Cell-type, allelic and genetic signatures in the human pancreatic beta cell transcriptome

Alexandra C Nica, Halit Ongen, Jean-Claude Irminger, et al.

Genome Res. published online May 28, 2013
Access the most recent version at doi:10.1101/gr.150706.112

Cell-type, allelic and genetic signatures in the human pancreatic beta cell transcriptome

Alexandra C. Nica[1,2,3], Halit Ongen[1,2,3], Jean-Claude Irminger[1,2], Domenico Bosco[4], Thierry Berney[4], Stylianos E. Antonarakis[1,2], Philippe A. Halban[1,2] , Emmanouil T. Dermitzakis[1,2,3]

1. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, 1211 Switzerland
2. Institute of Genetics and Genomics of Geneva (iGE3), 1211 Switzerland
3. Swiss Institute of Bioinformatics, 1 Rue Michel-Servet, 1211, Geneva, Switzerland
4. Cell Isolation and Transplant Center, Department of Surgery, Geneva University Hospitals and University of Geneva, Geneva, 1211 Switzerland

ABSTRACT

Elucidating the pathophysiology and molecular attributes of common disorders as well as developing targeted and effective treatments hinges on the study of the relevant cell type and tissues. Pancreatic beta cells within the islets of Langerhans are centrally involved in the pathogenesis of both type 1 and type 2 diabetes. Describing the differentiated state of the human beta cell has been hampered so far by technical (low resolution microarrays) and biological limitations (whole islet preparations rather than isolated beta cells). We circumvent these by deep RNA sequencing of purified beta cells from 11 individuals, presenting here the first characterization of the human beta cell transcriptome. We perform the first comparison of gene expression profiles between beta cells, whole islets and beta cell depleted islet preparations, revealing thus beta cell specific expression and splicing signatures. Further, we demonstrate that genes with consistent increased expression in beta cells have neuronal-like properties, a signal previously hypothesized. Finally, we find evidence for extensive allelic imbalance in expression and uncover genetic regulatory variants (eQTLs) active in beta cells. This first molecular blueprint of the human beta cell offers biological insight into its differentiated function including expression of key genes associated with both major types of diabetes.

INTRODUCTION

Phenotypic differences among cell-types, individuals and populations (Stranger et al. 2007; Dimas et al. 2009; Nica et al. 2011) are determined by variation in gene expression. A substantial proportion of this variability is driven by DNA polymorphisms residing in regulatory elements proximal or distal to the affected genes (Price et al. 2011; Grundberg et al. 2012). Numerous such variants have now been mapped for a variety of tissues, highlighting their tissue dependent properties and hence the acute need for expression profiling of a diverse panel of cell-types (Nica et al. 2011; Grundberg et al. 2012). This became even more evident in the context of functionally elusive results from genome-wide association studies (GWAS), as transcript abundance has been shown to provide a direct and causal link between genotype and disease susceptibility (Emilsson et al. 2008; Nica et al. 2010). This connection has been mostly attainable in disease relevant tissues, often in concordance with our present knowledge about the etiology of complex diseases (Nica et al. 2011; Grundberg et al. 2012). With the substantial improvement in the accuracy and resolution of transcriptome profiling by direct RNA sequencing (RNA-seq) (Montgomery et al. 2010; Pickrell et al. 2010), it is now possible to explore these relations comprehensively in an unbiased manner, with no theoretical limitation for dynamic range of expression detection provided sufficient sequencing depth.

Insulin-secreting pancreatic beta cells within the islets of Langerhans have been consistently involved in the pathogenesis of diabetes via autoimmune mediated apoptosis (Tisch and McDevitt 1996) (type 1) or insulin deficiency (Saltiel and Kahn 2001) (type 2). The genetic landscape of both common forms of the disease has been substantially broadened, with now over 60 known loci robustly associated with either type 1 (Barrett et al. 2009) or type 2 diabetes (Morris et al. 2012). As already attested (Gaulton et al. 2010), regulatory changes will likely explain a proportion of these associations, but uncovering them is entirely dependent on first describing the transcriptional profile of the beta cell and understanding its genetic determinants. In this context, we interrogate here the human beta cell transcriptome in multiple whole-genome sequenced individuals and uncover beta cell specific features in the context of other pancreatic endocrine cell-types.

RESULTS

Following ethical guidelines at the University Hospital in Geneva, we obtained human islets from 11 cadaveric pancreata from individuals without documented diabetes (description in

Supplementary Table 1). The islet preparations were of high purity (mean ± sd: 84.6 ± 10.3%) as measured by dithizone staining, indicating little contamination with exocrine tissue. The islet cells were sorted by FACS as previously documented (Parnaud et al. 2008) in order to obtain a highly purified population of fully functional beta cells for each individual: 86.7 ± 6.8% beta cell purity assessed by immunofluorescence staining for insulin (*INS*). We identified any contaminating alpha and delta cells by co-staining for glucagon (*GCG*) (4.4 ± 3.4%) and somatostatin (*SST*) (2.7 ± 0.8%). For five of the samples we also generated beta cell depleted "nonbeta" populations consisting primarily of glucagon secreting alpha cells (59.8 ± 14.1%), with 4.3 ± 2.6% beta cells. RNA was extracted immediately after FACS sorting from 23 cell preparations (11 beta, 7 islet and 5 nonbeta), cDNA libraries constructed (Thomas and Ansel 2010) and sequenced at very high depth.

We obtained a median of 209 million total reads (paired end, 49 bases) per sample (Supplementary Figure 1a), attaining thus an exceptional transcriptome resolution. The reads were mapped to the latest version of the human genome assembly (GRCh37/hg19) using BWA (Li and Durbin 2009) and subsequently filtered for mapping quality and correct pairing. On average, 75% of the filtered reads (median 121 million reads per sample) mapped to known exons annotated by GENCODE (Harrow et al. 2006) (version 10). We quantified all genes expressed in > 90% of individuals in either of the three cell type preparations (N = 19,975), normalizing the read counts to exonic gene length and sequencing depth (RPKM – reads per kilobase per million mapped reads (Mortazavi et al. 2008)).

Principal Component Analysis (PCA) on RPKM units (Figure 1a) indicates that beta cells and islets cluster closely together and markedly separate from nonbeta cells, with *INS*, insulin-like growth factor 2 (*IGF2*), *GCG*, transthyretin (*TTR*), regenerating islet-derived 1 alpha (*REG1A*), and *SST* driving most of this separation (Supplementary Figure 2). We notice a clear clustering of the islet derived RNA-seq data in the context of 18 other tissues (obtained from Ambion's FirstChoice® Human Total RNA Survey Panel and processed alike), with liver bearing most similarity to the islet samples. Unsurprisingly, islet purity influences gene expression (lowest purity preparation P786i clusters less well), yet this effect is not very large in our samples of overall high quality. To illustrate this, we quantified the proportion of true positives estimated from the enrichment of significant p-values (pi1 (Storey and Tibshirani 2003)) resulting after correlating each gene's RPKM with purity (pi1=0-0.2, Supplementary Figure 3).

Overall, we observe a high ranked correlation between beta cell and islet-expressed genes (rho=0.94, Figure 1b) and we estimate that 87% of the variance in beta cell gene expression can be explained by using islet expression as proxy. Given the estimated ~75% beta cell composition of the human islet (Pisania et al. 2010) and the quality of the material used for this study, the similarity between these two cell-types is not surprising. As expected, *INS* was by far the most abundantly transcribed gene, followed by *INS-IGF2* and *IGF2*, making up ~ 38%, 10% and 2% of the total nuclear beta cell transcriptome respectively (Figure 2a). The corresponding relative percentages are lower in the islet, by approximately two-fold (*INS* 21%, *INS-IGF2* 5.8%, *IGF2* 1.1%).

To uncover transcriptional signatures specific to beta cells, we next tested for differential gene expression after appropriate statistical modelling of the raw counts (DESeq (Anders and Huber 2010)). When comparing the beta cell sequence count data to 18 non-islet tissues, we discovered 2980 differentially expressed (DE) genes (~12% of genes tested) at 10% false discovery rate (FDR), 417 of which were significantly over-expressed in beta cells. These genes underlie the general features of endocrine pancreatic activity (DAVID (Huang da et al. 2009) enrichment

results - Supplementary Table 2). We next sought to identify beta cell specific genes only in the context of the islet and its other hormone-producing cell-types, while controlling for islet purity (included as covariate). Of the 19,975 genes tested, we found 5555 DE genes between the beta cell and islet preparations and 4380 DE genes when comparing beta and nonbeta cells (Table 1a). The difference in power is due to the smaller sample size of the nonbeta population and its greater variance in composition of cell-types. We find in both cases a larger proportion (~2/3) of over-expressed genes in the more heterogeneous of the cell populations compared (log_2FoldChange > 0, Figure 2 b,c). The remaining one third of the DE genes are significantly over-expressed in beta cells, with smaller fold changes in islets as expected (log_2FoldChange mean ± sd: -1.06 ± 0.37 islet/beta, -1.64 ± 0.87 nonbeta/beta). Notably, we observe a significant enrichment of annotated lincRNAs in beta cell over-expressed genes, corroborating their cell-type specific expression properties reported previously in other tissues (Cabili et al. 2011): we detect 132 overexpressed lincRNAs in beta vs islet (1.59 fold enrichment over protein coding genes, Fisher's p-value: 1.1E-04) and 148 overexpressed lincRNAs in beta vs nonbeta (1.9 fold enrichment, Fisher's p-value:1.02E-07). These are likely underestimates as we limited ourselves to noncoding RNAs currently present in the annotation (Moran et al. 2012).

To portray the unique biological characteristics of beta cells in the islet context, we defined a stringent "beta cell specific" set of genes (N=526) as the intersection of the beta cell over-expressed genes from the two pairwise comparisons (1987 beta-islet and 1583 beta-nonbeta) ranking in the following order of expression: beta>islet>nonbeta. Similarly, we defined a set of "nonbeta cell specific" genes (N=614) consistently and significantly under-expressed in beta cells (nonbeta>islet>beta). A functional annotation analysis reveals that beta cell specific genes have neuron-like properties (Table 2), a similarity noted earlier in studies implicating glucose-sensing hypothalamic neurons in nutrient homeostasis (Schwartz et al. 2000) or pancreatic hormone release (Ahren 2000). Nonbeta cell specific genes are mostly enriched for cell surface components (N-glycoproteins) central to both intercellular and cell-environment communication (Danzer et al. 2012), secretory proteins (signal peptide) and extracellular matrix components.

The two gene sets ("beta" and "nonbeta") offer a more accurate depiction of the transcripts that define the molecular identity of the major islet cell-types: Table 3 shows the top 30 beta and nonbeta cell enriched transcripts respectively, in descending order of RPKM and fold change (RPKM>1 cutoff was used). This approach successfully filters out highly expressed genes in contaminating cell-types (e.g. *SST, GCG* from somatostatin and glucagon cells contaminating the beta cell population), otherwise mistaken as key players in the expression signature of beta cells. In addition to known beta cell specific transcripts (*INS*, *IGF2*, *PDX1*) we highlight further targets, some featured already in a microarray analysis of sorted islet cells (Dorrell et al. 2011b) (e.g. *RGS16* - negative regulator of G-protein signaling, involved in endocrine pancreas development and re-expressed in adult cells in response to GLP-1 (Villasenor et al. 2010); *ADCYAP1* – pituitary adenylate cyclase activating polypeptide 1, involved in insulin secretion and beta cell regeneration/proliferation (Sakurai et al. 2011); *HADH* - hydroxyacyl-CoA dehydrogenase, negative regulator of insulin secretion (Hardy et al. 2007) associated with Alzheimer's (Nicolls et al. 2003) which is in turn associated with diabetes. Many other genes however have not been described before in the context of beta cells, including: *NPTX2* - neuronal pentraxin 2, found in neuronal cells and gliomas but also shown to be frequently down-regulated in pancreatic cancers (Zhang et al. 2012); *TSPAN1* - tetraspanin 1, which can associate with alpha6.beta1 integrin and promote FAK phosphorylation (Huang et al. 2008) shown by us to be involved in insulin secretion (Rondas et al. 2011); *GPM6A* - neuronal membrane glycoprotein of unknown function but identified as beta cell marker in sorted mouse islet cells (Dorrell et al. 2011a); *BMP5* - bone morphogenic protein 5, implicated in pancreas and fetal beta cell development (Jiang et al. 2002);

P2RY1 - purinergic receptor through which ADP and ATP modulate insulin secretion (Fernandez-Alvarez et al. 2001) etc.

Alternative splicing (AS), a common feature of most (~94%) eukaryotic genes contributing to tissue specificity (Pan et al. 2008) is also significantly enriched in beta cell specific genes (N=200 genes, 1.22 fold enrichment). To assess the extent of AS between islet cell-types, we quantified all the exon links (reads spanning pairs of exons only) (Ongen 2012) in the beta, islet and nonbeta preparations and tested for differential expression those exon pairs making a minimum of 5 links in at least 90% of the samples (N=154,190). We find 28,183 DE links (10% FDR) between beta and islets and 15,926 DE links between beta and nonbeta (Table 1b), corresponding to 5072 and 3025 genes respectively. In 2167 and 998 of these genes respectively, the DE links observed between the different islet cell-types were not due to DE of the underlying genes. This implicates 8-18% of the 12,334 genes thus tested as candidates for islet AS.

Mapping RNA-seq data to the genome fails to identify reads that span exon junctions. To quantify this effect in our data and discover potential unannotated transcripts, we constructed all possible 96-mer exon junctions for each gene (read length-1=48 bases on either side of the junction) and mapped all reads against them. On average, 2.9 million reads per beta cell sample did not map to the genome but were recovered by junction mapping (Supplementary Table 3), a small percentage (~1.36%) of the total number of reads. Nevertheless, we used these together with reads mapping better to exon junctions than the genome (better mapping quality and larger alignment length) to assess the percentage of previously unobserved junctions. We identified 46,096 junctions in 7717 genes covered by at least 10 good quality reads (map quality > 10) in all beta cell samples, 657 of which (1.4% of all junctions) were not present in the transcriptome annotation (GENCODE v10). These correspond to 465 genes enriched for posttranslational functions (acetylation: 8.9E-27, phosphoprotein: 1.3E-17, ribosome: 1.6E-14) and possibly containing new beta cell specific transcripts.

We next aimed to study the genetic control of the variation in beta cell and islet gene expression by integrating genome sequence level information when available. We extracted DNA from 7 of the donors and sequenced them to medium (16x) coverage (Supplementary Figure 1b). The reads were aligned to the hg19 reference genome with BWA, we applied GATK (McKenna et al. 2010) base quality score recalibration, indel realignment and duplicate removal followed by SNP discovery and genotyping across all 7 samples simultaneously using variant quality score recalibration (DePristo et al. 2011). Subsequently, we imputed the variants on the 1000 Genomes reference panel and phased them with Beagle 3.3.2 (Browning and Browning 2007), resulting in 5,748,462 good confidence autosomal SNPs for analysis. For each individual, we tested the subset of heterozygous SNPs with good coverage in the RNA-seq data (beta, islet and nonbeta) for allele specific expression (ASE) (median 23,358 sites per sample). On average, we observe that 9.3% of the heterozygous sites tested show significant ASE at 10% FDR (median 2626 sites per sample, Supplementary Figure 4). These correspond to a median of 1742 genes out of 6756 genes tested per individual, harboring at least one ASE SNP (equivalent to 59.18% of the total number of tested genes in all samples). A subset of these genes are of particular functional interest, having been linked to diabetes in GWAS: 15 of 23 T1D genes tested show ASE, 20 of 28 T2D genes and 15 of 18 genes associated with fasting glucose or insulin levels. Except for the subset of genes associated to fasting glucose or insulin (Fisher's p-value 0.02), this enrichment was not statistically significant. A substantial number of diabetes susceptibility genes are however expressed in beta cells, corroborating the GWAS predictions (32/40 T1D, 37/39 T2D, 19/20 glucose or insulin levels - Supplementary Figure 5). Several of these (N=9) are in fact beta cell specific transcripts: T1D-associated *INS*, imprinted *MEG3-DLK1*, *GLIS3*; T2D-associated

VEGFA, *SLC30A8*; fasting glucose/insulin levels-associated *ADRA2A*, *G6PC2*, *SLC2A2*. Interestingly, two other genes, *SH2B3* (T1D-linked) and *IRS1* (T2D-linked) are significantly over-expressed in the nonbeta population – given our data, it is tempting to speculate that alpha cell misregulation would be detrimental in these cases. In some instances we find evidence of significant allelic imbalance for the exact GWAS SNPs reported e.g. rs5215 (*KCNJ11*), a missense SNP associated to T2D risk is also an ASE variant in 4 beta cell samples, rs689 (*INS)*, associated to T1D autoantibodies is an ASE SNP in 4 samples and rs11558471 (*SLC30A8*) associated with fasting glucose-related traits and proinsulin levels is an ASE SNP in 1 sample (the rest of the samples were homozygous for the respective SNPs hence could not be tested for ASE). These observations suggest that a subset of diabetes-associated loci could contribute to disease progression via a change in beta cell gene expression.

As expected for true positive ASE sites, the direction of effect is almost always consistent for the same individual between different tissues (Figure 3). The high overall concordance between beta cell and islet expression data noted before is recapitulated at the sequence level. Taking difference of power into account (using pi1 to measure sharing of statistical significance), we estimate a median 89.11% enrichment of significant beta cell ASE p-values in the islet and a 88.11% median enrichment vice versa (significant islet ASE in the beta cell). One of the explanations for the ASE events discovered could be the presence of nearby regulatory variants (expression quantitative trait loci - eQTLs). The current sample size is prohibitively small for eQTL discovery; however, in an effort to find those potentially active in beta cells, we integrated ASE sites with eQTLs discovered in the best-powered multi-tissue dataset to date (MuTHER (Grundberg et al. 2012)), which reported 3148 eQTLs in fat, 3953 eQTLs in LCLs and 2515 eQTLs in skin. For each gene and tissue we phased the top eQTL SNP and top ASE SNP per individual, being thus able to test in our data 1806 of the eQTLs discovered in fat, 2272 eQTLs discovered in LCLs and 1400 eQTLs discovered in skin. We observe a significantly greater deviation from the expected reference/total allelic ratio (0.5) in individuals heterozygous for the eQTLs compared to homozygotes (Mann-Whitney p-value < 6.18E-11), indicating that a proportion of these regulatory variants are also active in beta cells (Figure 4). To shortlist them, we selected cases where the direction of the eQTL effect was in agreement with the significant (ASE p-value<0.01) allelic imbalance prediction. We thus report 254 candidate beta eQTL genes using the data from fat cells, 294 genes from LCLs and 198 from skin. Reassuringly, 303 of the 510 total genes detected as such (59.4%) are shared in at least two of the MuTHER tissues, a significant enrichment (Fisher's p-value: 6.2E-10) over the 44.8% eQTL sharing we began with. This suggests that these eQTLs are not highly tissue specific and therefore likely to be also present in beta cells.

DISCUSSION

Pancreatic islets are essential for regulating and maintaining glucose homeostasis. This functional role is fulfilled by at least five distinct hormone-producing cell-types, which act differently but synergistically to help maintain appropriate glycemic levels in healthy individuals. Therefore, uncovering the molecular identity of ideally each of these cell-types would ensure a better understanding of the mechanisms through which they become defective in diabetic patients. We present here the first ever unbiased transcriptome analysis of the differentiated human beta cell, the most abundant pancreatic islet cell-type that secretes insulin and is severely affected in common forms of diabetes. The advanced cell sorting protocol, immediate RNA extraction and unprecedented sequencing resolution makes this unique dataset the most faithful representation of the human beta cell transcriptome to date. Importantly, we compare this to the expression profile of whole islets and beta cell depleted islet preparations ("nonbeta"), uncovering unique beta cell specific features in contrast to other pancreatic endocrine cell-types. In the absence of the three-

way beta-islet-nonbeta transcriptome comparison enabled by our study design, such discoveries would remain concealed.

Overall, we observe that islets are a good proxy for beta cell gene expression, provided they are of high enough purity (>85%). However, important biological insights into cell-type specific expression signatures are revealed when comparing detailed RNA profiles of purified beta cells with those of whole islets or “nonbeta” cells. We find substantial evidence for differential expression between beta cells and islets, with >5000 genes showing significant changes in transcript abundance between the two. Genes consistently over-expressed in beta cells are enriched for neuron-like properties, a similarity suspected earlier which we now independently confirm. These important biological insights would not be revealed by islet expression studies alone. Further, we corroborate current GWAS results by reporting beta cell expression of a large number of genes (80-95%) previously associated with type 1 or type 2 diabetes. Differential splicing is also present among islet cell-types, with up to 18% of genes undergoing this process in our data. Altogether, this draws attention to the general limitations of analyzing a mixture of cell populations as opposed to preparations highly enriched for a single cell-type of interest.

Finally, we describe here the first set of genetic variants controlling inter-individual variability in beta cell gene expression. Despite the small sample size and consequent limited number of informative genetic variants, we are able to report allelic and genotypic expression differences, offering a first glance at the genetics behind human beta cell function. Our evidence for diabetes-associated loci expressed in an allele specific manner flags potential contributions to disease effects by genetically or epigenetically driven expression changes in beta cells. These are encouraging results, which remain to be refined and augmented by a much-awaited beta cell eQTL study.

Taken together, the data will serve the diabetes research community manifold ways, including towards the validation of future stem-cell derived or other surrogate beta cells for research purposes or regenerative medicine.

METHODS

Sample collection
Islets isolated from cadaveric pancreas were provided by the Cell Isolation and Transplant Center, Department of Surgery, Geneva University Hospital (Drs. T. Berney and D. Bosco) through the Juvenile Diabetes Research Foundation (JDRF) award 31-2008-416 (ECIT Islet for Basic Research Program).

Beta cell sorting
Islets were dispersed into single cells, stained with Newport Green and sorted into “beta” and “nonbeta” populations as described previously (Parnaud et al. 2008). The proportion of beta (insulin), alpha (glucagon) and delta (somatostatin) cells in each population (as percentage of total cells) was determined by immunofluorescence.

RNA extraction
Sorted beta cells, nonbeta cells and islets were centrifuged, the supernatant was removed and the pellet disrupted in RLT buffer (RNeasy, Qiagen). Total RNA was prepared according to the standard RNeasy protocol.

DNA extraction
DNA was prepared according to the standard DNeasy (Qiagen) protocol.

Library preparation and sequencing
Total RNA and genomic DNA libraries were constructed following customary Illumina TruSeq protocols for next-generation sequencing. PolyA-selected mRNAs were purified, size-fractioned and subsequently converted to single-stranged cDNA by random hexamer priming. Following second-strand synthesis, double-stranded cDNAs were blunt-end fragmented and indexed using adapter ligation after which they were amplified and sequenced according to protocol. RNA libraries were 49-bp paired end sequenced with one or a maximum of two samples per HiSeq 2000 lane. DNA libraries for whole-genome sequencing were constructed similarly, but starting directly from fragmented genomic DNA. The 7 DNA libraries were pooled and sequenced in a total of 9 HiSeq lanes each (1 control lane of 49-bp paired end reads and 8 lanes of 95-bp paired end sequencing). Standard quality checks for material degradation (Bioanalyzer) and concentration (Qubit) were done before and after library construction, ensuring that samples are suitable for sequencing.

Read mapping
Paired-end reads were mapped to the human genome (assembly version GRCh37/hg19) using BWA and allowing a maximum insert size of 500 kb (for instances when not enough good alignments could be used to infer the insert size distribution). RNA-seq reads were subsequently filtered for correct orientation of the mapped mate pairs with an insert size < 500 kb and a minimum mapping quality score of 10.

Expression quantification
GENCODE annotation v10 was used to assign filtered reads to their corresponding exons and genes. We filtered the annotation to include transcribed biotypes (coding and non-coding) i.e. all protein coding genes, lincRNAs, processed transcripts, non-coding, polymorphic pseudogenes, transcribed processed pseudogenes and transcribed unprocessed pseudogenes. For each gene, we processed the exons from all transcripts which we quantified by taking into account only filtered reads as above where both mates of a pair map to exons of the same gene. The gene counts were

incremented non-redundantly i.e. reads overlapping two exons are counted once to the total count sum per gene. Resulting raw gene counts were normalized to gene length (sum of exons) and sequencing depth i.e. RPKM – reads per kilobase per million mapped reads.

Differential expression analysis
Raw gene/exon link counts were modeled using a negative binomial distribution and tested for differential gene expression with the DESeq R package. Following author's guidelines, we estimated the size factor for each sample from the count data, followed by dispersion estimations for each gene/link. Subsequently, we tested for differential expression between conditions. We grouped all beta cell, islet, nonbeta and the 18 different tissues (adipose, bladder, brain, cervix, colon, esophagus, heart, kidney, liver, lung, ovary, placenta, prostate, spleen, testes, thymus, thyroid, trachea) in separate conditions (beta, islet, nonbeta, other) and tested for differential expression: beta vs other, beta vs islet and beta vs nonbeta. For the pair-wise DESeq tests of islet-derived data only, we included islet purity as a covariate: we fitted two generalized linear models and compared them to see whether the purity factor improved the fit (i.e. had a significant effect). The p-values of the comparisons were then adjusted for multiple testing using the Benjamini-Hochberg method. Adjusted p-values< 0.1 were considered significant. Introducing the islet purity as a covariate improved the fit in all cases.

Quantification of exon links
We made use of the paired end nature of our RNA-seq experiment to relatively quantify splicing events. Specifically, we used filtered mate pairs (correctly paired and 10 minimum mapping quality) where one mate maps to one exon and the other mate to a different exon to count "links" between two exons. The first exon in a link was referred to as the "primary exon". Overlapping exons were grouped into "exon groups" and unique portions of each exon in an exon group were identified. These unique portions were subsequently used to assign reads to an exon.

Generation of exon junctions
We have constructed all possible 96-mer junctions (flanking sequence on either side = read length - 1) per gene using the transcriptome information in Ensembl v65. For each gene, we first kept track of all the unique start and end positions of all its exons. Subsequently, for non-redundant exon-exon pairs with valid positions (i.e. starting position of second exon greater than ending position of first exon in junction), we constructed every possible pair-wise combination within the length limits of the exons involved. For junctions failing this criterion due to short exons on one or both ends the sequence was expanded into flanking exons. For every valid junction, we recorded whether it forms part of any known transcript in or not.

Functional annotation analysis
Gene functional analyses were performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7. We have used the Functional Annotation Chart tool to analyze our genes of interest (e.g. beta cell specific) in the context of the whole human genome. DAVID's default annotation categories were used and enrichment results sorted by significance of Benjamini corrected p-values.

SNP calling
Variant calling was performed with the Genome Analysis Toolkit (GATK) 1.5.31 following the Best Practice Variant Detection v3. Reads were aligned to the hg19 reference genome with BWA and bam files from each lane were merged into one bam per sample. For each sample we removed duplicates, realigned the reads around known indels from the 1000 Genomes, applied GATK base quality score recalibration followed then by SNP discovery and genotyping across all 7 samples

simultaneously with variant quality score recalibration. We used a confidence score threshold of 30 for variant detection and a minimum base quality of 17 to consider a base for calling. Good confidence (1% FDR) SNP calls were then imputed on the 1000 Genomes reference panel and phased with Beagle 3.3.2.

ASE analysis

Allele specific expression (ASE) analysis was performed per-individual on the subset of heterozygous sites with a minimum coverage of 16 reads per site filtered for mapping quality of 10. Problematic positions were filtered out as follows to exclude sites susceptible to allelic mapping bias: 1) sites with UCSC 50-bp mapability <1, implying that the 50bp flanking region of the site is non-unique in the genome and collapsed repeat regions were excluded; 2) simulated RNA-seq reads overlapping the site and showing >5% difference in the mapping of reads that carry the reference or non-reference allele were also excluded. Next, we calculated the expected reference allele ratio for each individual by summing up reads across all sites separately for each SNP allele combination after down-sampling reads of sites in the top 25th coverage percentile in order to avoid the highest covered sites having a disproportionally large effect on the ratios. These expected REF/TOTAL ratios (in the range of 0.489-0.518) correct for remaining slight genome-wide mapping bias in each individual. In a binomial test of the REF/NONREF allele counts, we used these individual-specific expected ratios to weight the occurrence of each allele and determine sites with ≥ 16 reads in each individual deviating from the biased-corrected expected allelic ratios. We called sites in significant ASE if resulting p-values < 0.01.

eQTL integration

We obtained *cis* eQTLs (1% FDR) detected in the MuTHER study, where 856 female twins were profiled for gene expression with microarrays in three tissues: fat, lymphoblastoid cell lines (LCLs) and skin. These amounted to 3148 genes in fat, 3953 genes in LCLs and 2515 genes in skin having *cis* eQTLs. We phased the top SNPs per gene in each of these datasets with top beta cell ASE sites for the same genes. We contrasted the beta cell allelic ratios for individuals heterozygous for the MuTHER eQTLs compared to homozygotes, expecting smaller effects in homozygotes where the eQTL itself (if active also in beta cells) would have no effect on expression.

DATA ACCESS

Genotype and sequence data has been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under accession number EGAS00001000442. Anonymized sequence files (matching the reference genome) are freely and publicly available on our ftp server (ftp://jungle.unige.ch/) as well as Ensembl.

ACKNOWLEDGEMENTS

This work was supported by grant number 17-2011-284 from JDRF and the Louis-Jeantet Foundation. We thank Stephane Dupuis, Ismael Padioleau, Alexandra Planchon, Luciana Romano and Alisa Yurovsky for expert technical assistance.

FIGURE LEGENDS

Figure 1. High overall similarity between beta cell and islet gene expression.
a) PCA analysis of gene RPKMs for beta (N=11), islet (N=7), nonbeta (N=5) preparations and 18 other tissues from unrelated individuals. Beta cell and islet samples cluster together, separating from nonbetas. The other tissues cluster separately, with liver being the most similar to the islet-derived RNA-seq data b) Scatterplot of beta cell versus islet median RPKMs on $\log_{10}$ scale

Figure 2. Expression differences between beta, islet and nonbeta samples.
a) Dot chart of top 10 highest expressed genes and their contribution to the nuclear transcriptome by cell-type b) Histogram of $\log_2$FoldChange (islet/beta) for differentially expressed genes (10% FDR) in islets and beta cells c) Histogram of $\log_2$FoldChange (nonbeta/beta) for differentially expressed genes (10% FDR) in nonbeta and beta cells.

Figure 3. Sharing of significant ASE effects between beta cells and islets (sample P775). Left panel histograms show the enrichment (pi1) of significant ASE p-values in the islets for ASE sites discovered in beta cells and vice-versa (beta cell ASE p-values of ASE sites discovered in islets). Right panel scatterplots display the direction of ASE effects (ratio of reference allele count to the total number of reads covering that site) between the two cell-types, almost always in concordance.

Figure 4. Candidate beta cell *cis* eQTLs discovered initially in other tissues (fat, LCL, skin). Top panel boxplots show the deviations of allelic ratios (reference/total) from the expected 0.5 in ASE individuals grouped by eQTL genotype, with heterozygotes having markedly higher effects on ASE ratios compared to homozygotes. Bottom scatterplots show the beta coefficients of the MuTHER eQTLs and the corresponding beta ASE ratios for the selected candidate beta cell regulatory variants.

Figure 1

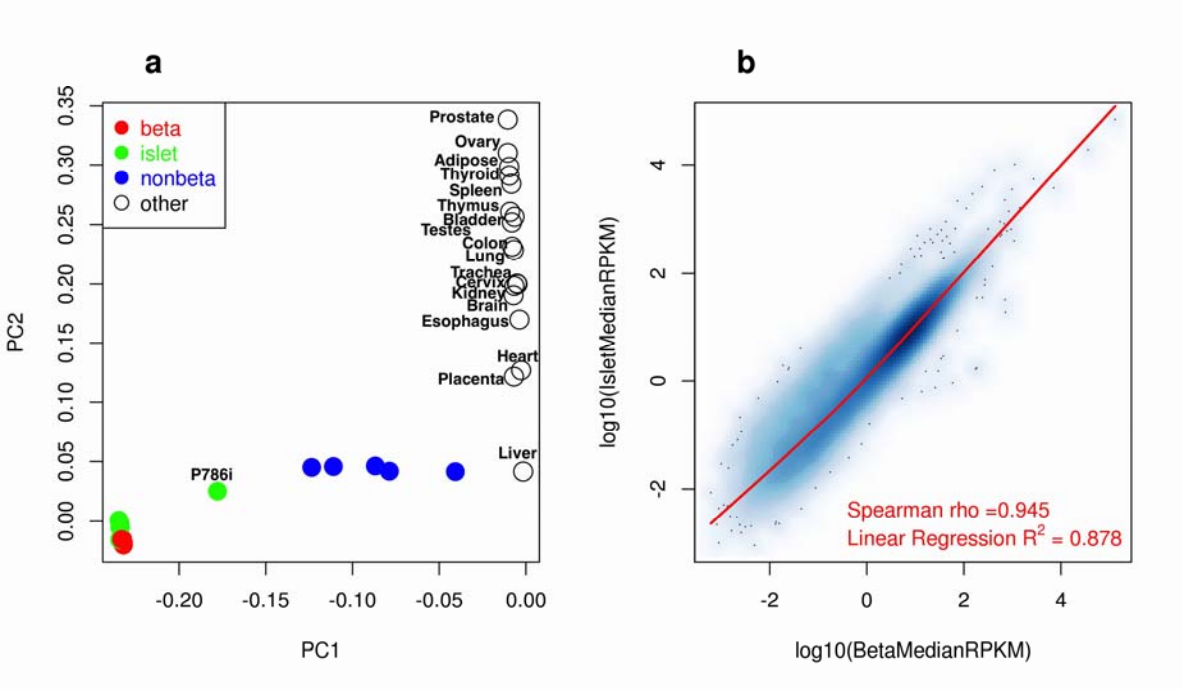


Figure 2

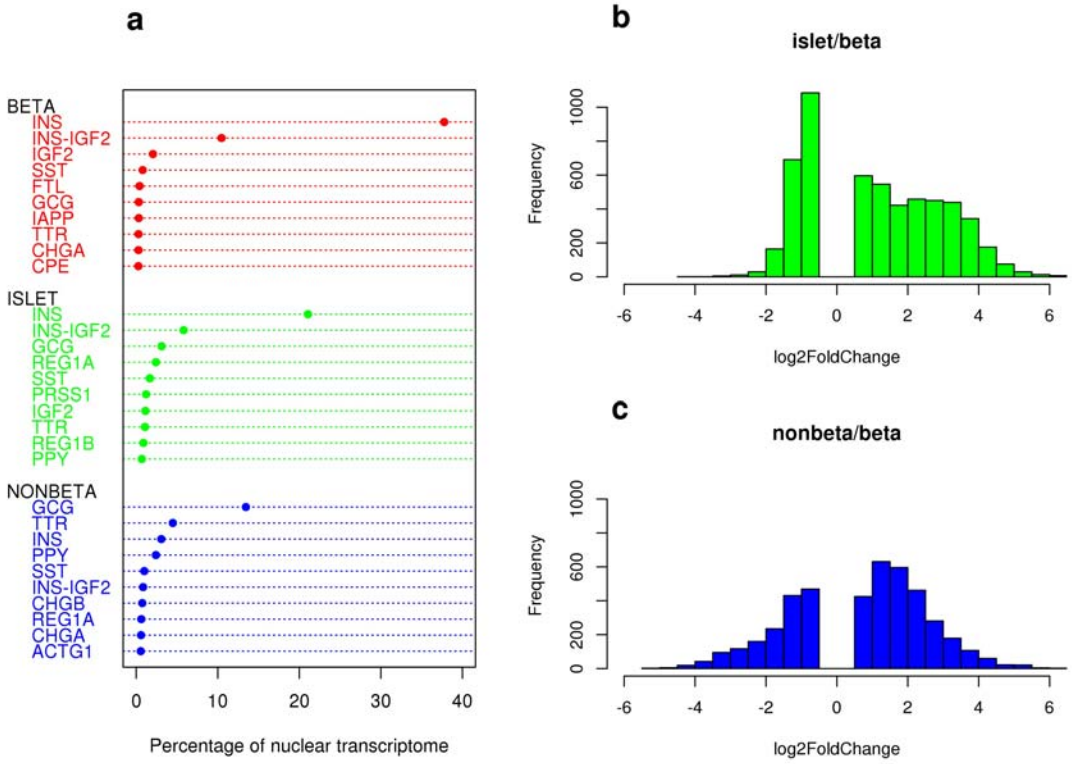


Figure 3

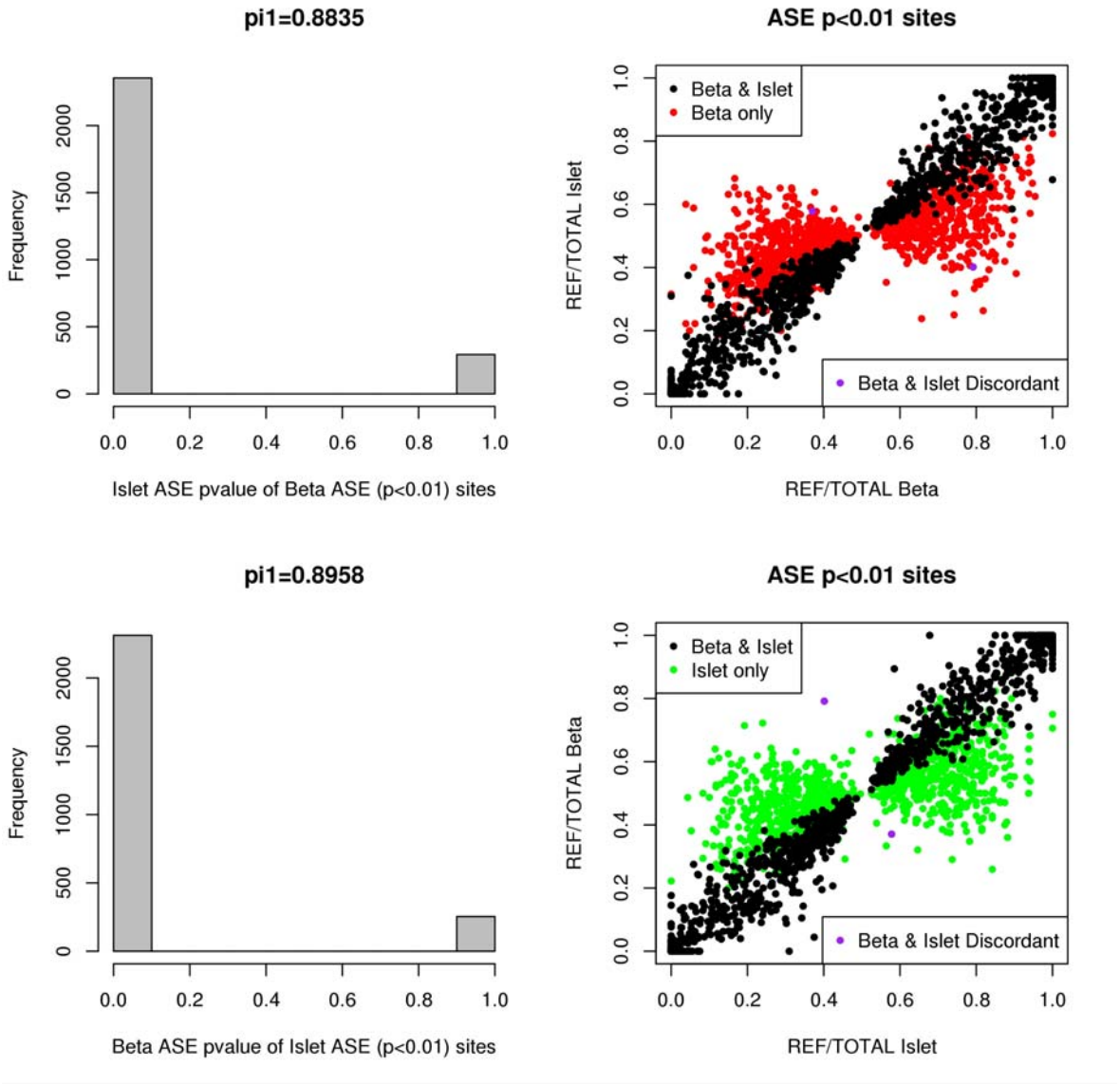


Figure 4

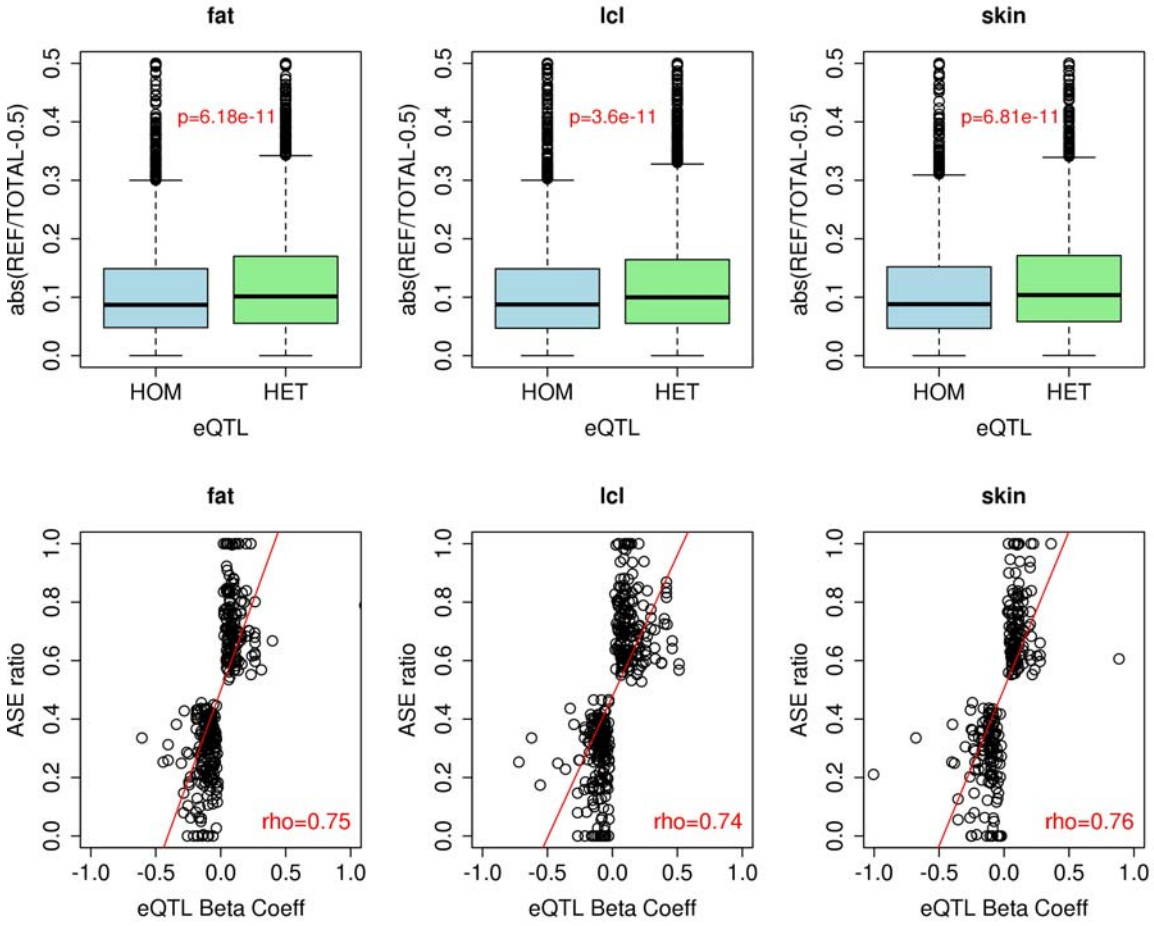


Table 1. Differentially expressed (DE) genes and exon links between beta cells, islets and nonbetas.

Pairwise DESeq results (beta versus islet, beta versus nonbeta) of significant changes in magnitude of expression of a) Genes b) Exon links (reads spanning exon junctions)

a) **Differentially expressed genes (10% FDR)** **19,975 genes tested**			
	Beta over-expressed	**Beta under-expressed**	**Total**
Islet	1987	3568	5555
Nonbeta	1583	2797	4380
b) **Differentially expressed exon links (10% FDR), corresponding gene counts and subset of candidate alternatively spliced genes (genes not DE but having DE links)** **12,334 genes tested**			
	Beta over-expressed	**Beta under-expressed**	**Total**
Islet			
links	11843	16340	28183
gene count	2130	2964	5072
gene count excluding DE genes			2167
Nonbeta			
links	5688	10238	15926
gene count	1199	1844	3025
gene count excluding DE genes			998

Table 2. Functional characteristics of beta cell specific and non-specific genes in islet context.
Top 20 most significantly over represented functional annotation terms (DAVID).

Beta cell specific genes in islet context (N=526)				**Beta cell non-specific genes in islet context (N=614)**			
Term	**Benjamini p-value**	**Count**	**Fold Enrichment**	**Term**	**Benjamini p-value**	**Count**	**Fold Enrichment**
neuron projection	1.13E-05	26	3.53	glycoprotein	5.26E-14	207	1.70
axon	1.35E-05	18	5.26	signal	5.26E-14	169	1.84
cell projection	5.33E-05	37	2.47	signal peptide	7.64E-13	169	1.83
synapse	5.79E-05	25	3.27	glycosylation site:N-linked (GlcNAc...)	2.58E-11	194	1.65
synapse part	7.50E-05	20	3.78	Secreted	5.77E-10	99	2.07
passive transmembrane transporter activity	1.77E-04	27	3.02	disulfide bond	1.50E-08	140	1.69
substrate specific channel activity	2.26E-04	26	3.02	extracellular region part	1.18E-06	67	2.14
intrinsic to membrane	2.29E-04	155	1.31	extracellular region	2.42E-06	110	1.68
ion channel activity	2.59E-04	26	3.11	plasma	4.64E-06	16	6.08
metal ion transmembrane transporter activity	2.68E-04	23	3.24	response to wounding	7.07E-05	42	2.62
integral to membrane	3.22E-04	150	1.32	membrane	8.91E-05	230	1.30
channel activity	5.08E-04	27	3.03	extracellular matrix	9.77E-05	23	3.37
ionic channel	7.98E-04	23	3.30	extracellular space	2.30E-04	47	2.11
gated channel activity	1.15E-03	21	3.13	plasma membrane part	2.51E-04	109	1.52
calcium ion binding	1.18E-03	41	2.06	growth factor	2.65E-04	16	4.32
ligand-gated channel activity	1.26E-03	13	4.73	transmembrane	5.07E-04	186	1.32
ligand-gated ion channel activity	1.26E-03	13	4.73	polymorphism	6.75E-04	374	1.14
cation channel	1.74E-03	19	3.19	proteinaceous extracellular	1.06E-03	27	2.59

activity				matrix			
ion transport	2.19E-03	31	2.44	extracellular matrix	1.19E-03	28	2.49
glycoprotein	2.35E-03	131	1.38	cell fraction	1.22E-03	61	1.73

Table 3. Top 30 genes enriched in the beta cell population and nonbeta cell population ordered by RPKM and fold change respectively.

Table contains genes with RPKM>1 consistently and significantly over-expressed in beta cells (order of expression: beta>islet>nonbeta) and nonbeta cells respectively (order of expression: nonbeta>islet>beta) ordered in descending order of RPKM and fold change.

Top 30 beta cell enriched genes				Top 30 nonbeta cell enriched genes			
Ordered by RPKM		Ordered by Fold Change		Ordered by RPKM		Ordered by Fold Change	
Gene	Beta RPKM	Gene	Fold Change Beta/Nonbeta	Gene	Nonbeta RPKM	Gene	Fold Change Nonbeta/Beta
INS	131369.34	INS-IGF2	20.17	GCG	37234.91	IRX2	23.79
INS-IGF2	36383.24	IGF2	19.99	TTR	12418.51	GCG	22.33
IGF2	7121.83	INS	18.88	CLU	895.89	LOXL4	18.31
SLC30A8	393.03	CAPN13	16.71	GC	415.26	GC	16.89
G6PC2	292.68	WSCD2	14.84	SERPINA1	410.05	FAP	16.55
DLK1	276.55	NPTX2	13.98	MUC13	220.59	PAPPA2	15.66
ERO1LB	272.02	RGS16	13.07	PTP4A3	176.18	ARX	13.89
PCSK1	207.63	ADCYAP1	12.18	TM4SF4	171.99	MUC13	13.57
NPTX2	200.05	HADH	11.50	TMEM176A	144.97	F10	13.07
SCD	166.61	TSPAN1	11.19	EZR	124.04	VSTM2L	12.88
VEGFA	164.97	ASB9	11.12	TPM4	116.04	FEV	12.75
HADH	163.95	DLK1	10.88	CRYBA2	113.63	IGFBP2	12.20
EIF4A2	157.07	GPM6A	10.84	IGFBP2	110.25	MYO10	11.94
TIMP2	145.81	GLP1R	10.48	COTL1	95.74	KCNJ3	11.50
MEG3	140.99	BMP5	10.15	LOXL4	94.37	APOH	11.44
UCHL1	136.25	P2RY1	9.99	TMEM176B	81.95	TM4SF4	9.80
SYT13	122.22	ENTPD3	9.96	F10	79.35	KCTD12	9.78
PDX1	119.25	MEG3	9.28	KCTD12	77.78	TTR	9.62
ADCYAP1	99.62	CYYR1	8.65	APOH	73.74	SEPT9	9.21
GAD2	96.59	VAT1L	8.53	LSR	72.98	NPNT	9.01
ALCAM	95.79	C8orf47	8.38	C19orf77	67.72	FOSL2	8.61
NKX6-1	91.78	PDX1	8.27	FOSL2	64.69	MAMLD1	8.31
GJD2	87.30	SLC17A6	8.12	LMNA	58.09	SERPINA1	7.93
TSC22D1	86.05	O3FAR1	8.07	IRX2	56.92	SMOC1	7.69
CASR	79.54	SCD5	8.05	PAPPA2	51.76	FXYD5	7.59
SLC6A6	77.41	OLFM1	8.05	ARX	50.97	NEDD9	7.49
RP11-713B9.1	70.46	SLC35D3	7.95	VWA1	50.20	KBTBD10	6.96
BTG3	64.76	FAM115C	7.90	SMOC1	40.44	AHNAK	6.60
CABP7	59.66	TMEM150C	7.89	RAP1GAP	40.29	LBH	6.36
SORL1	56.36	CASR	7.78	KBTBD10	39.63	DUSP4	6.20

REFERENCES

Ahren B. 2000. Autonomic regulation of islet hormone secretion--implications for health and disease. *Diabetologia* **43**: 393-410.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C et al. 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* **41**: 703-707.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**: 1084-1097.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**: 1915-1927.

Danzer C, Eckhardt K, Schmidt A, Fankhauser N, Ribrioux S, Wollscheid B, Muller L, Schiess R, Zullig R, Lehmann R et al. 2012. Comprehensive description of the N-glycoproteome of mouse pancreatic beta-cells and human islets. *J Proteome Res* **11**: 1598-1608.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**: 491-498.

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**: 1246-1250.

Dorrell C, Grompe MT, Pan FC, Zhong Y, Canaday PS, Shultz LD, Greiner DL, Wright CV, Streeter PR, Grompe M. 2011a. Isolation of mouse pancreatic alpha, beta, duct and acinar populations with cell surface markers. *Mol Cell Endocrinol* **339**: 144-150.

Dorrell C, Schug J, Lin CF, Canaday PS, Fox AJ, Smirnova O, Bonnah R, Streeter PR, Stoeckert CJ, Jr., Kaestner KH et al. 2011b. Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54**: 2832-2844.

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S et al. 2008. Genetics of gene expression and its effect on disease. *Nature* **452**: 423-428.

Fernandez-Alvarez J, Hillaire-Buys D, Loubatieres-Mariani MM, Gomis R, Petit P. 2001. P2 receptor agonists stimulate insulin release from human pancreatic islets. *Pancreas* **22**: 69-71.

Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D et al. 2010. A map of open chromatin in human pancreatic islets. *Nat Genet* **42**: 255-259.

Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A et al. 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**: 1084-1089.

Hardy OT, Hohmeier HE, Becker TC, Manduchi E, Doliba NM, Gupta RK, White P, Stoeckert CJ, Jr., Matschinsky FM, Newgard CB et al. 2007. Functional genomics of the beta-cell: short-chain 3-hydroxyacyl-coenzyme A dehydrogenase regulates insulin secretion independent of K+ currents. *Mol Endocrinol* **21**: 765-773.

Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Lagarde J, Gilbert JG, Storey R, Swarbreck D et al. 2006. GENCODE: producing a reference annotation for ENCODE. *Genome Biol* **7 Suppl 1**: S4 1-9.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Huang Y, Sook-Kim M, Ratovitski E. 2008. Midkine promotes tetraspanin-integrin interaction and induces FAK-Stat1alpha pathway contributing to migration/invasiveness of human head and neck squamous cell carcinoma cells. *Biochem Biophys Res Commun* **377**: 474-478.

Jiang FX, Stanley EG, Gonez LJ, Harrison LC. 2002. Bone morphogenetic proteins promote development of fetal pancreas epithelial colonies containing insulin-positive cells. *J Cell Sci* **115**(Pt 4): 753-760.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-1303.

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**: 773-777.

Moran I, Akerman I, van de Bunt M, Xie R, Benazra M, Nammo T, Arnes L, Nakic N, Garcia-Hurtado J, Rodriguez-Segui S et al. 2012. Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab* **16**: 435-448.

Morris AP Voight BF Teslovich TM Ferreira T Segre AV Steinthorsdottir V Strawbridge RJ Khan H Grallert H Mahajan A et al. 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**: 981-990.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621-628.

Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**: e1000895.

Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K et al. 2011. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet* **7**: e1002003.

Nicolls MR, D'Antonio JM, Hutton JC, Gill RG, Czwornog JL, Duncan MW. 2003. Proteomics as a tool for discovery: proteins implicated in Alzheimer's disease are highly expressed in normal pancreatic islets. *J Proteome Res* **2**: 199-205.

Ongen H, Dermitzakis E.T. 2012. Altrans: a method for quantification of splicing events (In review).

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413-1415.

Parnaud G, Bosco D, Berney T, Pattou F, Kerr-Conte J, Donath MY, Bruun C, Mandrup-Poulsen T, Billestrup N, Halban PA. 2008. Proliferation of sorted human and rat beta cells. *Diabetologia* **51**: 91-100.

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**: 768-772.

Pisania A, Weir GC, O'Neil JJ, Omer A, Tchipashvili V, Lei J, Colton CK, Bonner-Weir S. 2010. Quantitative analysis of cell composition and purity of human pancreatic islet preparations. *Lab Invest* **90**: 1661-1675.

Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. 2011. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet* **7**: e1001317.

Rondas D, Tomas A, Halban PA. 2011. Focal adhesion remodeling is crucial for glucose-stimulated insulin secretion and involves activation of focal adhesion kinase and paxillin. *Diabetes* **60**: 1146-1157.

Sakurai Y, Shintani N, Hayata A, Hashimoto H, Baba A. 2011. Trophic effects of PACAP on pancreatic islets: a mini-review. *J Mol Neurosci* **43**: 3-7.

Saltiel AR, Kahn CR. 2001. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**: 799-806.

Schwartz MW, Woods SC, Porte D, Jr., Seeley RJ, Baskin DG. 2000. Central nervous system control of food intake. *Nature* **404**: 661-671.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**: 9440-9445.

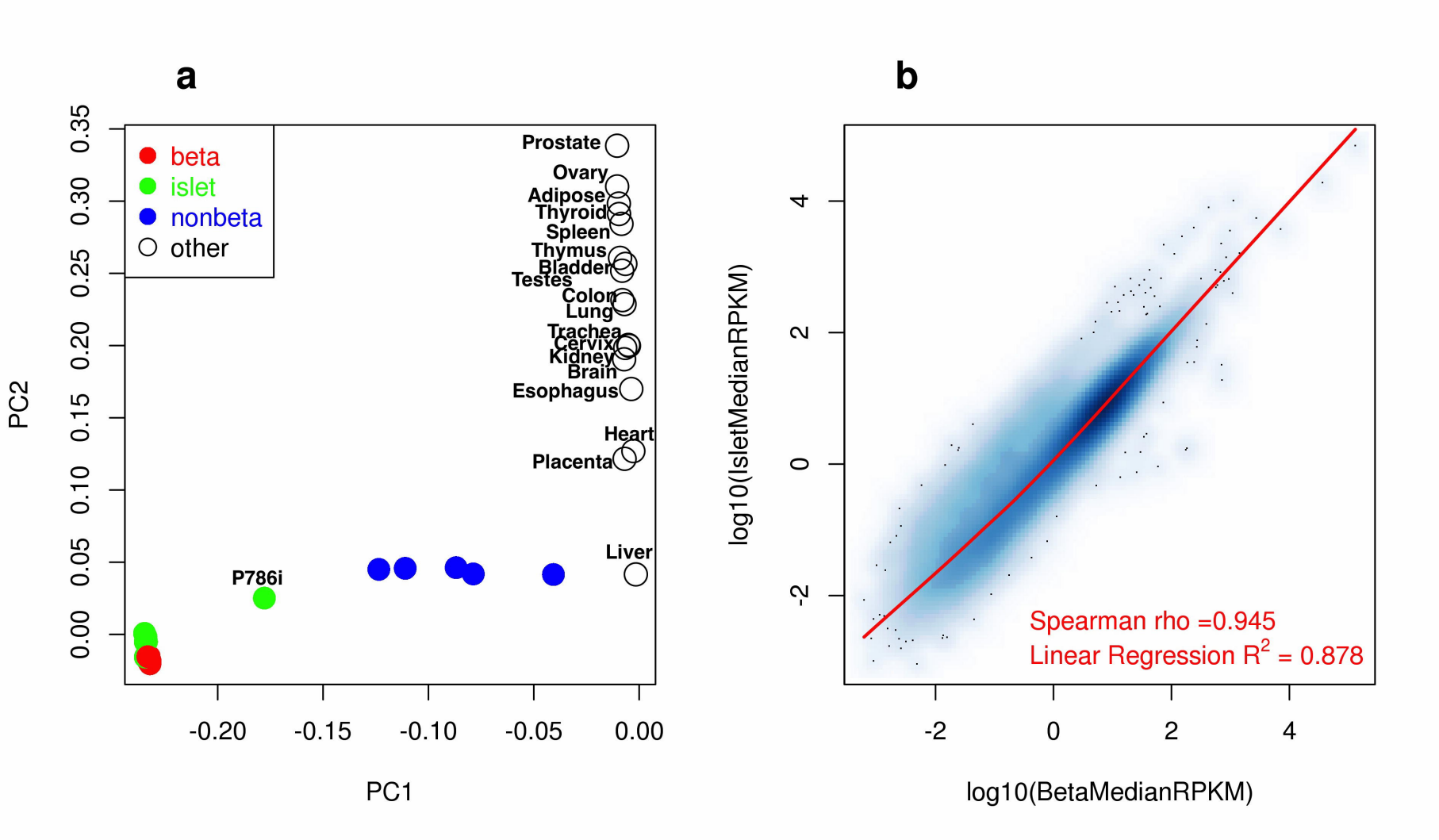
Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D et al. 2007. Population genomics of human gene expression. *Nat Genet* **39**: 1217-1224.

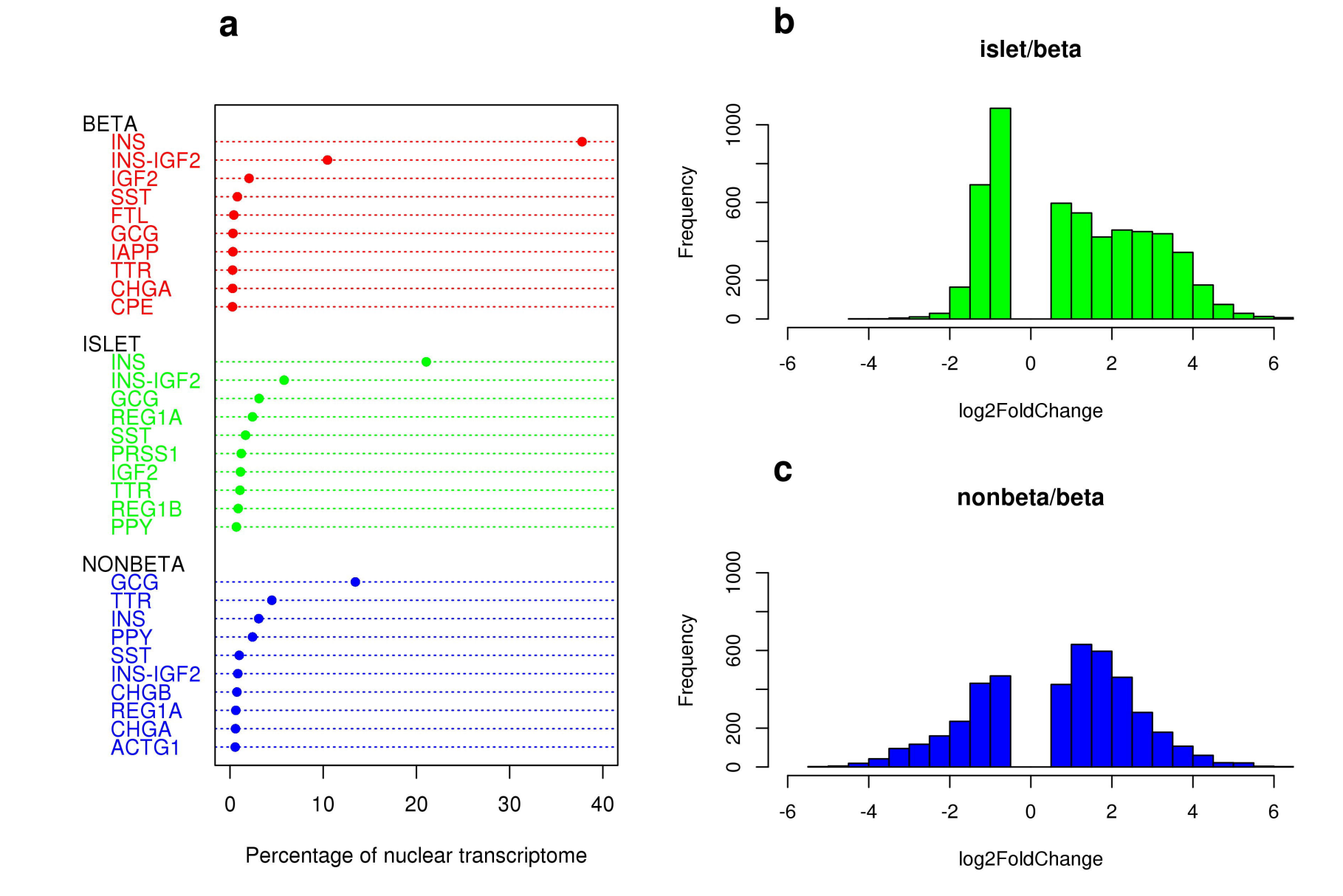
Thomas MF, Ansel KM. 2010. Construction of small RNA cDNA libraries for deep sequencing. *Methods Mol Biol* **667**: 93-111.

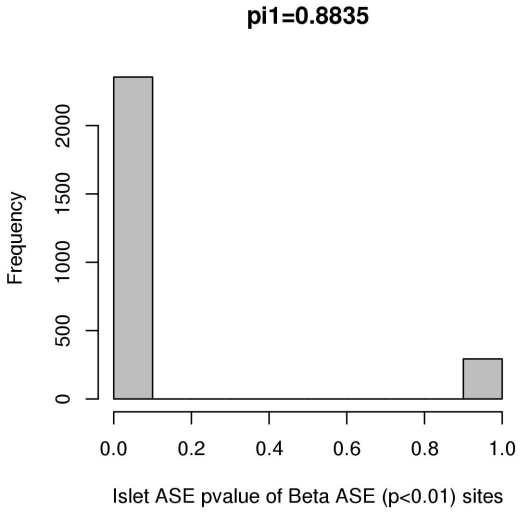
Tisch R, McDevitt H. 1996. Insulin-dependent diabetes mellitus. *Cell* **85**(3): 291-297.

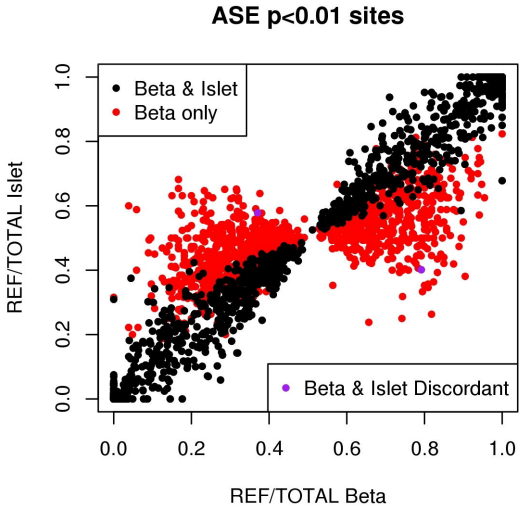
Villasenor A, Wang ZV, Rivera LB, Ocal O, Asterholm IW, Scherer PE, Brekken RA, Cleaver O, Wilkie TM. 2010. Rgs16 and Rgs8 in embryonic endocrine pancreas and mouse models of diabetes. *Dis Model Mech* **3**: 567-580.

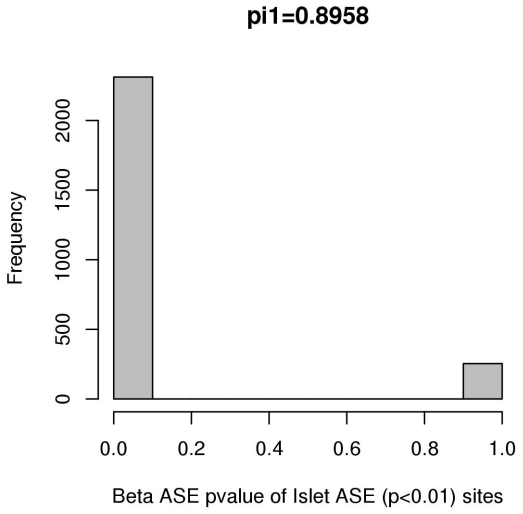
Zhang L, Gao J, Li Z, Gong Y. 2012. Neuronal pentraxin II (NPTX2) is frequently down-regulated by promoter hypermethylation in pancreatic cancers. *Dig Dis Sci* **57**: 2608-2614.

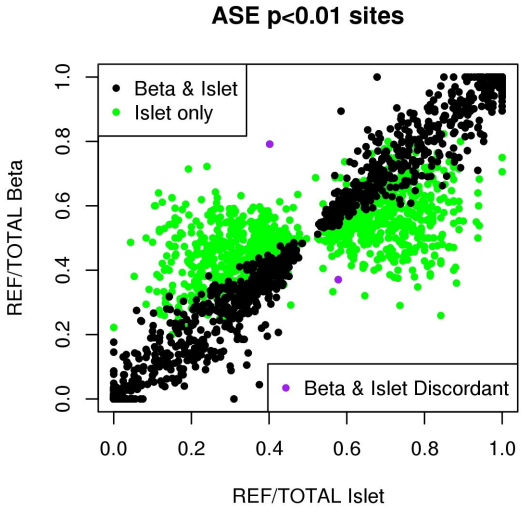

a
b
beta
islet
nonbeta
other
Prostate
Ovary
Adipose
Thyroid
Spleen
Thymus
Bladder
Testes
Colon
Lung
Trachea
Cervix
Kidney
Brain
Esophagus
Heart
Placenta
Liver
P786i
PC1
PC2
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
-0.20
-0.15
-0.10
-0.05
0.00
log10(IsletMedianRPKM)
log10(BetaMedianRPKM)
-2
0
2
4
Spearman rho =0.945
Linear Regression R2 = 0.878

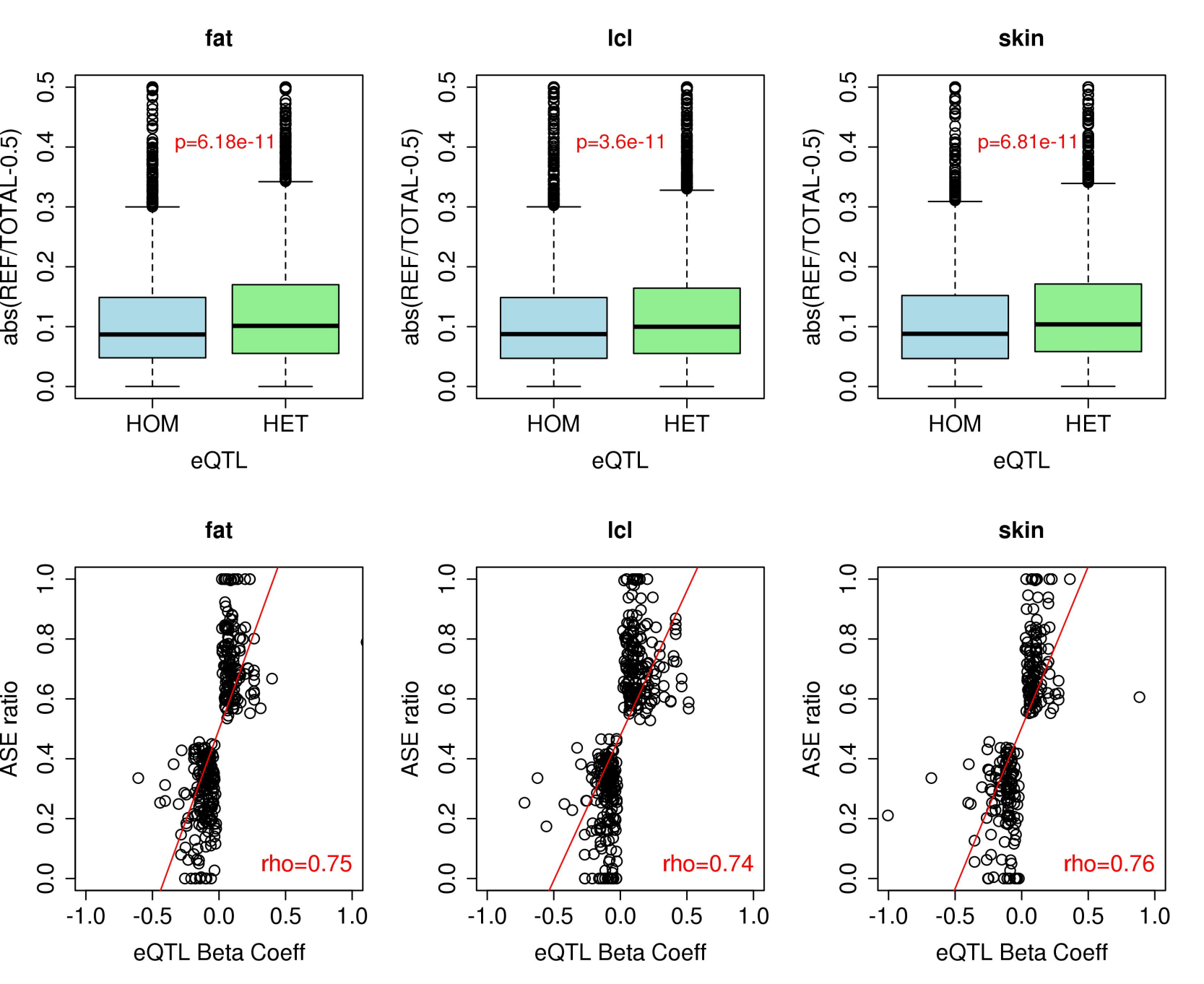

a
BETA
INS
INS-IGF2
IGF2
SST
FTL
GCG
IAPP
TTR
CHGA
CPE
ISLET
INS
INS-IGF2
GCG
REG1A
SST
PRSS1
IGF2
TTR
REG1B
PPY
NONBETA
GCG
TTR
INS
PPY
SST
INS-IGF2
CHGB
REG1A
CHGA
ACTG1
0
10
20
30
40
Percentage of nuclear transcriptome
b
islet/beta
Frequency
0
200
600
1000
-6
-4
-2
0
2
4
6
log2FoldChange
c
nonbeta/beta
Frequency
0
200
600
1000
-6
-4
-2
0
2
4
6
log2FoldChange


pi1=0.8835
Frequency
0
500
1000
1500
2000
0.0
0.2
0.4
0.6
0.8
1.0
Islet ASE pvalue of Beta ASE (p<0.01) sites


ASE p<0.01 sites
Beta & Islet
Beta only
Beta & Islet Discordant
REF/TOTAL Islet
REF/TOTAL Beta
0.0
0.2
0.4
0.6
0.8
1.0


pi1=0.8958
Frequency
0
500
1000
1500
2000
0.0
0.2
0.4
0.6
0.8
1.0
Beta ASE pvalue of Islet ASE (p<0.01) sites


ASE p<0.01 sites
Beta & Islet
Islet only
Beta & Islet Discordant
REF/TOTAL Beta
REF/TOTAL Islet
0.0
0.2
0.4
0.6
0.8
1.0


fat
lcl
skin
abs(REF/TOTAL-0.5)
p=6.18e-11
p=3.6e-11
p=6.81e-11
HOM
HET
eQTL
ASE ratio
rho=0.75
rho=0.74
rho=0.76
eQTL Beta Coeff