-wide association studies (GWASs) remains modest, raising a problem analogous to the astrophysicists' search for 'dark matter'. Secondly, the overwhelming majority of single nucleotide polymorphisms (SNPs) associated with disease lie in intra- or intergenic regions; there has been a striking dearth of coding SNPs. Finally, almost none of the association studies have led to the identification of the causal allele(s), and thus neither the 'disease-associated' gene, nor indeed the type of lesion (gain-of-function, loss-of-function, and so on), can be identified unequivocally. The sheer number of disorders subjected to GWAS suggests that these deficiencies are unlikely to be random findings, or to represent idiosyncratic genetic architecture of specific phenotypes. Rather, it is likely that these observations result from biological principles that will require a significant shift in our approach to understand them.

The second challenge, that of the modest risk conferred by common variants, is in some ways a minor problem, despite the unwelcome consequence that it potentially hinders, and might even defeat, the utility of high-density genotyping as a prognostic clinical tool. The fact that the relative risk conferred by most SNPs associated with Crohn's disease lingers in the 1.2-1.5 range (with the exception of alleles of *NOD2*, which encodes nucleotide oligomerization domain protein 2, and *IL23R*, which encodes the interleukin-23 receptor) does not detract from the establishment of the hypothesis that defective autophagy is part of the cause of this disorder [10], which has, in turn, enabled the development of mammalian models to study the disease process [11,12]. Likewise, the association of the same genomic regions with discrete disorders can highlight not only suspected relationships (rheumatoid arthritis and lupus), but also novel and surprising ones (diabetes and colon cancer; see [1] for more details).

The first and third challenges mentioned above, those of the genetic 'dark matter' and allele causality (coupled to mutational models), are more complex and are probably interlinked. Reasonable hypotheses have been put forth that suggest that the failure to detect the majority of the genetic load in complex disease is a reflection of insufficient statistical power, as evidenced by the accelerated locus discovery by meta-analyses. At the same time, evolutionary arguments have been proposed to explain the preferential enrichment of non-coding variants in complex disease. Most prominently, it has been suggested that strong (often coding) mutations that have a significant impact on fitness would be more likely to be associated with Mendelian disorders, whereas milder alleles that affect spatiotemporal patterns of gene expression are more likely candidates for exerting modest effects in complex disease [13].

It is reasonable to predict that understanding of the total genetic load in complex disease will be enhanced by higher density genotyping, increased sample size and expanded ethnic diversity, and medical resequencing of risk regions. Importantly, however, none of these tools will unequivocally enable the transition from association to causality.

The sobering reality remains that, despite the hundreds of loci identified, the number of *bona fide* genes associated with complex disease identified through GWASs is modest. Regrettably, a guilty-by-association view has emerged and the definitions of 'locus' and 'gene' are becoming dangerously interchangeable. At best, most arguments for the role of specific genes on conferring susceptibility put forth to date are correlative; although many will turn out to be correct, this is unlikely to be the ubiquitous truth. Recent work in age-related macular degeneration (AMD) highlights the problem. Case-control association studies pointed to a region on 10q26 that conferred strong susceptibility to AMD, and further work suggested that variation in the promoter region of *HTRA1*, which encodes a multi-functional serine protease, might explain the effect [14]. However, additional

genotyping [15] and, independently, complete resequencing of the risk haplotype, pointed to a deletion in the 3' untranslated region (UTR) of predicted transcript *LOC387715* that destabilizes its message, suggesting that loss of that transcript, and not *HTRA1*, drives the susceptibility to AMD [16]. Unfortunately, the presence of a haplotype that is not associated with AMD that carries a premature termination codon in *LOC387715* confounds that hypothesis as well [17].

Similar questions should be raised about the direct involvement in obesity of the *FTO* gene, given that the associated SNP lies in intron 1 of that gene's transcript [18]. Even though the *FTO* protein is now being investigated mechanistically to understand energy regulation [19], there is no actual evidence that *FTO* is the gene that underlies the association. Notably, *FTO*, and the neighboring gene *FTM* (or *RPGRIP1L*), which lies less than 1 kilobase away in a head-to-tail orientation, have been shown to be co-regulated by the transcription factor *CUTL1*, and at least one of the associated SNPs maps to a predicted *CUTL1* binding site [20], raising the possibility that dysfunction at *FTM*, not *FTO*, might be the driver of the phenotype. Intriguingly, *FTM* is mutated in several ciliopathies, which are hallmarked, among other features, by hyperphagia-driven obesity [21].

There is no doubt that additional, denser genotyping, admixture mapping and deep resequencing will refine current loci and uncover new ones. However, it is important to consider the relative value of these data without physiologically relevant functional interpretation. In some instances, population- and genetic-based arguments can suggest that certain alleles cause certain phenotypes, as has been shown by medical resequencing of candidate genes in patients at the extremes of high-density lipoprotein (HDL) and low-density lipoprotein (LDL) plasma levels [22,23]. Nonetheless, even in these examples, *in vitro* assays that tested the functional consequences of the mutations were necessary for the researchers to present compelling arguments.

Given these observations, it might be prudent to refocus our efforts on how to evaluate the physiological impact of genetic and genomic variation on gene function, because without that ability, the contribution of additional GWASs on understanding disease mechanisms will remain limited. Given the plethora of *in vitro* and *in vivo* tools available to the community, it should be possible to develop assays that can test directly the effect of gene variation in appropriate cell lines, cell types and animal models; without such tools, we might be left with an unexplorable catalog of associated SNPs but gain little wisdom about how to develop diagnostics and therapeutics for complex traits.

### Abbreviations

AMD, age-related macular degeneration; GWAS, genome-wide association study; SNP, single nucleotide polymorphism.

## Competing interests

The author has no competing interests to declare.

## Acknowledgements

I thank Sara Katsanis and Erica Davis for helpful discussion and editing of the manuscript. This work was supported by grants R01HD04260 from the National Institute of Child Health and Development, R01DK072301 and R01DK075972 from the National Institute of Diabetes, Digestive, and Kidney disorders and P20MH084018 from the National Institute of Mental Health.

## References

- Altshuler D, Daly MJ, Lander ES: **Genetic mapping in human disease.** *Science* 2008, **322**:881-888.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges.** *Nat Rev Genet* 2008, **9**:356-369.
- A Catalog of Published Genome-wide Association Studies** [<http://www.genome.gov/26525384>]
- Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, et al.: **Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.** *Nat Genet* 2008, **40**:638-645.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dasopoulou T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhardt AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, et al.: **Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.** *Nat Genet* 2008, **40**:955-962.
- Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, Downes K, Barrett JC, Healy BC, Mychaleckyj JC, Warram JH, Todd JA: **Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci.** *Nat Genet* 2008, **40**:1399-1401.
- Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadóttir A, Ingason A, Steinthorsdóttir V, Ólafsdóttir EJ, Ólafsdóttir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Pedersen O, Aben KK, Witjes JA, Swinkels DW, den Heijer M, Franke B, Verbeek AL, Becker DM, Yanek LR, Becker LC, et al.: **Many sequence variants affecting diversity of adult human height.** *Nat Genet* 2008, **40**:609-615.
- Hazra A, Kraft P, Selhub J, Giovannucci EL, Thomas G, Hoover RN, Chanock SJ, Hunter DJ: **Common variants of FUT2 are associated with plasma vitamin B12 levels.** *Nat Genet* 2008, **40**:1160-1162.
- Cooper GM, Johnson JA, Langae TY, Feng H, Stanaway IB, Schwarz UI, Ritchie MD, Stein CM, Roden DM, Smith JD, Veenstra DL, Rettie AE, Rieder MJ: **A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose.** *Blood* 2008, **112**:1022-1027.
- Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhardt AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR: **Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.** *Nat Genet* 2007, **39**:596-604.
- Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK, Kishi C, Kc W, Carrero JA, Hunt S, Stone CD, Brunt EM, Xavier RJ, Sleckman BP, Li E, Mizushima N, Stappenbeck TS, Virgin HW 4th: **A key role for autophagy and the autophagy gene Atg1611 in mouse and human intestinal Paneth cells.** *Nature* 2008, **456**:259-263.
- Saitoh T, Fujita N, Jang MH, Uematsu S, Yang B-G, Satoh T, Omori H, Noda T, Yamamoto N, Komatsu M, Tanaka K, Kawai T, Tsujimura T, Takeuchi O, Yoshimori T, Akira S: **Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1[bgr] production.** *Nature* 2008, **456**:264-268.
- Reich DE, Lander ES: **On the allelic spectrum of human disease.** *Trends Genet* 2001, **17**:502-510.
- Marx J: **Gene offers insight into macular degeneration.** *Science* 2006, **314**:405.
- Kanda A, Chen W, Othman M, Branham KEH, Brooks M, Khanna R, He S, Lyons R, Abecasis GR, Swaroop A: **A variant of mitochondrial protein LOC387715/ARMS2, not HTRA1, is strongly associated with age-related macular degeneration.** *Proc Natl Acad Sci USA* 2007, **104**:16227-16232.
- Fritsche LG, Loenhardt T, Janssen A, Fisher SA, Rivera A, Keilhauer CN, Weber BHF: **Age-related macular degeneration is associated with an unstable ARMS2 (LOC387715) mRNA.** *Nat Genet* 2008, **40**:892-896.
- Allikmets R, Dean M: **Bringing age-related macular degeneration into focus.** *Nat Genet* 2008, **40**:820-821.
- Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harrries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, et al.: **A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.** *Science* 2007, **316**:889-894.
- Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, Yeo GS, McDonough MA, Cunliffe S, McNeill LA, Galvanovskis J, Rorsman P, Robins P, Prieur X, Coll AP, Ma M, Jovanovic Z, Farooqi IS, Sedgwick B, Barroso I, Lindahl T, Ponting CP, Ashcroft FM, O'Rahilly S, Schofield CJ: **The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase.** *Science* 2007, **318**:1469-1472.
- Stratigopoulos G, Padilla SL, LeDuc CA, Watson E, Hattersley AT, McCarthy MI, Zeltser LM, Chung WK, Leibel RL: **Regulation of Fto/Ftm gene expression in mice and humans.** *Am J Physiol Regul Integr Comp Physiol* 2008, **294**:R1185-R1196.
- Badano JL, Mitsuma N, Beales PL, Katsanis N: **The ciliopathies: an emerging class of human genetic disorders.** *Annu Rev Genomics Hum Genet* 2006, **7**:125-148.
- Cohen JC, Kiss RS, Pertsemilidis A, Marcel YL, McPherson R, Hobbs HH: **Multiple rare alleles contribute to low plasma levels of HDL cholesterol.** *Science* 2004, **305**:869-872.
- Kotowski IK, Pertsemilidis A, Luke A, Cooper RS, Vega GL, Cohen JC, Hobbs HH: **A spectrum of PCAK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol.** *Am J Hum Genet* 2006, **78**:410-422.