

Shana E. McCormack<sup>1,2</sup> and Struan F.A. Grant<sup>1,2,3</sup>

## Allelic Expression Imbalance: Tipping the Scales to Elucidate the Function of Type 2 Diabetes–Associated Loci



*Diabetes* 2015;64:1102–1104 | DOI: 10.2337/db14-1836

Genome-wide association studies (GWAS) have identified many genetic locations harboring variation that increases susceptibility to type 2 diabetes (T2D) (1). However, in order to leverage these exciting findings into rational personalized treatment strategies for patients, one needs to understand these loci in much greater detail. To begin with, it is far from clear how mechanistically these genetic differences drive T2D risk; indeed, GWAS typically report variation that is in itself not causal but rather closely “travels” down the generations with the culprit variant. Furthermore, it has proven challenging to elucidate the actual causal gene at each location. Studies of obesity genetics highlight this point. For some time, attention has been focused on understanding *FTO*, as intronic variation within this gene was implicated in obesity through consistent GWAS (2,3). However, it was recently reported that these variants actually act at a distance to influence the expression of the neighboring gene, *IRX3* (4). There is much interest, therefore, in experimental strategies that can elucidate the functional significance of T2D GWAS variants while avoiding misattribution of biological risk.

In this issue of *Diabetes*, Locke et al. (5) applied a logical molecular biology approach to tackle this issue. They sought to discover the regional effects of previously identified T2D risk loci resulting from multiple GWAS efforts, the largest and most recent being from the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (6). Specifically, they investigated a particular mechanism by which nucleotide changes could impact T2D risk, namely by changing the transcription of genes in the proximity of a given signal. Their approach, “targeted allelic expression profiling,” aimed to identify imbalances in gene expression related to T2D risk–associated alleles. The presence of possible expression differences was thus hypothesized to tip the scales in favor of

a transcriptional explanation for at least some of the GWAS results.

The authors’ strategy is illustrated in Fig. 1. Many genetic variants associated with increased T2D risk are single nucleotide polymorphisms (SNPs) that lie in regions of genes (introns) that are never transcribed into mature messenger (m)RNA. As a result, the effect of these intronic SNPs on gene expression can be difficult to assess directly. For each “lead” intronic SNP (i.e., those variants that capture the association most optimally) identified in major GWAS reports of T2D, the investigators searched for “proxy” exonic SNPs (i.e., variants inherited together with the lead SNPs but located in an exon instead of an intron and thus much more amenable to expression analyses). For example, as shown in Fig. 1, lead SNP rs2007084 is located in the intron of the gene *ANPEP* but is in linkage disequilibrium (i.e., inherited together) with proxy SNP rs17240240, located in one of the exons of *ANPEP*. The quantity of mature mRNA carrying the C allele (acting as a proxy for the risk allele of the lead SNP) can then be measured and compared with the amount carrying the T allele (acting as a proxy for the nonrisk-conferring allele at the lead SNP). In this way, the transcriptional effects attributable to the risk allele can be isolated using transcription yielded from the other allele as a within-experiment control.

A suitable proxy exonic SNP partner could not be found for every lead SNP. Indeed, of the 65 loci identified in the original GWAS, ultimately only 18 unique exonic SNPs could be leveraged. Samples of islet tissue from 36 deceased, white donors without diabetes were used for the gene expression studies. For the allelic expression profiling to be feasible for a given lead SNP, donors needed to be heterozygous for that SNP (i.e., have a copy of each allele, as illustrated in Fig. 1).

<sup>1</sup>Division of Endocrinology and Diabetes, Children’s Hospital of Philadelphia, Philadelphia, PA

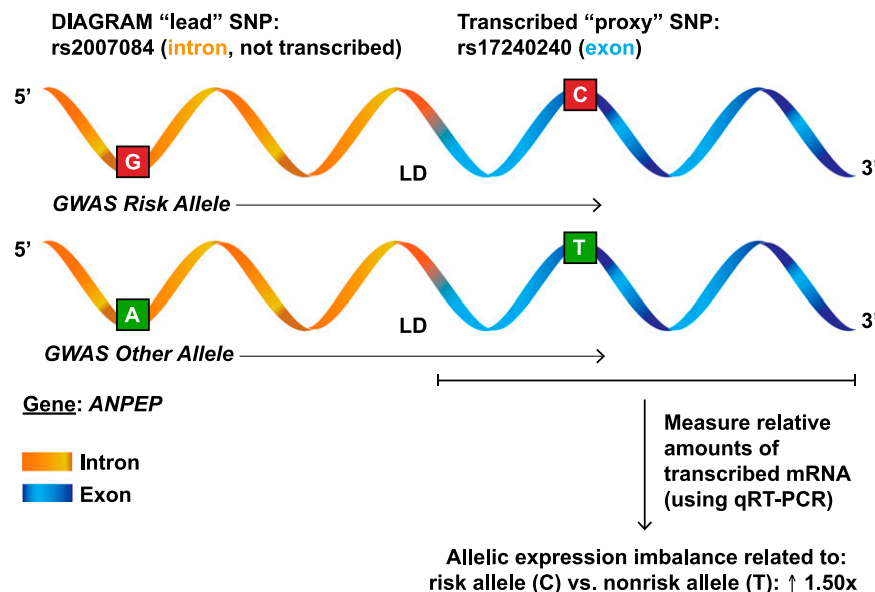
<sup>2</sup>Department of Pediatrics and Institute for Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

<sup>3</sup>Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, PA

Corresponding author: Struan F.A. Grant, grants@chop.edu.

© 2015 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered.

See accompanying article, p. 1484.



**Figure 1**—Allelic expression profiling to investigate the role of lead intronic SNPs influencing expression of nearby exons. The two strands of DNA of one of the genes studied by the investigators, *ANPEP*, are shown to illustrate this approach. First, investigators chose a sample heterozygous for the lead SNP of interest, here rs2007084, located in an intron (yellow area). The risk allele is shown in red, the other allele in green. Next, they identified a transcribed proxy SNP, located in an exon (blue area), inherited together (i.e., in linkage disequilibrium [LD]) with the lead SNP, here rs17240240, as shown by the arrows. The proxy SNP has a C nucleotide on the same DNA strand as the lead SNP risk allele, and a T nucleotide on the same DNA strand as the lead SNP other allele. In this way, the transcribed mRNA is tagged as originating from the DNA strand with or without the lead SNP risk allele. The relative amounts of transcribed mRNA can then be measured and compared using quantitative RT-PCR (qRT-PCR).

For five of the genes with available data, differential gene expression related to genotype at the proxy exonic SNP was identified and confirmed using other linked exonic SNPs. This short list includes genes with well-characterized function in islets. For example, *KCNJ11* encodes an ATP-sensitive  $K^+$  channel that couples glucose-stimulated energy production to insulin secretion in the  $\beta$ -cell; mutations in *KCNJ11* have been associated with neonatal diabetes (7). With others, there is a clear association with diabetes, and gene function is beginning to be better understood. For example, *WFS1* is mutated in Wolfram syndrome, a complex multisystem disorder that includes diabetes precipitated by nonimmune-mediated pancreatic  $\beta$ -cell death. Mutant *WFS1* may cause  $\beta$ -cell endoplasmic reticulum stress (8). In contrast, *ANPEP* (9), whose status as the causal gene was supported by additional expression quantitative trait loci experiments, is a transmembrane metalloprotease with a posited role in angiogenesis (10) whose involvement in diabetes pathogenesis remains to be explored.

The choice to use pancreatic islet tissue for these proof-of-principle experiments is a logical one, as pancreatic  $\beta$ -cell failure is a clinical hallmark of T2D. In addition, many T2D risk variants appear to exert their effects by altering insulin processing and secretion (11). However, many of the neighboring genes are also widely expressed outside the pancreas, and evidence of potentially significant regulatory variation at important T2D risk loci (e.g., *TCF7L2*) in nonpancreatic

tissues is accumulating (12–14). Studying these other tissues may yield a more complete picture. Indeed, Locke et al. (5) acknowledge that their experiments do not elucidate whether or how these differences in gene expression influence T2D risk. They point out that there is a precedent for even apparently small changes in expression affecting biology. For example, haploinsufficiency (i.e., carrying one mutated copy) of *SLC30A8*, a gene that encodes an islet zinc transporter, appears sufficient to substantially reduce risk for T2D (15). A risk allele in the 3' untranslated region of *SLC30A8* also produced allelic expression imbalance in their study.

Despite not being able to assess every locus due to a lack of an available exonic proxy and the limitation of a single tissue, these experiments demonstrate one promising strategy for identifying how GWAS loci tip the scales of gene expression. Allelic expression profiling therefore may be one incremental step in translating findings from GWAS into a better understanding of T2D pathogenesis.

**Acknowledgments.** The authors would like to thank Angela Knott (Children's Hospital of Philadelphia) for assistance in generating Fig. 1.

**Funding.** The authors received funding from National Institutes of Health grant K12 DK094723-01 (S.E.M.), and from the Daniel B. Burke Endowed Chair for Diabetes Research (S.F.A.G.).

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

## References

1. Mahajan A, Go MJ, Zhang W, et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;46:234–244
2. Frayling TM, Timpson NJ, Weedon MN, 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
3. Bradfield JP, Taal HR, Timpson NJ, et al.; Early Growth Genetics Consortium. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 2012;44:526–531
4. Smemo S, Tena JJ, Kim KH, et al. Obesity-associated variants within FTO form long-range functional connections with IIRX3. *Nature* 2014;507:371–375
5. Locke JM, Hysenaj G, Wood AR, Weedon MN, Harries LW. Targeted allelic expression profiling in human islets identifies cis-regulatory effects for multiple variants identified by type 2 diabetes genome-wide association studies. *Diabetes* 2015;64:1484–1491
6. Morris AP, Voight BF, Teslovich TM, et al.; Wellcome Trust Case Control Consortium; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; Asian Genetic Epidemiology Network–Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012;44:981–990
7. Gloyn AL, Pearson ER, Antcliff JF, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* 2004;350:1838–1849
8. Fonseca SG, Fukuma M, Lipson KL, et al. WFS1 is a novel component of the unfolded protein response and maintains homeostasis of the endoplasmic reticulum in pancreatic beta-cells. *J Biol Chem* 2005;280:39609–39615
9. Watt VM, Willard HF. The human aminopeptidase N gene: isolation, chromosome localization, and DNA polymorphism analysis. *Hum Genet* 1990; 85:651–654
10. Rangel R, Sun Y, Guzman-Rojas L, et al. Impaired angiogenesis in aminopeptidase N-null mice. *Proc Natl Acad Sci USA* 2007;104:4588–4593
11. Dimas AS, Lagou V, Barker A, et al.; MAGIC Investigators. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 2014;63:2158–2171
12. Boj SF, van Es JH, Huch M, et al. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* 2012;151:1595–1607
13. Bailey KA, Savic D, Zielinski M, et al. Evidence of non-pancreatic beta cell-dependent roles of Tcf7l2 in the regulation of glucose metabolism in mice. *Hum Mol Genet*. 14 November 2014
14. Kaminska D, Kuulasmaa T, Venesmaa S, et al. Adipose tissue TCF7L2 splicing is regulated by weight loss and associates with glucose and fatty acid metabolism. *Diabetes* 2012;61:2807–2813
15. Flannick J, Thorleifsson G, Beer NL, et al.; Go-T2D Consortium; T2D-GENES Consortium. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 2014;46:357–363