

Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism

João Fadista^{a,1}, Petter Vikman^a, Emilia Ottosson Laakso^a, Inês Guerra Mollet^a, Jonathan Lou Esguerra^a, Jalal Taneera^a, Petter Storm^a, Peter Osmark^a, Claes Ladenvall^a, Rashmi B. Prasad^a, Karin B. Hansson^a, Francesca Finotello^b, Kristina Uvebrant^a, Jones K. Ofori^a, Barbara Di Camillo^b, Ulrika Krus^a, Corrado M. Cilio^a, Ola Hansson^a, Lena Eliasson^a, Anders H. Rosengren^a, Erik Renström^a, Claes B. Wollheim^{a,c}, and Leif Groop^{a,1}

^aLund University Diabetes Centre, Department of Clinical Sciences, Skåne University Hospital Malmö, Lund University, 20502 Malmö, Sweden; ^bDepartment of Information Engineering, University of Padova, 35131 Padova, Italy; and ^cDepartment of Cell Physiology and Metabolism, University of Geneva, 1211 Geneva 4, Switzerland

Edited by Tak W. Mak, The Campbell Family Institute for Breast Cancer Research at Princess Margaret Cancer Centre, Ontario Cancer Institute, University Health Network, Toronto, Canada, and approved August 14, 2014 (received for review February 11, 2014)

Genetic variation can modulate gene expression, and thereby phenotypic variation and susceptibility to complex diseases such as type 2 diabetes (T2D). Here we harnessed the potential of DNA and RNA sequencing in human pancreatic islets from 89 deceased donors to identify genes of potential importance in the pathogenesis of T2D. We present a catalog of genetic variants regulating gene expression (eQTL) and exon use (sQTL), including many long noncoding RNAs, which are enriched in known T2D-associated loci. Of 35 eQTL genes, whose expression differed between normoglycemic and hyperglycemic individuals, siRNA of tetraspanin 33 (TSPAN33), 5'-nucleotidase, ecto (NT5E), transmembrane emp24 protein transport domain containing 6 (TMED6), and p21 protein activated kinase 7 (PAK7) in INS1 cells resulted in reduced glucose-stimulated insulin secretion. In addition, we provide a genome-wide catalog of allelic expression imbalance, which is also enriched in known T2D-associated loci. Notably, allelic imbalance in paternally expressed gene 3 (PEG3) was associated with its promoter methylation and T2D status. Finally, RNA editing events were less common in islets than previously suggested in other tissues. Taken together, this study provides new insights into the complexity of gene regulation in human pancreatic islets and better understanding of how genetic variation can influence glucose metabolism.

Type 2 diabetes (T2D) is an increasing global health problem (1). Although genome-wide association studies (GWAS) have yielded more than 70 loci associated with T2D or related traits (2, 3), they have not provided the expected breakthrough in our understanding of the pathogenesis of the disease. They have nonetheless pointed at a central role of the pancreatic islets and β -cell dysfunction in the development of the disease (4, 5). It therefore seems pertinent to focus on human pancreatic islets to obtain insights into the molecular mechanisms causing the disease (6, 7). Given that most SNPs associated with T2D lie in noncoding regions, the majority of causal variants are likely to regulate gene expression rather than protein function per se. Therefore, combination of DNA and RNA sequencing in the same individuals may help to disentangle the role these SNPs play in the pathogenesis of the disease (8). Although the human pancreatic islet transcriptome has been previously described (6, 9–18), using microarrays or RNA sequencing of a limited number of nondiabetic individuals, this has not allowed a more global analysis of the complexity of the islet transcriptome in T2D. Here we combined genotypic imputation, expression microarrays, and exome and RNA sequencing (Exome-Seq and RNA-Seq) in a large number of human pancreatic islets from deceased donors with and without T2D. This study identified a number of novel genes, including long intergenic noncoding RNAs (lincRNAs), whose expression and/or splicing influences insulin secretion and is associated with glycemia. In addition, we provide a catalog of RNA editing and allele-specific expression events in human pancreatic islets (*SI Appendix, Fig. S1*).

Results

Genes Showing Differential Expression Between Islets from Normoglycemic and Hyperglycemic Donors. To obtain a profile of gene expression variation in human islets, we sequenced the polyadenylated fraction of RNA from 89 individuals with different degrees of glucose tolerance, using 101 base pairs paired-end on an Illumina HiSeq sequencer (*Dataset S1* and *SI Appendix, Fig. S1*). Each individual transcriptome yielded, on average, 38.2 ± 4.4 (mean \pm SD) million paired-end reads mapped to the human genome, with $\sim 88\%$ mapping to known exons from the RefSeq Gene database. Because any expression cutoff is arbitrary, we considered a gene to be expressed if it was observed in at least 5% of the samples. Applying this definition, we detected 91% of RefSeq genes. However, most of these genes are expressed at low levels, supporting the view of pervasive transcription and leakage in the human transcriptome (*SI Appendix, Fig. S2*) (19). Moreover, in support of previous results (20), we observed a good correlation between gene expression based on RNA-seq and microarrays in the 89 samples ($r = 0.83$; $P < 0.001$) (*SI Appendix, Fig. S3*). To evaluate how well the RefSeq genes in our RNA-seq dataset are covered, we sequenced one sample at high depth of coverage (~ 150 million paired-end reads). As seen in *SI Appendix, Fig. S4*, our average sample coverage of 38.2 million paired-end reads is deep enough to detect the majority of known genes, transcripts, exons, and junctions. As expected, glucagon, insulin, and other

Significance

We provide a comprehensive catalog of novel genetic variants influencing gene expression and metabolic phenotypes in human pancreatic islets. The data also show that the path from genetic variation (SNP) to gene expression is more complex than hitherto often assumed, and that we need to consider that genetic variation can also influence function of a gene by influencing exon usage or splice isoforms (sQTL), allelic imbalance, RNA editing, and expression of noncoding RNAs, which in turn can influence expression of target genes.

Author contributions: J.F., P.V., O.H., E.R., and L.G. designed research; J.F., E.O.L., J.L.E., J.T., P.O., R.B.P., K.B.H., K.U., J.K.O., C.M.C., L.E., and A.H.R. performed research; J.F. and U.K. contributed new reagents/analytic tools; J.F., L.G.M., P.S., C.L., R.B.P., F.F., and B.D.C. analyzed data; and J.F., C.B.W., and L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Database deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE50398).

¹To whom correspondence may be addressed. Email: joao.fadista@med.lu.se or Leif.Groop@med.lu.se.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402665111/-DCSupplemental.

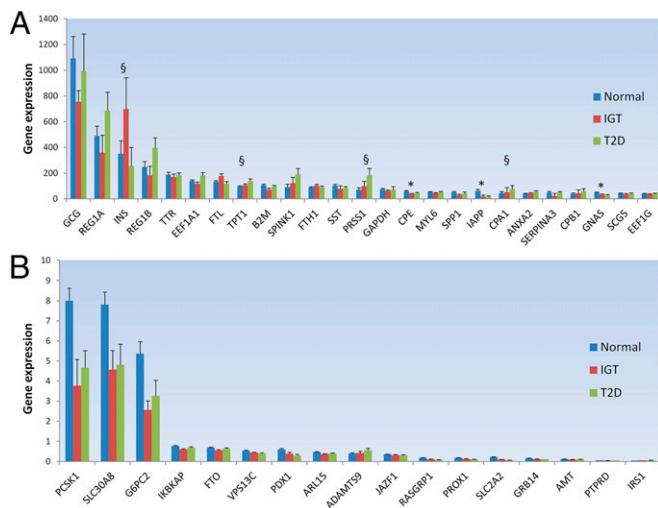


Fig. 1. Genes expressed in 89 human pancreatic islets stratified by glucose tolerance status. (A) RNA-seq normalized median expression of the top 25 nonribosomal genes expressed in islets. (B) RNA-seq normalized median expression of the 17 genes that show significant islet expression association with glucose tolerance status and are putatively associated with established T2D and glycemic associated loci (2, 3, 23–28). Genes are ordered by decreasing median expression in all 89 islet donors. Normal corresponds with normoglycemic donors (HbA1c < 6%; $n = 51$), IGT corresponds with impaired glucose-tolerant donors ($6\% \leq \text{HbA1c} < 6.5\%$; $n = 15$), and T2D corresponds with diabetic donors ($\text{HbA1c} \geq 6.5\%$; $n = 12$). Error bars represent SEM values. *Genes that show significant expression association with glucose tolerance status detected both by expression arrays and RNA-seq with both nominal and permutation P values < 0.05 (after performing 10,000 permutations). §Additional genes that show significant expression association with glucose tolerance status detected only with RNA-seq (at permuted P value < 0.05).

known pancreatic genes showed the highest expression (Fig. 1A and Dataset S1). To identify genes whose expression is influenced by glycemia (cause or consequence), we related gene expression to HbA1c, a measure of long-term glycemia, and compared expression in islets from donors with normal glucose tolerance (HbA1c <

6%), impaired glucose tolerance (IGT; $6\% \leq \text{HbA1c} < 6.5\%$), and T2D ($\text{HbA1c} \geq 6.5\%$). By using a linear model adjusting for age and sex, we detected 1619 genes associated with HbA1c levels in both RNA-seq and microarrays (Database S1 and SI Appendix). Briefly, genes were kept if both microarray and RNA-seq gene expression were nominally associated with HbA1c levels, with both nominal and permutation P values < 0.05. In addition, 271 genes showed specific exon associations, with HbA1c levels not detected at the gene level (Database S1). Of the genes associated with HbA1c levels, 70 were also associated with in vitro insulin secretion in human islets, further highlighting their role in glucose metabolism (Database S1). Of particular interest are the genes whose expression is associated with lower HbA1c levels and higher insulin secretion, such as RAS guanyl releasing protein 1 (*RASGRF1*) (6), transcription factor *RFX3* (21), and nicotinamide nucleotide transhydrogenase (*NNT*) (22), all of which have been suggested to regulate insulin secretion (SI Appendix, Figs. S5–S7). *RFX3* has also been suggested to regulate the glucokinase promoter, and thereby its expression in a mouse insulinoma cell line MIN6 (21). In line with these findings, we observed a clear coexpression between the *RFX3* and *GCK* genes in human islets (SI Appendix, Fig. S19). Of the established T2D and glycemic associated loci (2, 3, 23–28) whose gene expression proxies were associated with HbA1c levels, solute carrier family 30 (zinc transporter), member 8 (*SLC30A8*), glucose-6-phosphatase, catalytic, 2 (*G6PC2*), and proprotein convertase subtilisin/kexin type 1 (*PCSK1*) showed the highest expression using RNA-seq (Fig. 1B and Database S1). This is in line with our previous findings using microarrays (6). *SLC30A8*, *G6PC2*, and *PCSK1* also showed a strong positive correlation with glucagon expression (SI Appendix, Fig. S8). By using the RABT Cufflinks transcript assembly method (29), we also detected 445 potential novel genes with exon–exon junctions in addition to the existing GENCODE (30), UCSC, and Ensembl gene structure annotations (Database S1). Of these potential novel gene loci, 28 (6%) show coding potential, as assessed by the CPAT tool (31), and 391 (88%) are within 5 kb of known human islet active chromatin DNase, FAIRE, or H3K4me3 peaks (32–34), pointing to candidate nearby promoters for those genes (Database S1). One of these potential novel genes, although not showing any coding potential nor close to any known islet open chromatin mark, was also associated with HbA1c, and this new gene locus is in

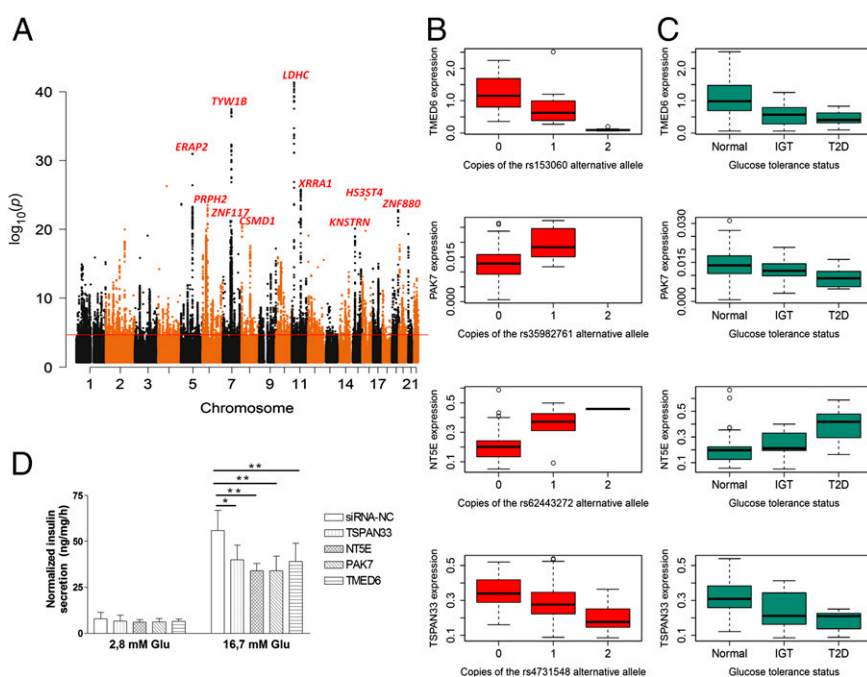


Fig. 2. Landscape of cis-eQTLs (≤ 250 kb) in 89 human pancreatic islets and eQTL genes validated to interfere with glucose metabolism and insulin secretion. (A) Manhattan plot of the best P value per SNP, showing the top 10 eQTL genes (FDR < 1% line drawn in black at P value = 2.267×10^{-5}). (B) eQTL SNP genotype versus eQTL gene expression ($n = 89$). All eQTLs pass FDR < 1% and 10,000 permutations. (C) eQTL gene expression stratified by glucose tolerance status [normal, HbA1c < 6% ($n = 51$); IGT, $6\% \leq \text{HbA1c} < 6.5\%$ ($n = 15$); T2D, HbA1c $\geq 6.5\%$ ($n = 12$)]. All genes have a nominal and permutation P value < 0.05 in both RNA-seq and microarrays (after 10,000 permutations). (D) Insulin secretion in response to 2.8 and 16.7 mM glucose 72 h after siRNA transfection, as measured during 1 h static incubation. Data are shown from three independent experiments for each siRNA. Data are normalized for protein content. Bars represent mean \pm SD. ** $P < 0.01$ and * $P < 0.05$ versus control siRNA. The knockdown efficiency of *NTSE*, *PAK7*, and *TMED6* were above 80%, whereas for *TSPAN33* it was 50%.

Table 1. Genome-wide significant GWAS T2D/glycemic hits as eQTLs in human pancreatic islets

SNP	Nearest gene	eQTL gene	eQTL <i>P</i> value	Allele change	eQTL direction	GWAS trait	GWAS effect allele
rs1019503	<i>ERAP2</i>	<i>ERAP2</i>	7.1e−24	G > A	+	2-h glucose (3)	A
rs2028299	<i>AP3S2</i>	<i>AP3S2</i>	1.1e−14	C > A	−	T2D (24)	A
rs505922	<i>ABO</i>	<i>ABO</i>	5.3e−08	T > C	+	Disposition index (25)	C
rs10830963	<i>MTNR1B</i>	<i>MTNR1B</i>	8.6e−08	C > G	+	T2D, fasting glucose (2, 26)	G
rs1535500	<i>KCNK16</i>	<i>KCNK17</i>	1.2e−06	G > T	+	T2D (27)	T

a ~10-kb region nominally significant in the MAGIC database for fasting glucose (23) (*SI Appendix, Fig. S9*).

Effect of SNPs on Gene Expression (eQTLs) and Splicing (sQTLs) in Human Pancreatic Islets. Because many SNPs are located in non-coding regions, suggesting they may influence gene expression, we analyzed whether any SNP genotyped in our islet samples and further imputed to the 1000 Genomes reference panel (35) would influence RNA-seq gene expression (eQTL) or exon use (sQTL) in cis (within 250 kb of the SNP). For analysis purposes, we identified a single best “sentinel” SNP for each gene or exon, defined as the SNP with the lowest *P* value per eQTL or sQTL gene. Applying these criteria and thresholds, we identified 616 cis eQTLs for known genes (Fig. 2A and Database S1), whereas 24 eQTLs were detected in previously unknown transcribed loci (Database S1; *Materials and Methods*). Notably, 54% of these eQTLs would have been missed in a microarray because the gene is not probed on the array or shows low expression (*SI Appendix, Fig. S10*).

Our sample size permitted us to detect significant eQTLs at >90% power with an effect size of 0.5 (beta) or more (*SI Appendix, Fig. S11*). Notably, only in about half of the cases did the eQTL SNPs influence expression on the nearest gene. The strongest eQTLs were detected in the lactate dehydrogenase C (*LDHC*), tRNA-yW synthesizing protein 1 homolog B (*Saccharomyces cerevisiae*) (*TYW1B*), and endoplasmic reticulum aminopeptidase 2 (*ERAP2*) genes (Fig. 2A). *LDHC* encodes the enzyme lactate dehydrogenase C, which catalyzes the glycolytic conversion of lactate to pyruvate. Although expression of another lactate dehydrogenase, *LDHA*, is suggested to be repressed in pancreatic β cells because of a minor role of anaerobic glycolysis in the adult β cell (36), *LDHC* is expressed at similar levels in α and β cells (17). Knock-down of *LDHC* and *TYW1B* in INS-1 cells using siRNA did not affect insulin secretion. *ERAP2* has been ascribed a role in autoimmunity and type 1 diabetes (37), and its eQTL sentinel SNP is in high linkage disequilibrium ($r^2 > 0.8$) with the genome-wide significant GWAS SNP rs1019503 for glucose levels 2 h after an oral glucose challenge (3). Furthermore, we observed 371 splicing QTLs (sQTLs) not reflected by changes in expression at the gene level (Database S1 and *SI Appendix, Fig. S13*). There was eQTL and sQTL enrichment in regions of islet active chromatin, such as those characterized by DNase I hypersensitivity [Fisher exact test, *P* value < 2.2e−16 (odds ratio = 2.1) for eQTLs; and *P* value = 3.8e−11 (odds ratio = 1.9 for sQTLs)], H3K4m3 [*P* value < 2.2e−16 (odds ratio = 3.1), for eQTLs and *P* value < 2.2e−16 (odds ratio = 2.3) for sQTLs], and FAIRE [*P* value = 2.1e−07 (odds ratio = 2.7) for eQTLs and *P* value = 0.02 (odds ratio = 1.9) for sQTLs] (32–34). There was no indication that these eQTLs and sQTLs were enriched in evolutionarily conserved sites.

Because GWAS for T2D only enabled identification of loci, rather than genes, we examined whether SNPs known to associate with T2D or related traits (glucose, insulin) would have a cis effect on gene expression or exon use. We found enrichment for GWAS T2D/glycemic trait loci in eQTLs (Fisher exact test, *P* value 4.1e−03; odds ratio = 5.0), with five GWAS SNPs showing a cis eQTL effect, and in the case of rs1535500, the effect was not on the nearest gene (Table 1 and *SI Appendix, Fig. S14*). Notably, of the 1,619 genes whose expression correlated with HbA1c, 35 (2%) had an eQTL (Database S1). We examined whether the eQTL SNPs in these genes were associated with insulin and glucose concentrations in the DIAGRAM and MAGIC databases (2, 3, 23–28). The sentinel eQTL SNP for sorting nexin 19 (*SNX19*), rs3751034, was nominally

associated with HbA1c in MAGIC (*P* value < 0.01) (28). *SNX19* has also been shown to regulate insulin secretion in a mouse pancreatic β -cell line (38). Finally, we tested whether the three eQTL genes [tetraspanin 33 (*TSPAN33*), 5′-nucleotidase, ecto (*NT5E*), and transmembrane emp24 protein transport domain containing 6 (*TMED6*)] showing the strongest effect on HbA1c levels would also influence insulin secretion by disrupting their expression in INS-1 cells. We also tested p21 protein activated kinase 7 (*PAK7*), a gene associated with HbA1c levels in both RNA-seq and microarray but only detected as an eQTL gene by RNA-seq. Down-regulation of the expression of these genes was associated with significantly reduced glucose-stimulated insulin secretion (Fig. 2B–D). Taken together, we present a list of SNPs influencing gene expression in human pancreatic islets with a likely role in regulating glucose homeostasis.

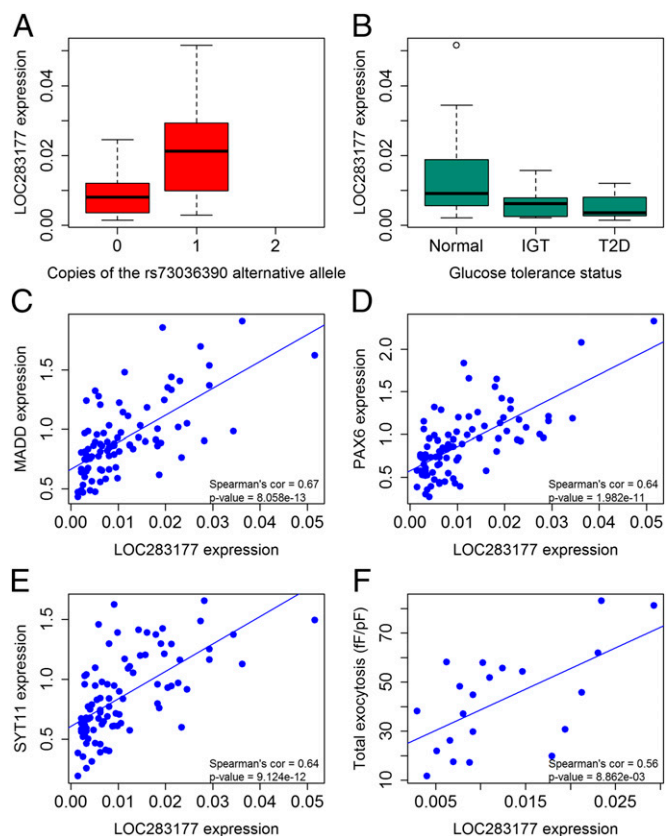


Fig. 3. Expression analysis of the lincRNA *LOC283177* in 89 human pancreatic islets. (A) The lincRNA *LOC283177* has an eQTL ($n = 89$, significant at FDR < 1% and 10,000 permutations), (B) which is associated with HbA1c in the islet donors (nominal and permutation *P* value < 0.05 after 10,000 permutations). Normal corresponds to normoglycemic donors (HbA1c < 6.5%; $n = 51$), IGT corresponds to impaired glucose tolerant donors (6% ≤ HbA1c < 6.5%; $n = 15$), and T2D corresponds to diabetic donors (HbA1c ≥ 6.5%; $n = 12$). *LOC283177* is coexpressed with the diabetic genes (C) *MADD*, (D) *PAXG*, (E) *SYT11*, and (F) associates with depolarization-evoked insulin exocytosis (Spearman correlation test significance at FDR < 1%).

eQTLs and HbA1c Influence Expression of lincRNAs. lincRNAs have recently been ascribed a role in the regulation of gene transcription, including pancreatic islets (16). We identified 493 RefSeq lincRNAs expressed in the pancreatic islets, with 54 of those being influenced by eQTLs or sQTLs and/or related to HbA1c levels (Database S1). Of these 54 lincRNAs, seven (13%) have also been reported in a study of lincRNAs in human pancreatic β cells (15). Of the 616 eQTLs we identified (Fig. 2A), 33 (5%) influenced the expression of lincRNAs, eight of which have also been reported in other tissues. Moreover, six (2%) of 371 sQTLs were seen within lincRNAs. Notably, 17 lincRNAs were significantly associated with HbA1c levels, two of which also had an eQTL (*LOC283177* and *SNHG5*) (Database S1). To obtain insight into putative target genes of these two lincRNAs, we performed a coexpression analysis linking their expression with all other genes in pancreatic islets. This analysis showed a strong coexpression of the MAP-kinase activating death domain (*MADD*), synaptotagmin 11 (*SYT11*), and paired box 6 (*PAX6*) genes with *LOC283177* (Fig. 3). All these genes have been ascribed a key role in islet function. Synaptotagmin 11 (*SYT11*) is known to regulate exocytosis of insulin (39) and *MADD* proinsulin synthesis (25), and *PAX6* is involved in development of pancreatic islets (40). In support of this, *LOC283177* expression was directly associated with insulin exocytosis in the islets (Fig. 3F). The lincRNA *ANRIL* (also known as *CDKN2B-AS1*), located in a locus on chromosome 9p, has been associated with both T2D (2) and cardiovascular disease. Although eQTLs for *ANRIL* have been reported in human blood (41), we could not detect any eQTL for *ANRIL* in human islets or any coexpressed genes.

Allelic Expression Imbalance. Both allelic expression imbalance (AEI) and cis-QTL analysis detect genetic effects on gene transcription, although they frequently do not capture the same loci. Whereas eQTL and sQTL refer to the effect of a SNP on expression of the gene or specific exons/isoforms, respectively, AEI refers to imbalance between expression of maternal and paternal alleles and, consequently, can only be detected in the case of heterozygosity. To further elucidate the cis-regulatory potential in islets, we searched for genes showing allelic expression imbalance. We compared transcriptome and exome sequencing from the same individuals, using Fisher test to define significant deviation from the expected 50/50 allelic distribution for the SNPs. Thereby,

we could detect 1,528 SNPs showing potential allelic imbalance in at least two samples at false discovery rate (FDR) $<1\%$ (Database S1). These encompass 1,102 genes, 14% of which have been previously suggested to be imprinted and/or showing imbalance of expression in other human tissues (Database S1). Only 3% of eQTLs and 0.5% of sQTLs were in strong linkage disequilibrium with an AEI SNP. We validated, with Sanger sequencing, an AEI in the *MMP7* gene and showed that the missense variant rs10502001 is nominally associated with exocytosis of insulin (SI Appendix, Fig. S15). To detect allelic imbalance sites relevant to T2D, we filtered AEI sites ascertained in at least 50% of the samples and found *PEG3* (paternally expressed gene 3). *PEG3* is a gene known to be imprinted in other tissues and to change its methylation levels in murine oocytes of diabetic females (42), suggesting a link between allelic imbalance and imprinting/methylation. Notably, we show clear differences in the degree of methylation in a region of the *PEG3* promoter, being hypomethylated in T2D islets, which do not have allelic imbalance (Fig. 4). These data indicate that differential methylation could be the cause of allelic imbalance, which in turn could influence susceptibility to T2D. Of SNPs associated with T2D or related glycemic traits, we found enrichment for allelic imbalance (Fisher exact test P value = $2.3e^{-06}$; odds ratio = 10.1), with eight showing evidence of allelic imbalance in solute carrier family 2, member 2 (*SLC2A2*), adaptor-related protein complex 3, sigma 2 subunit (*AP3S2*), thyroid adenoma associated (*THADA*), *MADD*, *ERAP2*, aminomethyltransferase (*AMT*), forkhead box A2 (*FOXA2*), and La ribonucleoprotein domain family, member 6 (*LARP6*) loci (Table 2).

RNA Editing. Finally, we assessed the frequency of RNA editing events in the pancreatic islet transcriptome, using a stringent pipeline to identify differences between DNA and RNA sequences, by comparing exome and RNA sequencing data (SI Appendix, Fig. S16). We found 65 loci showing potential RNA editing in at least two individuals overlapping 61 genes, two of which were in loci associated with T1D and T2D, *GLIS3* (44) and *ZFAND3* (2) (Database S1). Seven of the RNA editing events have also been reported before. As previously observed (45), the majority of RNA editing events were localized in the 3'UTR region or downstream of genes (67%), suggesting that RNA editing might play a role in miRNA-mediated regulation of gene expression by altering miRNA target sites or by affecting degradation of RNA. As

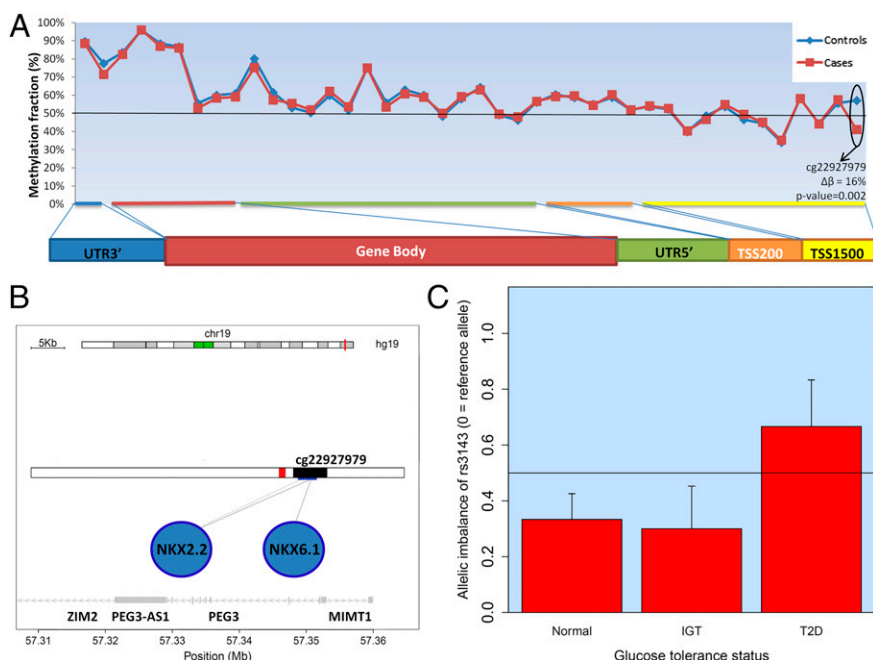


Fig. 4. Allelic imbalance in the *PEG3* gene is associated with diabetic status. (A) Methylation fraction of different parts of *PEG3* was assessed by Illumina's 450k chip in a subset of islets (adjusted t -test P value depicted). Cases were three patients with T2D, and controls were 21 normal patients or patients with IGT, sampled at random from the 89 total samples. Methylation of a CpG site in the promoter of *PEG3* was significantly higher in the nondiabetic islets. (B) The CpG site is in the islet active promoter of *PEG3* (43). The black bar corresponds to *PEG3* promoter, and the red bar corresponds to an islet active enhancer region. (C) Probed with the SNP rs3143, the RNA allelic imbalance is more pronounced in nondiabetic (normal and IGT, $n = 37$) than diabetic (T2D, $n = 9$) islets (P value = 0.03, Wilcoxon rank sum test). The y axis being 0.5 means no allelic imbalance, whereas y axis = 1 or 0 means expression of only one allele; that is, imprinting. Not all normal, IGT, or T2D were tested because of the harsh filtering criteria before allelic imbalance testing (SI Appendix). TSS1500, 1,500 base pairs from transcription start site. TSS200, 200 base pairs from transcription start site. UTR3', 3' untranslated region. UTR5', 5' untranslated region.

Table 2. Allelic imbalance loci in high linkage disequilibrium with genome-wide significant GWAS T2D/glycemic trait hits

AEI gene	AEI SNPs	Linkage disequilibrium $r^2 > 0.8$ with GWAS top SNP	GWAS trait	GWAS effect allele
<i>SLC2A2</i>	rs55679742, rs55989805	rs11920090	Fasting glucose (26)	T
<i>AP3S2</i>	rs2028299	rs2028299	T2D (24)	A
<i>LARP6</i>	rs3825970	rs1549318	Fasting proinsulin (25)	T
<i>THADA</i>	rs7578597	rs10203174	T2D (2)	C
<i>MADD</i>	rs35233100	rs35233100	Fasting proinsulin (25)	C
<i>ERAP2</i>	rs2287988, rs2548538	rs1019503	2-h glucose (3)	A
<i>AMT</i>	rs6997	rs11715915	Fasting glucose (3)	C
<i>FOXA2</i>	rs6048192	rs6113722	Fasting glucose (3)	G

described for other tissues, the vast majority of RNA editing events were canonical A-to-G events (83% if T-to-C events indicative of A-to-I editing on the opposite strand also are included) (*SI Appendix, Fig. S17*) (45). By randomly choosing nine editing events, we could validate three of the six A-to-G events by Sanger sequencing, but none of the three non A-to-G events (*SI Appendix, Fig. S18*). This low validation rate brings into question some previous reports of thousands of RNA editing events based simply on RNA sequencing (46).

Discussion

By combining RNA and exome sequencing of human pancreatic islets with in vitro and in vivo functional studies, we present novel insights into the molecular mechanisms by which impaired islet function can contribute to deregulated glucose metabolism. Coexpression analysis showed that expression of many genes correlated strongly with glucagon, not least *SLC30A8* encoding the zinc transport *ZnT8* (*SI Appendix, Fig. S84*). Rare loss-of-function variants in the *SLC30A8* gene have recently been associated with lowering of blood glucose and protection from T2D (47), but the mechanism for this glucose-lowering effect has been unclear, especially as disruption of the *SLC30A8* gene in mice has yielded the opposite phenotype: glucose intolerance (48).

The current data might thus shed some light on this paradox: the lower the expression in human pancreatic islets of *SLC30A8*, the lower the expression of glucagon. It remains to be shown whether carriers of these loss-of-function mutation carriers also show inappropriately low glucagon concentrations. Because most SNPs associated with T2D are intronic or intergenic, it has been assumed that most of them would influence expression, rather than function, of a gene. Although the nearest genes often have been suggested as targets, this has not previously been formally tested in human islets, which represent the culprit in the pathogenesis of T2D. Our current study in a large number of human islets allowed this analysis and showed enrichment of GWAS SNPs associated with T2D or glycemic traits in eQTLs (Table 1) and in genetic variants showing allelic imbalance (Table 2). Although we often assume that both parental alleles are expressed to the same degree, this was not the case for SNPs in 1,102 genes, including eight T2D-associated genes (Table 2). This could easily mask an association if the effect of the two parental alleles is bidirectional. We also found allelic imbalance to be often associated with DNA methylation. Genes found to be differentially methylated in human pancreatic islets of non-T2D versus T2D donors (49) were enriched to have allelic imbalance of expression in our dataset (Fisher exact test P value = $6.8e^{-4}$; odds ratio = 1.5). Moreover, *PEG3* was here detected to have its allelic imbalance associated with diabetic status (Fig. 4). *PEG3* encodes for a zinc finger protein that may play a role in cell proliferation and p53-mediated apoptosis (50), mechanisms that could be involved in the regulation of functional β -cell mass (51).

We also found several eQTLs and sQTLs associated with measures of β -cell function and glucose metabolism, most notably variation in the *TMED6*, *NT5E*, *PAK7*, and *TSPAN33* genes, whose disruption in INS-1 cells resulted in impaired insulin secretion (Fig. 2 B–D). These and the other identified genes with an eQTL

associated with in vitro and in vivo effects on glucose metabolism could be further explored as potential novel drug targets (*Database S1*). In addition, many eQTLs and sQTLs influenced expression of noncoding RNAs, many of which seem to target genes of importance for β -cell function. Among them, the lincRNA *LOC283177* was found to be coexpressed with key genes implicated in islet function (*PAX6*, *SYT11*, and *MADD*) (25, 39, 40), and its expression correlated with HbA1c levels and insulin exocytosis (Fig. 3). Finally, we also provide, to the best of our knowledge, the first genome-wide catalog of RNA editing events in human islets mostly related to A-to-G events, but our data also emphasize the need for validation rather than simply relying on RNA sequencing.

There are some limitations with the study we need to take into account. One caveat could be purity of human cadaver islets and differences in contribution of exocrine and endocrine tissue or different contribution of α and β cells between normoglycemic and hyperglycemic donors. We focused on whole islets, as sorting of islet cells would have limited the amount of tissue available for the different analyses. Furthermore, there is important additional information to gain from studying the microorgan islet as an entity, as shown by the expression of other pancreatic hormones and their coexpression. However, some information on cell-specific expression is available from three recent papers (16–18) on a small number of sorted β cells. As described in the *Materials and Methods*, *Database S1*, and *SI Appendix, Fig. S20*, there was no difference in purity between individuals with NGT, IGT, and T2D (Kruskal-Wallis rank sum test P value = 0.83). In addition, the contribution of exocrine and endocrine tissue did not significantly differ between diabetic and nondiabetic islets, as indicated by expression of pancreatic-specific exocrine (alpha 2 amylase) and endocrine (glucagon in alpha cells, *MAFA* in beta cells, and somatostatin in delta cells) genes (*SI Appendix, Fig. S21*). Moreover, beta cell content, as measured by FACS β/α cells ratio, was also not significantly different among NGT, IGT, and T2D (Kruskal-Wallis rank sum test P value = 0.14) (*SI Appendix, Fig. S22*). Acknowledging these limitations, only large enough numbers can outweigh the problems of heterogeneity and purity. To this end, the current study, to our knowledge, represents the largest collection of human islets published thus far. In conclusion, we provide a comprehensive catalog of novel genetic variants influencing gene expression in human pancreatic islets and metabolic phenotypes to facilitate diabetes research.

Materials and Methods

Detailed materials and methods, including all statistical analysis, are available in *SI Appendix*. Islets from 89 cadaver donors of European ancestry were provided by the Nordic Islet Transplantation Program and processed as previously described (6). Microarray analysis was performed using oligo (52) and sva (53) Bioconductor packages and processed with the standard Affymetrix protocol. Sample preparation for RNA-seq was performed using Illumina's TruSeq RNA Sample Preparation Kit. Output reads were aligned to the human reference genome (hg19) with TopHat v.2.0.2 (54), using Bowtie v.0.12.8 (55). The `dexseq_count` python script was used by counting uniquely mapped reads in each exon (56). Gene and exon expression normalizations were then performed using the TMM method (57), and further normalization was applied by adjusting the expression to gene or exon length,

respectively. A linear model adjusting for age and sex as implemented in the R Matrix eQTL package (58) was used to determine the expression of genes and exons associated with HbA1c class. Exome sequencing was performed using the Illumina exome sequencing protocols. Reads were aligned to the human genome (hg19) with BWA v.0.6.2 (59). Postalignment processing and SNP calling was done with GATK v.1.6.2 (60). Allelic imbalance of expression was analyzed by Fisher exact test to calculate the proportion of reference/alternative alleles in the exome sequencing versus RNA-seq for each sample. RNA editing sites were called on autosomes in

positions that were homozygous in the exome sequencing but heterozygous in the RNA-seq data. Genotyping was performed on the Illumina HumanOmniExpress 12v1 C chips, and all of the samples passed standard genotype QC metrics. Genotypes were imputed to 1000 Genomes data, using IMPUTE2 (61) and SHAPEIT (62). cis-eQTL and cis-sQTL associations were computed between gene expression levels (eQTL) or exon expression levels (sQTL) and all SNPs within 250 kb up- or downstream of each of these genes. We used a linear model adjusting for age and sex, as implemented in the R Matrix eQTL package (58).

1. Scully T (2012) Diabetes in numbers. *Nature* 485(7398):52–53.
2. Morris AP, 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 (2012) Large-scale association analyses provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44(9):981–990.
3. Scott RA, et al.; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44(9):991–1005.
4. Lysenko V, et al. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359(21):2220–2232.
5. Dimas AS, et al.; MAGIC Investigators (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 63(6):2158–2171.
6. Taneera J, et al. (2012) A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab* 16(1):122–134.
7. Mahdi T, et al. (2012) Secreted frizzled-related protein 4 reduces insulin secretion and is overexpressed in type 2 diabetes. *Cell Metab* 16(5):625–633.
8. Dermitzakis ET (2012) Cellular genomics for complex traits. *Nat Rev Genet* 13(3):215–220.
9. Dorrell C, et al. (2011) Transcriptomes of the major human pancreatic cell types. *Diabetologia* 54(11):2832–2844.
10. Eizirik DL, et al. (2012) The human pancreatic islet transcriptome: Expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet* 8(3):e1002552.
11. Maffei A, et al. (2004) Identification of tissue-restricted transcripts in human islets. *Endocrinology* 145(10):4513–4521.
12. Gunton JE, et al. (2005) Loss of ARNT/HIF1 β mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* 122(3):337–349.
13. Kutlu B, et al. (2009) Detailed transcriptome atlas of the pancreatic beta cell. *BMC Med Genomics* 2:3.
14. Lyttle BM, et al. (2008) Transcription factor expression in the developing human fetal endocrine pancreas. *Diabetologia* 51(7):1169–1180.
15. Marselli L, et al. (2010) Gene expression profiles of Beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS ONE* 5(7):e11499.
16. Morán I, et al. (2012) Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab* 16(4):435–448.
17. Bramswig NC, et al. (2013) Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *J Clin Invest* 123(3):1275–1284.
18. Nica AC, et al. (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res* 23(9):1554–1562.
19. Djebali S, et al. (2012) Landscape of transcription in human cells. *Nature* 489(7414):101–108.
20. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9):1509–1517.
21. Ait-Lounis A, et al. (2010) The transcription factor Rfx3 regulates beta-cell differentiation, function, and glucokinase expression. *Diabetes* 59(7):1674–1685.
22. Aston-Mourney K, et al. (2007) Increased nicotinamide nucleotide transhydrogenase levels predispose to insulin hypersecretion in a mouse strain susceptible to diabetes. *Diabetologia* 50(12):2476–2485.
23. Manning AK, et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue Human Expression Resource (MUTHER) Consortium (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 44(6):659–669.
24. Kooner JS, et al.; DIAGRAM; MuTHER (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 43(10):984–989.
25. Huyghe JR, et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 45(2):197–201.
26. Dupuis J, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42(2):105–116.
27. Cho YS, et al.; DIAGRAM Consortium; MuTHER Consortium (2012) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 44(1):67–72.
28. Soranzo N, et al.; WTCCC (2010) Common variants at 10 genomic loci influence hemoglobin A_{1c} levels via glycemic and nonglycemic pathways. *Diabetes* 59(12):3229–3239.
29. Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27(17):2325–2329.
30. Harrow J, et al. (2012) GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22(9):1760–1774.
31. Wang L, et al. (2013) CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 41(6):e74.
32. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
33. Stitzel ML, et al.; NISC Comparative Sequencing Program (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* 12(5):443–455.
34. Gaulton KJ, et al. (2010) A map of open chromatin in human pancreatic islets. *Nat Genet* 42(3):255–259.
35. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.
36. Sekine N, et al. (1994) Low lactate dehydrogenase and high mitochondrial glycerol phosphate dehydrogenase in pancreatic beta-cells. Potential role in nutrient sensing. *J Biol Chem* 269(7):4895–4902.
37. Fierabracci A, Millilo A, Locatelli F, Fruci D (2012) The putative role of endoplasmic reticulum aminopeptidases in autoimmunity: Insights from genomic-wide association studies. *Autoimmun Rev* 12(2):281–288.
38. Harashima S, et al. (2012) Sorting nexin 19 regulates the number of dense core vesicles in pancreatic β -cells. *J Diabetes Invest* 3(1):52–61.
39. Andersson SA, et al. (2012) Reduced insulin secretion correlates with decreased expression of exocytotic genes in pancreatic islets from patients with type 2 diabetes. *Mol Cell Endocrinol* 364(1–2):36–45.
40. Gosmain Y, et al. (2012) Pax6 is crucial for β -cell function, insulin biosynthesis, and glucose-induced insulin secretion. *Mol Endocrinol* 26(4):696–709.
41. Cunningham MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B (2010) Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* 6(4):e1000899.
42. Ge ZJ, et al. (2013) Maternal diabetes causes alterations of DNA methylation statuses of some imprinted genes in murine oocytes. *Biol Reprod* 88(5):117.
43. Pasquali L, et al. (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 46(2):136–143.
44. Nogueira TC, et al. (2013) GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. *PLoS Genet* 9(5):e1003532.
45. Chen L (2013) Characterization and comparison of human nuclear and cytosolic editomes. *Proc Natl Acad Sci USA* 110(29):E2741–E2747.
46. Pickrell JK, Gilad Y, Pritchard JK (2012) Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science* 335(6074):1302, author reply 1302.
47. Flannick J, et al.; Go-T2D Consortium; T2D-GENES Consortium (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 46(4):357–363.
48. Hardy AB, et al. (2012) Effects of high-fat diet feeding on Znt8-null mice: Differences between β -cell and global knockout of Znt8. *Am J Physiol Endocrinol Metab* 302(9):E1084–E1096.
49. Dayeh T, et al. (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* 10(3):e1004160.
50. Jiang X, et al. (2010) The imprinted gene PEG3 inhibits Wnt signaling and regulates glioma growth. *J Biol Chem* 285(11):8472–8480.
51. Zhou Y, et al. (2012) Survival of pancreatic beta cells is partly controlled by a TCF7L2-p53-p53INP1-dependent pathway. *Hum Mol Genet* 21(1):196–207.
52. Carvalho BS, Irizarry RA (2010) A framework for oligonucleotide microarray pre-processing. *Bioinformatics* 26(19):2363–2367.
53. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28(6):882–883.
54. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
55. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
56. Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22(10):2008–2017.
57. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11(3):R25.
58. Shabalin AA (2012) Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28(10):1353–1358.
59. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
60. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
61. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5(6):e1000529.
62. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.