Jonathan M. Locke, Gerald Hysenaj, Andrew R. Wood, Michael N. Weedon, and Lorna W. Harries

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 | DOI: 10.2337/db14-0957

Genome-wide association studies (GWAS) have identified variation at >65 genomic loci associated with susceptibility to type 2 diabetes, but little progress has been made in elucidating the molecular mechanisms behind most of these associations. Using samples heterozygous for transcribed single nucleotide polymorphisms (SNPs), allelic expression profiling is a powerful technique for identifying *cis*-regulatory variants controlling gene expression. In this study, exonic SNPs, suitable for measuring mature mRNA levels and in high linkage disequilibrium with 65 lead type 2 diabetes GWAS SNPs, were identified and allelic expression determined by real-time PCR using RNA and DNA isolated from islets of 36 white nondiabetic donors. A significant allelic expression imbalance (AEI) was identified for 7/14 (50%) genes tested (*ANPEP*, *CAMK2B*, *HMG20A*, *KCNJ11*, *NOTCH2*, *SLC30A8*, and *WFS1*), with significant AEI confirmed for five of these genes using other linked exonic SNPs. Lastly, results of a targeted islet expression quantitative trait loci experiment support the AEI findings for *ANPEP*, further implicating *ANPEP* as the causative gene at its locus. The results of this study support the hypothesis that changes to *cis*-regulation of gene expression are involved in a large proportion of SNP associations with type 2 diabetes susceptibility.

Genome-wide association studies (GWAS) have discovered germline genetic variation associated with type 2 diabetes risk (1–4). One of the largest GWAS, involving DNA taken from individuals of European descent and conducted by the DIAGRAM (DIAbetes Genetics Replication And Meta-analysis) consortium, identified 65 loci associated with type 2 diabetes risk (1). However, for most of these loci, the precise identity of the affected gene and the molecular mechanisms underpinning the altered risk are not known.

Similar to other complex polygenic diseases, the lack of protein-coding variation at the majority of these loci has led many to believe that the disease-associated variation must result in changes in the level of gene expression. This is predicted to affect the amount or nature of protein produced, resulting in compromised cellular function and ultimately altered disease risk. It is this hypothesis that has led many to conduct expression quantitative trait loci (eQTL) studies, where the disease-associated genotype is correlated with expression of *cis*-linked transcripts. While there have been some successes when using such an approach (5,6), the presence of multiple known and unknown confounding factors that can affect gene expression means few disease-associated loci are robustly associated. Allelic expression profiling provides an alternative and direct way to measure the

Institute of Biomedical & Clinical Science, University of Exeter Medical School, Exeter, U.K.

Corresponding author: Lorna W. Harries, l.w.harries@exeter.ac.uk.

Received 20 June 2014 and accepted 31 October 2014.

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db14-0957/-/DC1.

See accompanying article, p. 1102.

effect of *cis*-regulatory variation on gene expression. By measuring transcripts from each allele of a gene (using a single nucleotide polymorphism [SNP] as a marker to differentiate between the two mRNA copies), the effect of *trans*-acting factors on gene expression is essentially removed as the output from one allele serves as a within-sample control for the other.

The known tissue specificity of gene expression regulation means that the most informative studies will measure transcript levels in the specific tissue(s) relevant to the disease. In the case of type 2 diabetes, characterization of physiological responses (e.g., stimulus-induced insulin secretion, insulin sensitivity) suggests most loci are associated with defects in pancreatic β-cell function (2,3,7). Therefore there is a real need to measure gene expression in human β-cells (or whole islets, as these have been shown to be a suitable proxy [8]). There have, however, been very few reports linking type 2 diabetes–associated variation with islet gene expression using the classical eQTL approach (9,10).

In this report, we provide one of the first pieces of evidence to support the theory that for a large proportion of loci associated with predisposition to type 2 diabetes, the variants affect the expression of known protein-coding genes. Furthermore, by pinpointing genes with allelic expression imbalances (AEIs), we highlight the genes likely to be involved in disease pathogenesis.

RESEARCH DESIGN AND METHODS

Tissue and Nucleic Acid Extraction

Snap-frozen islets were purchased from ProCell Biotech (Newport Beach, CA), where islets had been collected with ethical permission at source. Donor characteristics are presented in Supplementary Table 1. Islet purity and viability was determined by dithizone and fluorescein diacetate/propidium iodide staining, respectively. RNA*later*-ICE (Life Technologies, Carlsbad, CA) was used to transition the tissue to a state where RNA could be extracted using the miRVana microRNA isolation kit, as per manufacturer instructions, and using the total RNA extraction protocol. Small amounts of genomic DNA were coeluted upon RNA extraction. Whole-genome amplification was subsequently carried out using the REPLI-g Mini kit (Qiagen, Venlo, the Netherlands).

Genotyping and Quantitative RT-PCR

To find heterozygous samples suitable for allelic expression measurements, whole-genome amplified genomic DNAs were first genotyped. Sanger sequencing or TaqMan SNP genotyping assays were used to confirm that in all cases where allelic expression was determined, the respective lead type 2 diabetes–associated SNP was also heterozygous. For allelic expression analyses, total RNA was DNase-treated using the Turbo DNA-free kit (Life Technologies). DNase-treated RNA was subsequently reverse transcribed using the SuperScript VILO cDNA synthesis system (Life Technologies). For each SNP, duplicate wells were used to amplify cDNA and ~5 ng genomic DNA in all heterozygous samples. SDS version 2.1 software (Life Technologies) and its automatic settings were used to determine cycle threshold (Ct) values in the log phase of PCR. Following amplification, wells with a Ct value >36 were excluded from our analysis and dCt values calculated as dCt = Ct(FAM) − Ct(VIC). Subsequently, ddCt values were calculated by determining the difference between the dCt and the mean dCt of all genomic DNA samples. Genomic DNA samples should show a 1:1 allelic ratio, and thus any departure from 0 illustrates unequal amplification of alleles that must be corrected for. These ddCt values were then transformed to $2^{-\mathrm{ddCt}}$ values to take into account the logarithmic nature of PCR. Mean average allelic expression measurements, determined from two independent cDNAs reverse transcribed and amplified on different days, are presented. All genotyping and allelic expression experiments were performed with TaqMan SNP genotyping assays and TaqMan Genotyping Master Mix (both Life Technologies). Inventoried TaqMan SNP genotyping assays were used where possible (amplicon mapping 100% to exonic sequence). If not available or not meeting our design specifications, custom TaqMan SNP genotyping assays were designed and purchased where possible. For the islet eQTL analysis, inventoried TaqMan Gene Expression assays and TaqMan Fast Advanced Master Mix (both Life Technologies) were used. Relative expression of target genes was normalized to the geometric mean of five housekeeping genes (*ACTB*, *B2M*, *GUSB*, *HMBS*, *RPL11*). All TaqMan assays were run on the ABI7900HT platform (Life Technologies). Assay identifications and primer/probe sequences for TaqMan SNP genotyping and TaqMan Gene Expression assays are shown in Supplementary Table 2.

Statistical Analysis

Paired two-tailed *t* tests, comparing genomic DNA and cDNA values from the same donor, were used to determine statistical significance for allelic expression. Allelic and total gene expression values were normally distributed upon $\log_2$ transformation and thus permitted the use of parametric statistics. Pearson correlation coefficients were used to ascertain the robustness of our allelic expression measurement.

RESULTS

Bioinformatic Analysis of Type 2 Diabetes GWAS Loci to Identify Exonic SNPs for Allelic Expression Profiling in Human Islets

Large-scale association studies conducted by DIAGRAM, in individuals overwhelmingly of European descent, have reported 65 lead SNPs associated with susceptibility to type 2 diabetes (1). Figure 1 illustrates how these SNPs and closely correlated proxy SNPs were systematically selected for allelic expression analysis. In brief, 1,525 proxy SNPs ($r^2 > 0.8$, CEU, 1,000 Genomes Phase 1) were found. Of these SNPs (lead + proxies), 45/1,590 (2.8%) map to exons of 23 human RefSeq genes. For 18 of these

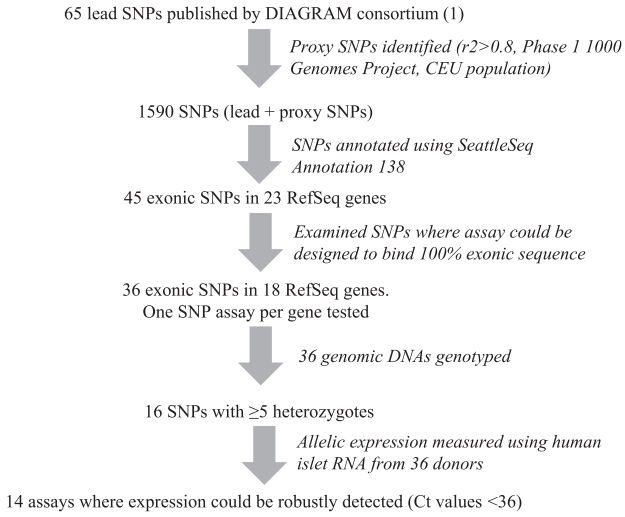

Figure 1—Systematic identification of exonic SNPs suitable for allelic expression analysis in human islets.

genes, TaqMan SNP genotyping assays could be designed to map entirely to exonic sequence, thus allowing for amplification and measurement of mature (i.e., spliced) mRNA species and normalization of allelic expression using genomic DNA from the same individual. After exclusion of SNPs with <4 heterozygotes (rs1801282, *PPARG*; rs3734621, *KIF6*) and assays where >50% cDNA samples yielded Ct values >36 (rs2793823, *ADAM30*; rs7377, *SRGN*), indicating very low levels of gene expression, allelic expression could be determined for 14 genes in samples from 36 white nondiabetic donors.

Validation of TaqMan SNP Genotyping Assays as a Robust Method for Determining Allelic Expression

The reproducibility of the allelic expression measurement was determined by comparing values generated from independent cDNAs that were reverse transcribed and amplified by real-time PCR on different days and PCR plates (r^2 = 0.72; $P = 1.8 \times 10^{-62}$) (Fig. 2*A*). To further demonstrate the robustness of this technique, allelic expression values for five genes were compared between two linked SNPs residing in the same gene (r^2 = 0.67; $P = 5.7 \times 10^{-16}$) (Fig. 2*B*).

Identification of AEI at GWAS Loci

Significant (unadjusted $P < 0.05$) AEI was observed for 7/14 genes tested (Table 1 and Fig. 3*A*, *C*, *E*, *G*, *I*, *K*, and *L*). For five of the genes where significant imbalance was observed (*ANPEP*, *KCNJ11*, *NOTCH2*, *SLC30A8*, *WFS1*), there was at least one other exonic proxy SNP that was used for validation of our findings. For all five of the tested SNPs, significant (unadjusted $P < 0.05$) AEI in a direction consistent with the results from the first SNP was observed (Table 1 and Fig. 3*B*, *D*, *F*, *H*, and *J*). We were unable to validate our results for *CAMK2B* and *HMG20A* due to the position of the only other proxy exonic SNP on the transcript (SNP either too close to a splice site to allow for assay to be designed to measure mature mRNA or near a sequence of low complexity precluding assay design).

Islet eQTL Analysis Provides Further Evidence to Support *ANPEP* as Being the Causal Gene at This Locus

One limitation of our approach is its reliance on the presence of exonic proxy SNPs, thus preventing testing of all candidate genes at a locus. Of the seven genes where we identified significant AEI, *ANPEP* exhibited by far the largest imbalance, with both SNPs tested showing ~50% difference between mRNA alleles. Therefore we predicted that we had most power to see an eQTL effect at this locus versus the other six where far lower AEIs were observed. Using real-time PCR, we measured the expression of *ANPEP* and four neighboring RefSeq genes, the two immediately 5′ of *ANPEP* (*MESP1*, *MESP2*) and the two immediately 3′ of *ANPEP* (*AP3S2*, *ARPIN*) in the 36 islet samples. Consistent with our AEI findings of increased levels of *ANPEP* mRNA originating from the risk allele, we measured 2.2-fold higher levels of *ANPEP* expression in individuals homozygous for the risk (G) allele of rs2007084 compared with heterozygous individuals (P = 0.03) (Fig. 4*A*). We found no association with rs2007084 genotype for three of the other genes at this locus (*AP3S2*, P = 0.61; *ARPIN*, P = 0.11; *MESP1*, P = 0.97) (Fig. 4*B*–*D*). Very low levels of *MESP2* transcripts (all samples either mean Ct >38 or not amplified) meant analysis of *MESP2* expression with respect to genotype was not possible. In summary, this eQTL analysis supports our AEI findings for *ANPEP* and provides further evidence that the type 2 diabetes–associated risk allele at this locus exerts its detrimental effect via increasing *ANPEP* expression.

DISCUSSION

In this study, we report that half of the type 2 diabetes SNPs examined had effects on allelic expression of nearby genes, consistent with a role for such variation in *cis*-regulation of islet gene expression. This figure is greater than genome-wide analyses of AEI, which have generally reported 4–8% of common SNPs associated with the expression of at least one transcript (11,12). An enrichment of significant genotype effects on *cis*-linked gene expression has been reported for complex disease traits (13,14), and our results suggest that effects on gene expression may play a significant role in mediating the genotype associations with type 2 diabetes.

We are aware of only one other published study reporting associations between gene expression levels and multiple type 2 diabetes–associated lead SNPs. Using global microarray data measuring gene expression in islets from 63 donors, Taneera et al. (10) reported a significant ($P < 0.001$) *cis*-eQTL for 5/47 (11%) type 2 diabetes–associated SNPs. Unfortunately, none of the *cis*-eQTLs identified in their study could be tested in our study due to a lack of SNPs meeting our inclusion criteria

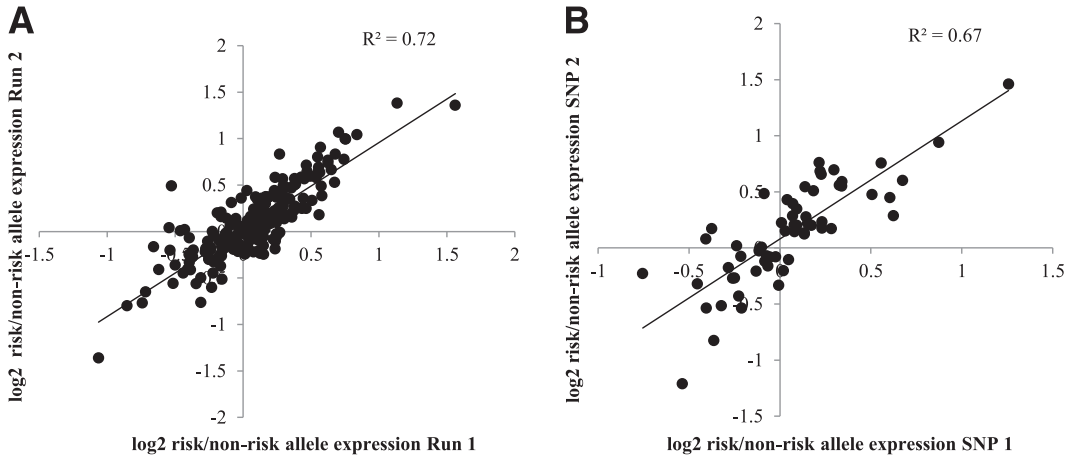

Figure 2—TaqMan SNP genotyping assays as a robust method to determine allelic expression. *A*: Correlation between allelic expression measurements determined from independent cDNAs reverse transcribed and amplified on different days (n = 221; 19 SNPs). *B*: Correlation between allelic expression measurements calculated from SNPs in high LD with each other and residing within the same gene (n = 61; 5 pairs of linked SNPs).

(exonic and assay mapping purely to exonic sequence). A comparison of the results from their study and ours illustrates some of the advantages and disadvantages of the respective techniques. While our targeted allelic expression approach identifies a higher percentage (50 vs. 11%) of genes regulated by *cis*-acting variation, our approach is limited to genes with exonic SNPs in high linkage disequilibrium (LD; $r^2 > 0.8$) with the lead type 2 diabetes–associated SNP. Indeed this dependence will also often preclude examining isoform-specific effects that may be substantial (15) and an examination of all candidate genes at a locus. The classical *cis*-eQTL approach is not dependent on transcribed exonic SNPs and thus can examine many more of the disease-associated loci and the

Table 1—Allelic expression results for 14 type 2 diabetes GWAS-implicated genes

DIAGRAM lead SNP	Transcribed SNP (gene)	Linkage (lead-transcribed SNP) r^2 (D′)	Number of samples with expression determined	Mean allelic expression (risk/nonrisk)	Unadjusted P value	Adjusted* P value
rs10923931	**rs699779 (NOTCH2)**	**0.99 (1.00)**	**10**	**0.79**	**4.8×10^{-5}**	**9.1×10^{-4}**
rs5215	**rs5219 (KCNJ11)**	**1.00 (1.00)**	**16**	**1.43**	**1.3×10^{-4}**	**1.2×10^{-3}**
rs2007084	**rs17240240 (ANPEP)**	**0.86 (0.95)**	**10**	**1.50**	**7.4×10^{-4}**	**4.7×10^{-3}**
rs7177055	**rs952472 (HMG20A)**	**0.91 (0.99)**	**13**	**1.14**	**2.1×10^{-3}**	**6.7×10^{-3}**
rs10278336	**rs1065359 (CAMK2B)**	**0.99 (1.00)**	**13**	**0.82**	**2.6×10^{-3}**	**7.1×10^{-3}**
rs3802177	**rs11558471 (SLC30A8)**	**0.97 (1.00)**	**14**	**1.12**	**6.2×10^{-3}**	**1.5×10^{-2}**
rs4458523	**rs1046320 (WFS1)**	**0.95 (0.99)**	**16**	**0.92**	**6.9×10^{-3}**	**1.5×10^{-2}**
rs12427353	rs55834942 (HNF1A)	0.95 (0.99)	12	0.92	0.10	0.15
rs10203174	rs7578597 (THADA)	0.91 (0.98)	10	1.04	0.19	0.26
rs10401969	rs72999033 (HAPLN4)	0.90 (1.00)	5	1.13	0.27	0.34
rs7202877	rs13337017 (BCAR1)	0.93 (0.97)	6	1.07	0.29	0.34
rs516946	rs750625 (ANK1)	0.96 (0.99)	16	0.94	0.45	0.50
rs1552224	rs11603334 (ARAP1)	1.00 (1.00)	7	1.05	0.54	0.57
rs5215	rs11024271 (NCR3LG1)	0.88 (0.96)	11	0.97	0.65	0.65
Validation						
rs2007084	**rs41276922 (ANPEP)**	**0.83 (0.95)**	**10**	**1.51**	**1.6×10^{-3}**	**6.7×10^{-3}**
rs10923931	**rs835575 (NOTCH2)**	**0.99 (1.00)**	**10**	**0.70**	**1.9×10^{-3}**	**6.7×10^{-3}**
rs5215	**rs5215 (KCNJ11)**	**1.00 (1.00)**	**15**	**1.12**	**1.6×10^{-2}**	**3.0×10^{-2}**
rs4458523	**rs1801206 (WFS1)**	**0.93 (0.99)**	**18**	**0.93**	**3.1×10^{-2}**	**5.4×10^{-2}**
rs3802177	**rs3802177 (SLC30A8)**	**1.00 (1.00)**	**15**	**1.05**	**3.9×10^{-2}**	**6.2×10^{-2}**

SNPs exhibiting significant AEI (unadjusted $P < 0.05$) shown in bold. *Adjusted using Benjamini-Hochberg false discovery rate.

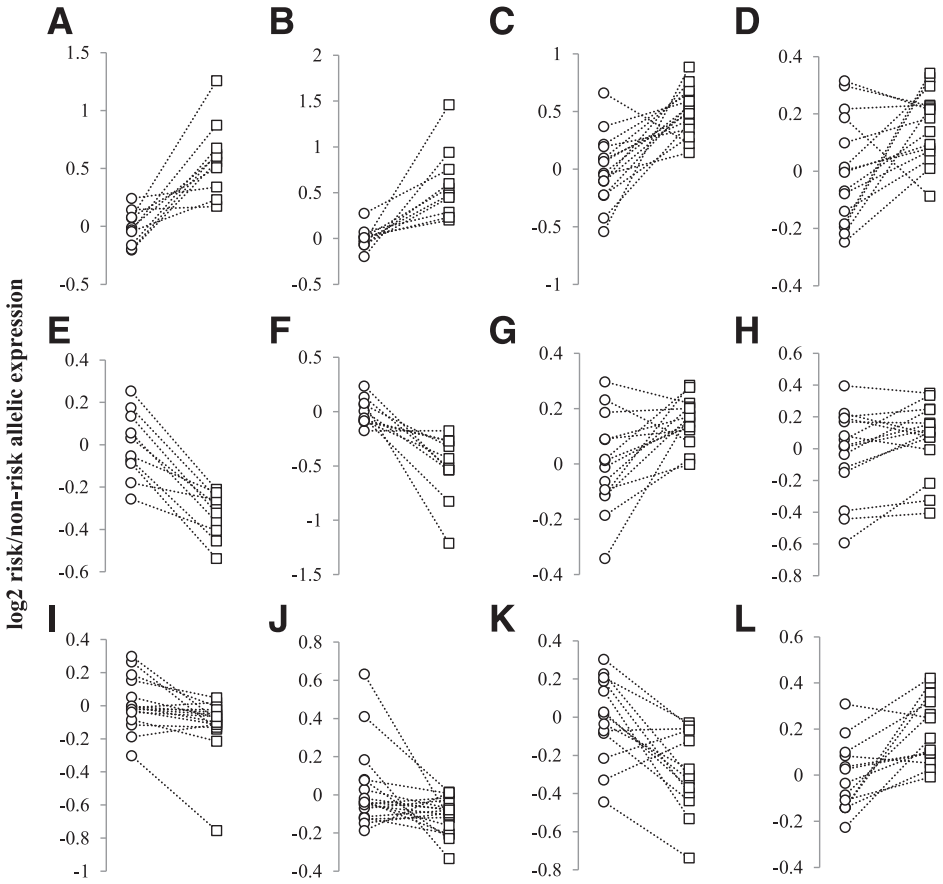

Figure 3—Significant ($P < 0.05$) AEI identified for SNPs in seven genes. *A:* rs17240240 (*ANPEP*); *B*: rs41276922 (*ANPEP*); *C*: rs5219 (*KCNJ11*); *D*: rs5215 (*KCNJ11*); *E*: rs699779 (*NOTCH2*); *F*: rs835575 (*NOTCH2*); *G*: rs11558471 (*SLC30A8*); *H*: rs3802177 (*SLC30A8*); *I*: rs1046320 (*WFS1*); *J*: rs1801206 (*WFS1*); *K*: rs1065359 (*CAMK2B*); *L*: rs952472 (*HMG20A*). White circles represent genomic DNA; white squares represent cDNA. Dotted lines show paired genomic DNA:cDNA samples.

expression of any of the closely residing genes and their isoforms. However, the presence of known and unknown *trans*-acting confounding factors, which cannot all be accounted for in regression analyses, adds significant variation to the gene expression measurements and likely contributes to the lower number of significant associations between genotype and expression. To enable a greater number of type 2 diabetes–associated loci to be examined by the more powerful allelic expression approach, future studies may consider using intronic SNPs to measure allelic expression. Allelic expression measurements calculated from intronic SNPs have been shown to be

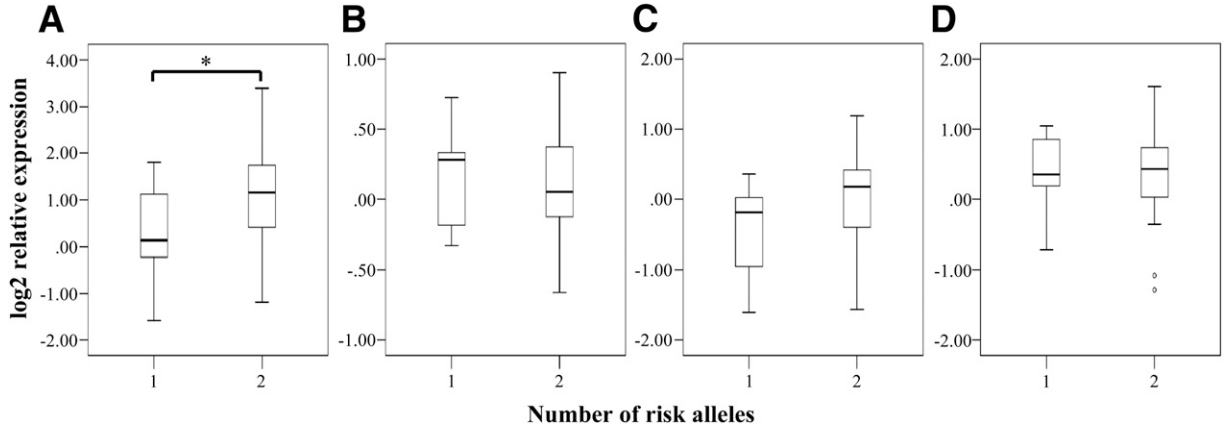

Figure 4—The rs2007084 genotype is significantly associated with (*A*) *ANPEP* expression but not expression of neighboring genes (*B*) *AP3S2*, (*C*) *ARPIN*, (*D*) *MESP1* in human islet tissue ($n = 36$). There were no individuals possessing zero risk alleles. *$P < 0.05$ (*t* test).

highly correlated with values calculated from linked exonic SNPs (16).

While this is, to our knowledge, the first report of a systematic use of allelic expression measurements in human islets to identify likely causal genes at type 2 diabetes GWAS loci, there are other published studies that have reported allelic imbalance in human islets at single loci. Kulzer et al. (17) report AEI for two SNPs in *ARAP1*, one of which (rs11603334) was also tested in our study. They report significant AEI for rs11603334 in six individuals with a mean AEI = 12.6%, whereas our study fails to identify significant AEI (P = 0.54) in the seven heterozygous individuals tested. It is difficult to understand the reasons for such disparate findings, given that allelic expression should remove much of the confounding effects of *trans*-acting factors, but it may be that differences in methodology (TaqMan vs. matrix-assisted laser desorption/ionization time-of-flight) and the small sample sizes (n = 6–7) explain the lack of replication. Another study by Nica et al. (8) analyzed allelic expression for type 2 diabetes GWAS-implicated genes using RNA-Seq in pancreatic β-cells from 11 donors. In support of our findings, they also report AEI for rs5215 (*KCNJ11*) and rs11558471 (*SLC30A8*), albeit only in four samples and one sample, respectively.

The lack of reports detailing AEIs for type 2 diabetes GWAS genes may be explained by the widespread use of nontargeted RNA-Seq methodologies. While whole transcriptome RNA-Seq approaches provide researchers with an unparalleled resource to investigate all transcripts, its lack of targeting of specific transcripts means the sequencing depth at loci of interest is not as high as could be achieved using targeted approaches. An examination of two recent studies using RNA-Seq methods to quantify the islet transcriptome (18,19) shows that the median reads per kilobase per million mapped reads for the genes at the 65 type 2 diabetes–associated GWAS loci is 8–9 reads per kilobase per million mapped reads. If one considers that current read lengths are ~50 bp and typical human islet sequencing depths vary, but are perhaps currently ~50 million mapped reads per sample, then an average 25 reads will map to one exonic SNP. Fontanillas et al. (20) calculate that 200 reads (equivalent to 8 samples) would give 80% power to detect AEI of 1.5, while to detect an AEI = 1.25, >500 mapped reads (>20 samples) would be needed (assuming a liberal type 1 error rate α = 5%). In our study, we found an AEI ≤1.25 for 5/7 loci where AEI was significant. Therefore to replicate our findings using current nontargeted RNA-Seq technologies would likely require a number of heterozygous individuals, which, given the difficulty in collecting large cohorts of human islets, is not likely to occur in the near future.

The significance of small AEIs (≤1.25) at five of the loci deserves discussion. Are such seemingly small imbalances of significant consequence to islet function? At one of these loci (*SLC30A8*), a missense SNP (rs13266634; W325R) is in complete LD (r^2 = 1; D′ = 1) with the lead SNP (rs3802177) also examined for AEI. There is some evidence that functional nonsynonymous variants have coevolved with regulatory variants so as to strengthen or diminish their functional impact (21). This would suggest that the increase in expression of the risk versus nonrisk allele is not directly altering disease risk in such cases, but rather moderating the effect of the nonsynonymous variant, which is primarily responsible for the molecular deficit increasing disease risk. However, with respect to *SLC30A8*, it has recently been reported that heterozygous loss-of-function mutations, which are likely null alleles, protect against type 2 diabetes (22). This is strong evidence that the precise levels of SLC30A8 are important for β-cell function. Given that we measured increased levels of *SLC30A8* transcripts originating from the risk, versus nonrisk, allele, one could speculate that it is the increase in SLC30A8 levels, rather than the presence of the missense variant, that increases type 2 diabetes risk. Indeed the sole study, to date, that examined the functional effects of the missense variant on zinc transporter activity reported only a mild attenuation in zinc ion transporter activity in cells overexpressing the risk (R) allele of W325R, compared with cells expressing the nonrisk (W) allele (23).

If we consider the other six loci where we observe significant AEI, two (*ANPEP*, *KCNJ11*) have at least one missense variant within that gene and in high LD (r^2 > 0.8) with the lead SNP, while at the *NOTCH2* locus, there is a missense SNP in a nearby gene (*ADAM30*). Whether the SNP in *ADAM30* is of consequence to the type 2 diabetes association is very questionable given its reported testis-specific expression (24), and the very low expression we and others (8,19) have observed in human islets. Furthermore, small changes in expression of NOTCH2, as identified in this study, may be of significant consequence to the β-cell with experiments in mice showing that when only one copy of *NOTCH2* is conditionally removed from Ngn3-positive cells (i.e., endocrine progenitors), the cells are diverted to an acinar fate (25). Therefore it may be hypothesized that carriers of *NOTCH2* risk alleles have reduced β-cell mass (and hence increased type 2 diabetes susceptibility), but difficulties in accurately measuring β-cell mass in large cohorts of individuals mean this theory is currently not easily amenable to testing. With regards to *ANPEP*, the one proxy SNP (rs17240268) resulting in a nonsynonymous substitution does not affect an evolutionary well-conserved residue. Therefore this increases the likelihood that the *ANPEP* expression changes we identified, via allelic expression and eQTL analyses, are causally involved in the association between SNPs at this locus and risk of type 2 diabetes. The *ANPEP* gene encodes the protein aminopeptidase N, a broad-specificity membrane-associated peptidase. Its widespread expression and involvement in many cellular processes (e.g., proliferation, apoptosis, antigen presentation, differentiation, angiogenesis, chemotaxis) (26) makes it difficult to predict exactly how increased expression predisposes to

type 2 diabetes. Indeed this may partially explain why investigators sought to proffer the neighboring gene, *AP3S2*, as the causal gene at this locus, this despite the fact that the lead and all proxy SNPs, as defined in this article, map within the *ANPEP* gene. The results of this study will hopefully stimulate research into the role of aminopeptidase N in the pathology of diabetes.

For three genes (*CAMK2B*, *HMG20A*, and *WFS1*), there are no nonsynonymous variants in high LD ($r^2 > 0.8$), strongly suggesting that at these loci, at least, alterations in gene expression are primarily involved in affecting type 2 diabetes risk. Our finding of a small AEI for *WFS1* may be of significant consequence to the β-cell given that homozygous null alleles in *WFS1* cause Wolfram syndrome, a rare, recessive disorder characterized by early-onset diabetes with a nonautoimmune origin (27), and a heterozygous mutation segregates with adult-onset diabetes in a large family (28). Our most unexpected finding is perhaps the significant AEI for *CAMK2B*. This gene is in close proximity to *GCK*, the gene that encodes glucokinase, a key enzyme involved in the first rate-limiting step of glucose metabolism in the β-cell and a gene where heterozygous inactivating mutations cause persistent mild hyperglycemia and homozygous inactivating mutations cause permanent neonatal diabetes (29). Characterization of the effects of the type 2 diabetes–associated SNP at this locus on glycemic traits show a strong association with fasting glucose (7), further implicating *GCK* as the causative gene. One cannot rule out, however, a role for variation in *CAMK2B* levels affecting β-cell function, given that this gene encodes a calcium/calmodulin-dependent protein kinase that has been shown to be involved in coupling Ca^{2+} entry with insulin secretion (30–32) and that this isoform is the predominant form in pancreatic β-cells (31,33). Little is known about HMG20A function in human islets, but its involvement in neuronal differentiation (34,35) may point toward a similar function in the neuronal-like β-cell.

While we have identified AEIs in genes implicated by type 2 diabetes GWAS in human islets, it is not altogether clear whether altered expression of these genes in islets is the primary mechanism leading to changes in disease susceptibility. The most recent and largest study to date examining the effect of type 2 diabetes GWAS variants on continuous glycemic traits reports 12/36 loci examined showing defects in β-cell function, while only 4/36 were associated with insulin sensitivity measures (7). Of the 12 loci associated with defects in β-cell function, we report AEI of *SLC30A8* and *CAMK2B* at the *SLC30A8* and *GCK* loci, respectively. At the remaining 20 loci, Dimas et al. (7) report no discernible effect of the SNPs on glycemic traits; this includes the *NOTCH2*, *KCNJ11*, and *WFS1* loci exhibiting AEI in our study. Given the raft of measures used to assay β-cell function, in particular, one might speculate that SNPs at these twenty loci impact on other cell types relevant to diabetes pathogenesis. However, it is worth noting that a number of loci harboring genes where mutations cause monogenic diabetes (*HNF1A*, *HNF1B*, *KCNJ11*, *WFS1*), and that do so through a detrimental effect on β-cell function, are included in this list with unknown effects on glycemic traits. Therefore one cannot rule out the possibility that the genes where we have uncovered AEI in islets, but where there is no data currently to support an effect of the type 2 diabetes–associated SNP on β-cell function (either through physiological analyses of SNPs on glycemic measures or presence of monogenic forms of diabetes), do so via an effect in this tissue. Furthermore, while regulation of gene expression can be highly tissue specific, the findings of significant numbers of *cis*-eQTLs that are tissue independent (shared across two or more tissues) mean that AEI observed in islets may be observed in other disease-relevant tissues (36–38).

In summary, we report AEIs that are consistent with type 2 diabetes–associated variation regulating the expression of *cis*-linked genes in human islets. For some of the genes where significant AEI was identified (e.g., *SLC30A8*, *WFS1*), there is strong evidence from human genetics that small changes in gene dosage may have significant consequences for the pancreatic β-cell. For other genes with significant AEI (e.g., *ANPEP*, *HMG20A*), their role is less well defined, and hence this study should provide a platform for further work examining the effects of carefully manipulating the expression of these genes in human islets.

Acknowledgments. The authors thank Dr. Tim McDonald and Tim Frayling (both at the University of Exeter Medical School) for helpful discussions during this study.

Funding. This work was supported by the Medical Research Council (grant number MR/J006777/1).

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

Author Contributions. J.M.L. conceived and designed the experiments, conducted experiments and/or analyzed data, interpreted the results, and wrote the manuscript. G.H. and A.R.W. conducted experiments and/or analyzed data and interpreted the results. M.N.W. conducted experiments and/or analyzed data, interpreted the results, and reviewed and edited the manuscript. L.W.H. conceived and designed the experiments, interpreted the results, and reviewed and edited the manuscript. L.W.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. 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

2. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105–116

3. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579–589
4. Kooner JS, Saleheen D, Sim X, et al.; DIAGRAM; MuTHER. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 2011;43:984–989
5. Musunuru K, Strong A, Frank-Kamenetsky M, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 2010;466:714–719
6. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature 2012;492:369–375
7. 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
8. Nica AC, Ongen H, Irminger JC, et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 2013;23:1554–1562
9. Lyssenko V, Lupi R, Marchetti P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest 2007;117:2155–2163
10. Taneera J, Lang S, Sharma A, et al. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab 2012;16:122–134
11. Ge B, Pokholok DK, Kwan T, et al. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. Nat Genet 2009;41:1216–1222
12. Heap GA, Yang JH, Downes K, et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. Hum Mol Genet 2010;19:122–134
13. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010;6:e1000888
14. Zhong H, Beaulaurier J, Lum PY, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. PLoS Genet 2010;6:e1000932
15. Kwan T, Benovoy D, Dias C, et al. Genome-wide analysis of transcript isoform variation in humans. Nat Genet 2008;40:225–231
16. Lalonde E, Ha KC, Wang Z, et al. RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. Genome Res 2011;21:545–554
17. Kulzer JR, Stitzel ML, Morken MA, et al. A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. Am J Hum Genet 2014;94:186–197
18. Cnop M, Abdulkarim B, Bottu G, et al. RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate. Diabetes 2014;63:1978–1993
19. Eizirik DL, Sammeth M, Bouckenooghe T, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet 2012;8:e1002552
20. Fontanillas P, Landry CR, Wittkopp PJ, et al. Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. Mol Ecol 2010;19(Suppl. 1):212–227
21. Montgomery SB, Lappalainen T, Gutierrez-Arcelus M, Dermitzakis ET. Rare and common regulatory variation in population-scale sequenced human genomes. PLoS Genet 2011;7:e1002144
22. 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
23. Nicolson TJ, Bellomo EA, Wijesekara N, et al. Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. Diabetes 2009;58:2070–2083
24. Cerretti DP, DuBose RF, Black RA, Nelson N. Isolation of two novel metalloproteinase-disintegrin (ADAM) cDNAs that show testis-specific gene expression. Biochem Biophys Res Commun 1999;263:810–815
25. Cras-Méneur C, Li L, Kopan R, Permutt MA. Presenilins, Notch dose control the fate of pancreatic endocrine progenitors during a narrow developmental window. Genes Dev 2009;23:2088–2101
26. Mina-Osorio P. The moonlighting enzyme CD13: old and new functions to target. Trends Mol Med 2008;14:361–371
27. Hofmann S, Philbrook C, Gerbitz KD, Bauer MF. Wolfram syndrome: structural and functional analyses of mutant and wild-type wolframin, the WFS1 gene product. Hum Mol Genet 2003;12:2003–2012
28. Bonnycastle LL, Chines PS, Hara T, et al. Autosomal dominant diabetes arising from a Wolfram syndrome 1 mutation. Diabetes 2013;62:3943–3950
29. Osbak KK, Colclough K, Saint-Martin C, et al. Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia. Hum Mutat 2009;30:1512–1526
30. Dadi PK, Vierra NC, Ustione A, Piston DW, Colbran RJ, Jacobson DA. Inhibition of pancreatic β-cell Ca2+/calmodulin-dependent protein kinase II reduces glucose-stimulated calcium influx and insulin secretion, impairing glucose tolerance. J Biol Chem 2014;289:12435–12445
31. Tabuchi H, Yamamoto H, Matsumoto K, et al. Regulation of insulin secretion by overexpression of Ca2+/calmodulin-dependent protein kinase II in insulinoma MIN6 cells. Endocrinology 2000;141:2350–2360
32. Wang Z, Ramanadham S, Ma ZA, et al. Group VIA phospholipase A2 forms a signaling complex with the calcium/calmodulin-dependent protein kinase IIbeta expressed in pancreatic islet beta-cells. J Biol Chem 2005;280:6840–6849
33. Rochlitz H, Voigt A, Lankat-Buttgereit B, et al. Cloning and quantitative determination of the human Ca2+/calmodulin-dependent protein kinase II (CaMK II) isoforms in human beta cells. Diabetologia 2000;43:465–473
34. Wynder C, Hakimi MA, Epstein JA, Shilatifard A, Shiekhattar R. Recruitment of MLL by HMG-domain protein iBRAF promotes neural differentiation. Nat Cell Biol 2005;7:1113–1117
35. Ceballos-Chávez M, Rivero S, García-Gutiérrez P, et al. Control of neuronal differentiation by sumoylation of BRAF35, a subunit of the LSD1-CoREST histone demethylase complex. Proc Natl Acad Sci U S A 2012;109:8085–8090
36. Nica AC, Parts L, Glass D, et al.; MuTHER Consortium. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 2011;7:e1002003
37. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. Nature 2008;452:423–428
38. Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009;325:1246–1250