



Headlong into a genomic singularity

Robert W. Williams*

Center for Integrative and Translational Genomics, Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN, USA

*Correspondence: rwilliam@nb.utmcm.edu

A commentary on

Expression genetics and the phenotype revolution.

by Williams, R. W. (2006). *Mamm. Genome* 17, 496–502.

Many biologists have been struck by how preposterous it is to apply the term “post-genomics” to any aspect of our field. Estimates of gene number have finally settled to a comfortably low asymptote, but the dam has broken on the RNA front, revealing an impressive level of ignorance and mystery. In an era in which fewer than 100 humans have been sequenced and the vast majority of species are represented by zero genomes, restraint or even humility is appropriate. Or maybe not. We are now venturing into new genomic territory made possible by ultra-high throughput sequence.

In a recent paper, Adams, and colleagues (Sudbery et al., 2009) tantalize us with what post-genomics might mean – when access to massive sequencing is taken for granted. The implications in neurogenomics are enormous, and investigators would be well advised to track this new technology and determine how best to exploit the new resources in their own research programs. This commentary briefly reviews the Sudbery paper and a recent paper by Blakely and colleagues (Carneiro et al., 2009). The intent is to highlight likely repercussions of sequencing on the analysis of complex traits, including behavior.

In a cutting-edge approach, the Wellcome Trust Sanger team flow-sorted one of the most interesting mouse chromosomes – chromosome 17, home of the major histocompatibility complex and many quantitative trait loci (QTLs) – from the genomes of two exceedingly different strains of mice, A/J and CAST/Ei. They sequenced these divergent versions of Chr 17 using a massively parallel sequencing system (Solexa) and applied a battery of genome assembly methods to stitch together 100 bp reads. In total they achieved coverage of between

22 and 34× – enough to drench all but the most repetitive 1.5% of this chromosome in high quality sequence. Accomplishing this task took a few weeks of machine time, but was preceded by months of planning and sample preparation and followed by several months of intense assembly and analysis.

With sequences in hand, the group applied a sophisticated chain of algorithms (and independent validation) to systematically harvest large numbers of known and novel SNPs, indels, CNVs, and structural rearrangements. The gain in number of SNPs between A/J and the genome of the reference strain C57BL/6J was a modest 30%. In contrast, the gain between CAST/Ei and C57BL/6J was an impressive 35-fold addition, with a harvest of 630,000 new SNPs.

Readers of the paper by Sudbery et al. (2009) may fairly ask why did they sequence just one chromosome? The answer is that this is a large pilot project leading to a truly massive sequencing program. David Adams, the senior author, and colleagues are assembling genomes for 17 strains of mice – by far the largest effort of this kind in mouse genomics. Other groups, including our own, are taking a focused approach, deeply sequencing key research strains (in our case DBA/2J, the other parent of the BXD family).

Now comes the fun part. Any practitioner of the dark art of QTL mapping knows that identifying causal sequence variants is the rate-limiting step that distinguishes QTL analysis from functional genomics. Various tricks – some clever and some just dirty – have been devised by the bright and the addled to make this process more efficient. No doubt, the best friend of the hapless QTL cloner is excellent sequence data of parental strains and progeny. Illustrating this point, in the paper Sudbery et al. (2009) we are treated to comprehensive sequence-based dissection of a QTL that causes variation in liver triglyceride levels in a cross between A/J and C57BL/6J. By systematically working through all variants in a well-delimited

QTL on Chr 17, the team efficiently cut the list of high-caliber candidates down to three, of which one, *Lmf1*, is a gem previously linked to hypertriglyceridemia.

This result illustrates one of the most intriguing implications of this work: complex trait analysis is transitioning from being a strictly forward genetic approach (from phenotype to gene) and is now adopting reverse genetic methods similar to those used to study knockouts. As sequencing becomes cheaper, this transition will accelerate. Eventually, researchers interested in complex phenotypes – in other words, almost all of us – will use forward and reverse genetic methods with equal facility. In contrast to other methods, reverse complex trait analysis will often start with natural sequence variants rather than with engineered alleles or mutagenized stock.

Reverse complex trait analysis will still require large numbers of progeny (sets of recombinant inbred strains will be ideal), but the analysis will now be far more focused, as presaged in a recent paper by Carneiro et al. (2009) who hammered away at the functions of a crucial serotonin transporter variant (5-HTT, *Slc6a4*) using as a resource a panel of 80 BXD strains made by crossing C57BL/6J and DBA/2J. Roughly half of the progeny strains have inherited the *B* allele, the other half the *D* allele. The result of combining sequence data with the BXD genetic reference panel is a robust and high-powered *t* test of the function of a known sequence variant. This is reverse genetics, functional genetics, and complex trait analysis rolled into one.

Many of us are already contemplating how to redesign and reinterpret genetic and functional experiments when the next wave of sequence data strikes, bringing a compendium of ~50 million murine sequence variants for our reading, reviewing, and experimenting pleasure (Roberts et al., 2007). The analysis of a backlog of QTLs could be most rewarding, the analysis of thousands of previously hidden major alleles even more so.

One last thought brings us back to the title of this commentary. The accelerating pace of progress in sequencing forces us to acknowledge the sharp transition that we are now entering – in essence, a genomics singularity that matches that in computer science, in which each new generation of technology has exponentially greater throughput and can almost instantly replicate all of what has gone before. In this environment, today's cutting-edge paper is practically obsolete before publication. What will experimental genetics be like when genomes are almost free? These genomes will be a gift to both complex trait analysis and mutagenesis. In the end, there will just be a continuum between major and minor sequence variants, and the technologies used to find them will be of minimal importance. In contrast, our success in

understanding relations between sequence difference and phenotypes will depend critically on the precision and depth of phenotyping (Williams, 2006) and on the sophistication with which we can model the intricacies and contingencies of complex biological systems. Post-genomics may one day be relabeled *systems genomics*.

REFERENCES

- Carneiro, A. M., Airey, D. C., Thompson, B., Zhu, C. B., Lu, L., Chesler, E. J., Erikson, K. M., and Blakely, R. D. (2009). Functional coding variation in recombinant inbred mouse lines reveals multiple serotonin transporter-associated phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2047–2052.
- Roberts, A., Pardo-Manuel de Villena, F., Wang, W., McMillan, L., and Threadgill, D. W. (2007). The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm. Genome* 18, 473–481.
- Sudbery, I., Stalker, J., Simpson, J. T., Keane, T., Rust, A. G., Hurles, M. E., Walter, K., Lynch, D., Teboul, L., Brown, S. D., Li, H., Ning, Z., Nadeau, J. H., Croniger, C. M., Durbin, R., and Adams, D. J. (2009). Deep short-read sequencing of chromosome 17 from the mouse strains A/J and CAST/Ei identifies significant germline variation and candidate genes that regulate liver triglyceride levels. *Genome Biol.* 10, R112.
- Williams, R. W. (2006). Expression genetics and the phenotype revolution. *Mamm. Genome* 17, 496–502.

Received: 01 November 2009; published online: 03 February 2010.

Citation: Williams RW (2010) Headlong into a genomic singularity. *Front. Neurosci.* 4:25. doi: 10.3389/neuro.15.001.2010

This article was submitted to *Frontiers in Neurogenomics*, a specialty of *Frontiers in Neuroscience*.

Copyright © 2010 Williams. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.