Diabetologia (2004) 47:284–299
DOI 10.1007/s00125-003-1300-8

Diabetologia

An expression profile of human pancreatic islet mRNAs by Serial Analysis of Gene Expression (SAGE)

C. Cras-Méneur[2, 4] · **H. Inoue**[1] · **Y. Zhou**[2] · **M. Ohsugi**[2] · **E. Bernal-Mizrachi**[2] · **D. Pape**[3] · **S. W. Clifton**[3] · **M. A. Permutt**[2]

[1] Kawasaki Medical School, Okayama, Japan
[2] Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, St. Louis, Missouri, USA
[3] Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri, USA
[4] Division of Endocrinology, Diabetes and Metabolism, Washington University School of Medicine, St. Louis, USA

Abstract

Aims/hypothesis. The Human Genome Project seeks to identify all genes with the ultimate goal of evaluation of relative expression levels in physiology and in disease states. The purpose of the current study was the identification of the most abundant transcripts in human pancreatic islets and their relative expression levels using Serial Analysis of Gene Expression.
Methods. By cutting cDNAs into small uniform fragments (tags) and concatemerizing them into larger clones, the identity and relative abundance of genes can be estimated for a cDNA library. Approximately 49 000 SAGE tags were obtained from three human libraries: (i) ficoll gradient-purified islets (ii) islets further individually isolated by hand-picking, and (iii) pancreatic exocrine tissue.
Results. The relative abundance of each of the genes identified was approximated by the frequency of the tags. Gene ontology functions showed that all three libraries contained transcripts mostly encoding secreted factors. Comparison of the two islet libraries showed various degrees of contamination from the surrounding exocrine tissue (11 vs 25%). After removal of exocrine transcripts, the relative abundance of 2180 islet transcripts was determined. In addition to the most common genes (e.g. insulin, transthyretin, glucagon), a number of other abundant genes with ill-defined functions such as proSAAS or secretagogin, were also observed.
Conclusion/interpretation. This information could serve as a resource for gene discovery, for comparison of transcript abundance between tissues, and for monitoring gene expression in the study of beta-cell dysfunction of diabetes. Since the chromosomal location of the identified genes is known, this SAGE expression data can be used in setting priorities for candidate genes that map to linkage peaks in families affected with diabetes. [Diabetologia (2004) 47:284–299]

Keywords SAGE · Human · Islet · Transcripts · Gene expression

Received: 31 July 2003 / Revised: 28 October 2003
Published online: 13 January 2004

C. Cras-Méneur (✉)
Division of Endocrinology, Diabetes and Metabolism, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8127, St. Louis, Missouri 63110, USA
E-mail: c.cras-meneur@wustl.edu

Abbreviations: SAGE, Serial Analysis of Gene Expression

Electronic Supplementary Material
Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00125-003-1300-8

Pancreatic islets are a unique tissue with highly specialized functions that determine many of the major metabolic responses to fasting and feeding in higher mammals. In the fed state, a robust secretion of insulin occurs followed by a compensatory increase in insulin biosynthesis. During prolonged insulin resistance, such as that occurring with obesity, compensatory hyperinsulinaemia emerges associated with islet hypertrophy and hyperplasia [1]. For many years, attempts to dissect the molecular steps involved in these pancreatic islet critical adaptive responses have relied on analysis of expression of individual genes through sequencing of cDNA clones and Northern blot analysis of islet mRNA. Later, more comprehensive means

of monitoring expression patterns in islets under various experimental conditions were introduced such as subtractive hybridization [2] and differential display of mRNA [1, 3, 4]. These methods provided important insights into differences in expression levels under various experimental conditions, yet did not provide a global analysis of gene expression.

With the development of human genomics, two major means of examining global gene expression have emerged, quantitative microarray analysis and serial analysis of gene expression (SAGE) [5, 6]. Several studies have recently documented the utility of microarray analysis in exploring a number of experimental issues in diabetes [7]. In contrast, while SAGE analysis has been applied in various other fields of biomedical research [8, 9], this methodology has not been used in the study of diabetes. SAGE analysis provides a method of examining comprehensive gene expression in a tissue of interest in a quantitative fashion that does not rely on prior knowledge of transcripts or on variable hybridization conditions [5, 6]. This method uses the identification of a short stretch of cDNA sequence sufficient to identify specific mRNAs. To facilitate efficient DNA sequencing, these short sequences, or sequence "tags", are concatamerized into cDNA clones, and the number of tags for a given transcript represent the relative expression level of the gene. The method uses poly-A RNA as the template for cDNA synthesis that is then cleaved with an anchoring restriction endonuclease such as *Nla*III, cleaving at four base sites (CATG) found in most cDNAs. The cDNAs are then ligated to linkers with a type IIS restriction enzyme site and then digested into small tags of usually 14 base pairs with the enzyme *Bam*HI. These tags are ligated to form ditags followed by concatemerization and subsequent cloning. Thus cDNAs with numerous small easily identifiable sequence tags of 14 bp each, preceded by the *Nla*III site, can be obtained in a single clone to facilitate sequencing. As a result sequencing as few as 1000 clones can generate enough sequence tags to obtain a broad view of the relative level of expression of the most abundant transcripts in a particular tissue.

In the current experiments, tissues from multiple pancreatic donors were pooled to construct SAGE libraries. To identify an expression profile of pancreatic islets, three libraries were constructed. These included (i) islets isolated by a standard gradient centrifugation protocol often utilized to prepare islets for transplantation [10], (ii) islets isolated by gradient centrifugation followed by further hand selection, and (iii) exocrine pancreas from which islets had been removed. Analysis of a total of 48 915 sequence tags obtained from the three libraries revealed expression profiles that could be compared with each other, and with SAGE libraries from other tissues. Additionally, by a digital subtraction of the exocrine pancreatic contaminants, the relative level of expression of more than 2000 of the most abundant transcripts expressed in human islets is reported. This data provided us with a comparative overview of the distribution of the molecular functions of transcripts in islets and exocrine tissue. We also were able to identify major islet transcripts located in chromosomal regions involved in linkage to diabetes.

Materials and methods

Tissues. SAGE libraries were constructed from three tissue sources. The first SAGE library, SAGEHISL1, was constructed with pancreatic islets from three normal human donors (HR39, HR46, HR50, Islet Isolation Core Facility at Washington University School of Medicine). Individual islets were isolated using a standard protocol of an intraductal Liberase perfusion, gentle mechanical dissociation, and a continuous gradient of Hypaque Euroficoll on a refrigerated COBE 2991 [10]. The islets, collected in fractions, were assayed for purity and fractions with a purity of greater than 90% were stored at −80°C for RNA extraction. The second library was prepared from islets from a single donor, HR96, prepared using the same protocol but with an additional purification by hand selection of 300 pancreatic islets (SAGEHR96R). The third SAGE library, SAGEHEXO1, was constructed with normal pancreatic exocrine cells, which were obtained during the process of islet purification from two donors (HR18, HR27). Informed consent was obtained for all tissue donors and the Washington University Human Studies Committee approved the study.

SAGE library construction and DNA sequencing. Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, Calif., USA), followed by RNA cleanup protocol with DNAse treatment using Qiagen RNAeasy kit (Qiagen, Valencia, Calif., USA) to remove any remaining genomic DNA or degraded RNA. The RNA concentration of the samples was measured spectro-photometrically at 260/280 nm and 1 μg of sample was run on a 1% non-denaturing agarose gel to assess quality.

The SAGE libraries were constructed with an I-SAGE system (Invitrogen) according to the manufacturer's instructions, with minor modifications (Anchor enzyme: *Nla*III). 10 μg of total RNA was used as a starting material. For the pooled SAGE libraries (SAGEHISL1, SAGEHEXO1), equal amounts of total RNA from different donors were mixed to add up to 10 μg. After ligating the DiTag concatamers into the *Sph*I site of pZero-1 (Invitrogen), electro-transformation into Electro-MAX DH10B competent cells (Invitrogen) was performed using Gene Pulser system (BioRad, Hercules, Calif., USA). Bacterial colonies from each library were handpicked and arranged into nineteen 96-well plates (1824 colonies).

Sequencing reactions were carried out using ABI BigDye terminator mix and were loaded and run on an ABI3700 automated DNA sequencer. The sequencing chromatograms were analyzed using the Phred base-caller [11, 12] with high quality right and left cut-off positions determined. The high quality region then was screened for vector-adaptor sequence using local alignments. The vector trimming was confirmed by blasting against a database of all vectors. Sequence was screened against a number of contaminant databases (structural RNA, non-self mitochondria, and bacterial sequence). There also was a check for low entropy sequence and for computer processing errors. Finally, the data was parsed into standard dbEST format, detailed at http://ncbi.nlm.nih.gov/dbEST/. The methods for the construction of the SAGE libraries are illustrated in Fig. 1.

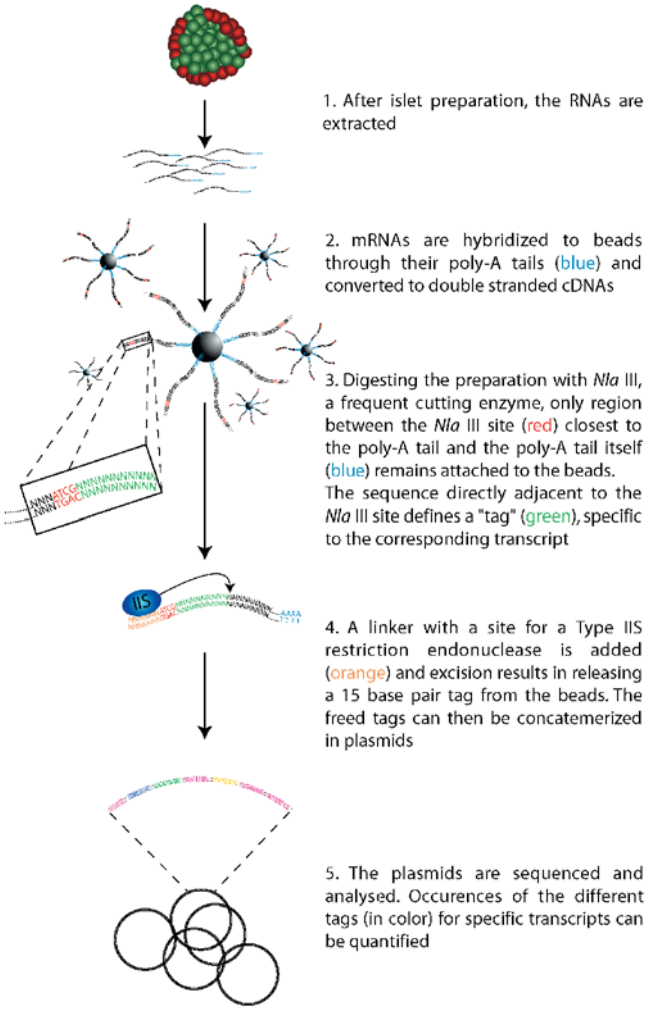


Fig. 1. Schematic representation of the construction of a SAGE library. The diagram represents the method used to construct a SAGE library from an islet preparation. α cells are represented in *red* and beta cells in *green*

SAGE library analysis. SAGE sequences were extracted and analyzed using the SAGE 2000 Software Version 4.12 available on the World Wide Web at http://www.sagenet.org/sage_protocol.htm), and the tags were then mapped using the "reliable" SAGEmap tag to gene mapping databases using build 160 available through http://www.ncbi.nlm.nih. gov/SAGE/ and ftp://ftp.ncbi.nih.gov/pub/sage/map/Hs/NlaIII/.

For the assessment of the relative exocrine content within the islet libraries (SAGEHISL1, SAGEHR96R), the contamination factor (β_i) was estimated with the proportions of the tags corresponding to 13 well-characterized exocrine unique genes. The individual β_i was calculated through $\beta_i = \frac{n_{bi}}{N_b} \times \frac{N_a}{n_{ai}}$ (n_{bi} and n_{ai}: number of tag for the exocrine unique i tag in the islet n_{bi} and exocrine n_{ai} libraries; N_b and N_a: total number of tags in the islet n_{bi} and exocrine n_{ai} libraries).

To provide an evaluation of gene expression profile of the islet libraries deprived of exocrine content, digital subtraction was done tag by tag using the calculated contamination proportion and the relative abundance of each tag in the exocrine and the islet library respectively. The following calculation was used $P_{b'i} = \left(1 + \frac{\bar{\beta}}{1-\bar{\beta}}\right) \cdot P_{bi} - \left(\frac{\bar{\beta}}{1-\bar{\beta}}\right) \cdot P_{ai}$ where P_{ai} is the proportion of the i tag in the exocrine library; P_{bi} is the proportion of the i tag in the islet library, $P_{b'i}$ is the proportion of the i tag in the subtracted islet library and $\bar{\beta}$ is the averaged contamination factor.

Estimation of transcripts specific to islets. Since their tags were not subtracted and the total number of tags diminished by removing the exocrine tags, transcripts whose proportion was increased through the subtraction were considered to be specific to the islets. Genes increased through the subtraction by 25.2%±(2×0.8%) (by over 23.6%) in the gradient purified islet library (HISL1) and 11.4%±(2×1.7%) (by over 8.0%) in the hand-selected islet library (HR96R) could be considered as relatively specific to the islets, or at least highly enriched in islets compared to the exocrine library (HEXO1).

Annotation, gene ontology functions and chromosome locations. All Gene Ontology annotations and chromosome and cytoband locations were gathered using the UniGene cluster IDs collected through the SAGEmap annotation for each transcript from the SOURCE database [13] (http://source.stanford.edu/). The different Gene Ontology categories were further grouped for the molecular functions using the classifications described by the Gene Ontology Consortium [14] (http://www.geneontology.org/). Some transcripts bear several functions and were therefore counted in every category to which they corresponded. The number of tags corresponding to a transcript found in a category were cumulated to represent the relative importance of each gene ontology class.

Results

SAGE analysis of islet and exocrine pancreas libraries. To provide an estimate of mRNA expression levels in pancreatic islets, three human SAGE libraries were constructed. An exocrine pancreas SAGE library (HEXO1, two donors) was compared to those constructed from RNAs extracted from islets either isolated by gradient centrifugation (HISL1, three donors) or islets isolated by gradient centrifugation followed by further purification as a result of hand selection (HR96R, one donor). A total of 48 915 sequence tags were obtained from the three libraries. The entire data set for the libraries has been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/GEO/) and can be retrieved from the SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/) and Endocrine Pancreas Consortium (http://www.cbil.upenn.edu/EPConDB/) web site. The following describes the contents of the libraries and the results of their analyses.

Only tags represented at least twice in a library were considered for analysis to minimize the effects of possible sequencing errors. To map tags to genes for the different libraries, we used the SAGEmap "reliable" tag-to-gene list (build #160). Among the tags used for subsequent analysis, several tags corresponded either to no gene or to several genes and could not be reliably identified. Between 6824 and 12 480 tags

for the three libraries were found to be "reliable" as they had perfect matches to a single gene determined by SAGEmap. All of the tags with a reliable match represented known genes in UniGene (http://www.ncbi.nlm.nih.gov/UniGene/), although some are named while others are currently only ESTs. Multiple tags representing the same gene were then clustered to provide a cumulative count per transcript.

Expression profiles in the exocrine and islet libraries. A total of 13 630 tags were sequenced for the exocrine library (HEXO1). After identification and elimination of all tags counted only once in the library, they corresponded to 1194 transcripts (1013 identified transcripts and 181 remaining different tags that could not be mapped). For the gradient purified islet library (HISL1), there were 1726 transcripts (1524 identified transcripts identified and 202 remaining unmapped different tags), and for the further purified hand selected islet library (HR96R), 1191 transcripts (1042 identified transcripts and 149 unmapped different tags). Amongst the identified tags, between 7.0% and 11.5% were derived from ribosomal or mitochondrial transcripts. While they are in the complete data sets deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) that can also be retrieved through SAGEMap (http://www.ncbi.nlm.nih.gov/SAGE/), they are not listed in the subsequent tables presented here.

The relative abundance of the top 50 genes, represented by the number of tags for each of the three libraries, is shown in Tables 1, 2 and 3. The table provides the original tag count and tags per million as normalization to compare abundance in the different pancreatic libraries, as well as abundance in other libraries available in SAGEmap. Most of the highly expressed genes in the exocrine library represented well known secreted acinar pancreatic enzymes or factors involved in protein synthesis (Table 1). Surprisingly, a gene whose function still remains unclear, regenerating islet-derived 1α was the most highly represented among the top 50 exocrine transcripts.

In the gradient purified islet library (Table 2), while insulin was seen as the most highly expressed gene, many exocrine transcripts were observed as well, presumably due to exocrine contamination during isolation. In contrast, in the gradient purified library that was further hand-selected, the relative exocrine content appeared to be diminished (compare Tables 2 and 3). Notably, many stress response related transcripts like reg-1α and β, clusterin or heat-shock proteins were also present at high levels in the different libraries, perhaps reflecting stress the cells were subjected to during the preparation of the tissues [15, 16, 17].

Estimation of expression profiles of the islet libraries after subtraction of the exocrine tags. To assess the relative exocrine content within the islet libraries, the proportions of the tags corresponding to known exocrine genes were analyzed. Since the relative expression of these transcripts can vary from individual to individual, and the three libraries were constructed from different donors, the analysis was based on 13 relatively abundant known exocrine transcripts. As no exocrine tag is supposed to be present in the islet libraries, the mean contamination factor was calculated as the mean for these 13 known exocrine genes (see Methods). The use of this number of different exocrine genes also helped to decrease the chances of over/under estimation of the exocrine content that could derive from the lack of precision of genes represented only by a few copies. These calculations estimated that 25.2%±0.8% of the tags from the gradient-purified islet library were due to exocrine contamination. The gradient-purified and hand selected islet library showed a much lower content with only 11.4%±1.7% of its tags represented as exocrine transcripts (Table 1S, Supplementary Materials).

To provide estimates of gene expression profiles of the islet libraries deprived of exocrine contamination, digital subtraction was performed tag by tag using the calculated contamination proportion and the relative abundance of each tag in the exocrine and the islet libraries, respectively. The consistency of the calculations was assessed by comparing the result of the subtraction for the two islet libraries one against the other. By comparing the expression levels of each individual gene in the two libraries, a linear regression of expression levels in HR96R and HISL1 provided a very high correlation with R^2=0.9664. Subsequently, merging the results of the two islet libraries provided estimates of relative levels of gene expression from four donors. The two libraries were merged using tags per million as a normalization factor and giving a weight of three to the gradient purified islet library (HISL1, three donors) and one for the hand selected islet library (HR96R, one donor). The top 50 transcripts in the merged islet library after subtraction is presented in Table 4 and the complete list of 2180 transcripts is tabulated by abundance in the Supplementary Materials (Table 2S).

The apparent level of the exocrine transcripts remaining after subtraction was decreased (for example compare Table 2 and 3 with Table 4). The validity of the analysis is suggested by inspection of the genes listed in Table 4, as many of these most abundant transcripts have been previously described as selectively expressed in islets. These include insulin, transthyretin [18], glucagon, the pro-protein convertase inhibitor (proSAAS [19, 20]) clusterin [21], carboxypeptidase-E [22] and islet amyloid polypeptide. Further, while some of the genes found among the most abundant transcripts had previously been described as expressed in pancreatic islets, their relative abundance was unknown, such as secretory granule, neurendocrine pro-

Table 1. The 50 most abundant identified transcripts with their relative levels of expression in the HEXO1 (exocrine library) SAGE libraries. Only the top 50 non-ribosomal or non-mitochondrial transcripts within the library are shown in this table. The transcripts are ranked by their relative abundance and represented with their tag counts within each libraries and the corresponding tags per million to allow comparisons.. When the tags could not be matched to any UniGene cluster ID, the name of the transcript has been replaced by its tag. The complete lists of the transcripts in these libraries can be retrieved through SAGEmap

UniGene ID	Symbol	Name	Cumulated count	Tags per million
Hs.49407	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	872	48.964
Hs.78546	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	681	38.239
Hs.2879	CPA1	carboxypeptidase A1 (pancreatic)	570	32.006
Hs.241561	PRSS2	protease, serine, 2 (trypsin 2)	525	29.479
Hs.419094	PRSS1	protease, serine, 1 (trypsin 1)	456	25.605
Hs.300280	AMY2A	amylase, alpha 2A; pancreatic	364	20.439
Hs.929	MYH7	myosin, heavy polypeptide 7, cardiac muscle, beta	357	20.046
Hs.180884	CPB1	carboxypeptidase B1 (tissue)	313	17.575
Hs.181289	ELA3A	elastase 3A, pancreatic (protease E)	298	16.733
Hs.1340	CLPS	colipase, pancreatic	269	15.105
Hs.89717	CPA2	carboxypeptidase A2 (pancreatic)	255	14.319
Hs.102876	PNLIP	pancreatic lipase	244	13.701
Hs.406160	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	232	13.027
Hs.425790	ELA3B	elastase 3B, pancreatic	184	10.332
Hs.48604	DKFZP434B168	DKFZP434B168 protein	157	8.816
Hs.74502	CTRB1	chymotrypsinogen B1	146	8.198
Hs.53985	GP2	glycoprotein 2 (zymogen granule membrane)	145	8.142
Hs.401448	TPT1	tumor protein, translationally-controlled 1	143	8.030
Hs.21	ELA2A	elastase 2A	124	6.963
Hs.992	PLA2G1B	phospholipase A2, group IB (pancreas)	100	5.615
Hs.133430		ESTs	99	5.559
		GGTTTACTGA	96	5.391
		GAACACACAA	93	5.222
Hs.4158	REG1B	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)	90	5.054
Hs.181286	SPINK1	serine protease inhibitor, Kazal type 1	89	4.997
		TCCCCGTACA	79	4.436
Hs.8709	CTRC	chymotrypsin C (caldecrin)	77	4.324
Hs.58247	PRSS3	protease, serine, 3 (mesotrypsin)	71	3.987
Hs.143113	PNLIPRP2	pancreatic lipase-related protein 2	68	3.818
Hs.107287	KIAA1411	KIAA1411 protein	64	3.594
Hs.234726	SERPINA3	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	60	3.369
Hs.90436	SPAG7	sperm associated antigen 7	57	3.201
Hs.335493	AMY2B	amylase, alpha 2B; pancreatic	56	3.144
		CCCATCGTCC	51	2.864
		TTCATACACC	47	2.639
Hs.299916	MGC46680	hypothetical protein MGC46680	46	2.583
Hs.75309	EEF2	eukaryotic translation elongation factor 2	45	2.527
Hs.422118	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	44	2.471
Hs.114648	ERG-1	estrogen regulated gene 1	44	2.471
Hs.268049	HSPC031	hypothetical protein HSPC031	39	2.190
Hs.149923	XBP1	X-box binding protein 1	38	2.134
Hs.95243	TCEAL1	transcription elongation factor A (SII)-like 1	36	2.021
Hs.182476	WBSCR21	Williams Beuren syndrome chromosome region 21	36	2.021
		CACCTAATTG	36	2.021
Hs.418650	FTH1	ferritin, heavy polypeptide 1	35	1.965
Hs.421608	EEF1B2	eukaryotic translation elongation factor 1 beta 2	34	1.909
		CACTACTCAC	34	1.909
		TGATTTCACT	34	1.909
Hs.89832	INS	insulin	32	1.797
Hs.74566	DPYSL3	dihydropyrimidinase-like 3	32	1.797

Table 2. The 50 most abundant identified transcripts with their relative levels of expression in the HISL1 (ficoll gradient-purified islets library) SAGE libraries. Only the top 50 non-ribosomal or non-mitochondrial transcripts within the library are shown in this table. The transcripts are ranked by their relative abundance and represented with their tag counts within each libraries and the corresponding tags per million to allow comparisons. When the tags could not be matched to any UniGene cluster ID, the name of the transcript has been replaced by its tag. The complete lists of the transcripts in these libraries can be retrieved through SAGEmap

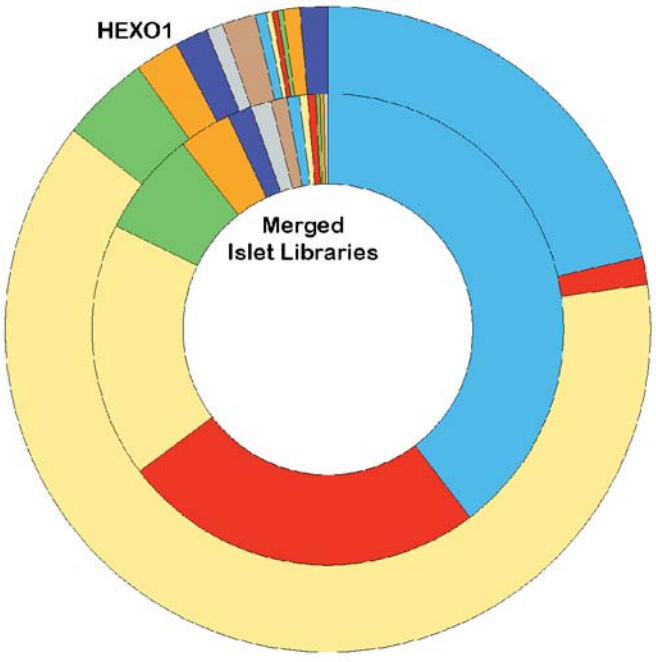
UniGene ID	Symbol	Name	Cumulated count	Tags per million
Hs.89832	INS	insulin	2188	116.179
Hs.49407	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	373	19.806
Hs.78546	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	250	13.275
Hs.427202	TTR	transthyretin (prealbumin, amyloidosis type I)	187	9.929
Hs.2879	CPA1	carboxypeptidase A1 (pancreatic)	158	8.390
Hs.419094	PRSS1	protease, serine, 1 (trypsin 1)	149	7.912
Hs.241561	PRSS2	protease, serine, 2 (trypsin 2)	134	7.115
Hs.133430		ESTs	129	6.850
		TCCCCGTACA	128	6.797
		TCCCTATTAA	110	5.841
Hs.300280	AMY2A	amylase, alpha 2A; pancreatic	99	5.257
Hs.401448	TPT1	tumor protein, translationally-controlled 1	97	5.151
Hs.180884	CPB1	carboxypeptidase B1 (tissue)	85	4.513
		CCCATCGTCC	77	4.089
		TGATTTCACT	76	4.035
Hs.406160	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	74	3.929
Hs.929	MYH7	myosin, heavy polypeptide 7, cardiac muscle, beta	73	3.876
Hs.181289	ELA3A	elastase 3A, pancreatic (protease E)	71	3.770
Hs.1340	CLPS	colipase, pancreatic	68	3.611
Hs.89717	CPA2	carboxypeptidase A2 (pancreatic)	65	3.451
Hs.102876	PNLIP	pancreatic lipase	55	2.920
Hs.422118	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	53	2.814
		CACCTAATTG	53	2.814
Hs.53985	GP2	glycoprotein 2 (zymogen granule membrane)	51	2.708
		TTCATACACC	49	2.602
Hs.425790	ELA3B	elastase 3B, pancreatic	47	2.496
Hs.48516	B2M	beta-2-microglobulin	45	2.389
Hs.181286	SPINK1	serine protease inhibitor, Kazal type 1	45	2.389
Hs.90436	SPAG7	sperm associated antigen 7	44	2.336
Hs.429437	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	42	2.230
Hs.418650	FTH1	ferritin, heavy polypeptide 1	42	2.230
Hs.107003	HEI10	enhancer of invasion 10	42	2.230
Hs.8709	CTRC	chymotrypsin C (caldecrin)	41	2.177
Hs.268049	HSPC031	hypothetical protein HSPC031	40	2.124
Hs.95243	TCEAL1	transcription elongation factor A (SII)-like 1	39	2.071
		GGTTTACTGA	39	2.071
Hs.423901	GCG	glucagon	38	2.018
Hs.234726	SERPINA3	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	38	2.018
Hs.75106	CLU	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	34	1.805
Hs.74502	CTRB1	chymotrypsinogen B1	32	1.699
Hs.379466	UBE2A	ubiquitin-conjugating enzyme E2A (RAD6 homolog)	32	1.699
Hs.992	PLA2G1B	phospholipase A2, group IB (pancreas)	30	1.593
Hs.77495	UBXD2	UBX domain containing 2	30	1.593
Hs.169476	GAPD	glyceraldehyde-3-phosphate dehydrogenase	30	1.593
		CACTACTCAC	30	1.593
Hs.75968	TMSB4X	thymosin, beta 4, X chromosome	29	1.540
Hs.423	PAP	pancreatitis-associated protein	29	1.540
Hs.4158	REG1B	regenerating islet-derived 1 beta (pancreatic stone protein, pancreatic thread protein)	29	1.540
		CAAGCATCCC	28	1.487
		CTAAGACTTC	28	1.487

Table 3. The 50 most abundant identified transcripts with their relative levels of expression in the HR96R (gradient-purified and handpicked islet library) SAGE libraries. Only the top 50 non-ribosomal or non-mitochondrial transcripts within the library are shown in this table. The transcripts are ranked by their relative abundance and represented with their tag counts within each libraries and the corresponding tags per million to allow comparisons. When the tags could not be matched to any UniGene cluster ID, the name of the transcript has been replaced by its tag. The complete lists of the transcripts in these libraries can be retrieved through SAGE map

UniGene ID	Symbol	Name	Cumulated count	Tags per million
Hs.89832	INS	insulin	1426	116.190
Hs.427202	TTR	transthyretin (prealbumin, amyloidosis type I)	287	23.385
Hs.423901	GCG	glucagon	138	11.244
		TCCCCGTACA	89	7.252
		TTCATACACC	87	7.089
		CCCATCGTCC	82	6.681
		CACCTAATTG	78	6.355
		TCCCTATTAA	77	6.274
Hs.300280	AMY2A	amylase, alpha 2A; pancreatic	67	5.459
Hs.2879	CPA1	carboxypeptidase A1 (pancreatic)	67	5.459
Hs.133430	–	ESTs	65	5.296
Hs.401448	TPT1	tumor protein, translationally-controlled 1	55	4.481
Hs.78546	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	48	3.911
Hs.429437	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	47	3.830
Hs.2265	SGNE1	secretory granule, neuroendocrine protein 1 (7B2 protein)	45	3.667
		TGATTTCACT	43	3.504
Hs.2281	CHGB	chromogranin B (secretogranin 1)	41	3.341
Hs.95243	TCEAL1	transcription elongation factor A (SII)-like 1	40	3.259
Hs.49407	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	40	3.259
Hs.419094	PRSS1	protease, serine, 1 (trypsin 1)	40	3.259
Hs.181289	ELA3A	elastase 3A, pancreatic (protease E)	39	3.178
Hs.436980		ESTs, Highly similar to NME3_HUMAN Glutamate [NMDA] receptor subunit epsilon 3 precursor (N-methyl D-aspartate receptor subtype 2C) (NR2C) (NMDAR2C) [*H. sapiens*]	35	2.852
Hs.75106	CLU	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	34	2.770
Hs.48516	B2M	beta-2-microglobulin	29	2.363
Hs.241561	PRSS2	protease, serine, 2 (trypsin 2)	27	2.200
Hs.75360	CPE	carboxypeptidase E	26	2.118
Hs.142255	IAPP	islet amyloid polypeptide	24	1.956
Hs.127179	CRYPTIC	cryptic gene	24	1.956
		ACTAACACCC	24	1.956
		AGAGGTGTAG	23	1.874
Hs.418650	FTH1	ferritin, heavy polypeptide 1	22	1.793
		CTAAGACTTC	22	1.793
		ACCCTTGGCC	22	1.793
		GGTCAGTCGG	22	1.793
Hs.404283	RAD23B	RAD23 homolog B (*S. cerevisiae*)	21	1.711
Hs.374523	GNAS	GNAS complex locus	21	1.711
Hs.268049	HSPC031	hypothetical protein HSPC031	20	1.630
Hs.181874	IFIT4	interferon-induced protein with tetratricopeptide repeats 4	20	1.630
Hs.116428	SCGN	secretagogin, EF-hand calcium binding protein	20	1.630
		CAAGCATCCC	20	1.630
Hs.406160	CEL	carboxyl ester lipase (bile salt-stimulated lipase)	19	1.548
Hs.323562	DKFZp564K142	hypothetical protein DKFZp564K142 similar to implantation-associated protein	19	1.548
		TCCCGTACAT	19	1.548
Hs.89655	PTPRN	protein tyrosine phosphatase, receptor type, N	18	1.467
Hs.75452	HSPA1A	heat shock 70 kDa protein 1A	18	1.467
Hs.422118	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	18	1.467
Hs.198281	PKM2	pyruvate kinase, muscle	18	1.467
Hs.119206	IGFBP7	insulin-like growth factor binding protein 7	17	1.385
		CACTACTCAC	17	1.385
Hs.90436	SPAG7	sperm-associated antigen 7	15	1.222

Table 4. List of the 50 most abundant transcripts (non mitochondrial and non ribosomal) within the exocrine-subtracted merged islet library. The transcripts are ranked according to their relative abundance using the counts in tags per million. The complete lists of the transcripts in this library can be found in Table 2S of the supplementary material

UniGene IDs	Symbol	Name	Tags per million
Hs.89832	INS	insulin	148.791
Hs.427202	TTR	transthyretin (prealbumin, amyloidosis type I)	16.508
		TCCCCGTACA	7.598
Hs.49407	REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	7.479
		TCCCTATTAA	7.260
Hs.133430		ESTs	6.779
Hs.423901	GCG	glucagon	5.196
		CCCATCGTCC	5.169
		TGATTTCACT	4.492
Hs.401448	TPT1	tumor protein, translationally-controlled 1	4.141
		CACCTAATTG	4.039
		TTCATACACC	3.857
Hs.78546	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	3.641
Hs.429437	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	3.317
Hs.48516	B2M	beta-2-microglobulin	2.903
Hs.422118	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	2.532
Hs.95243	TCEAL1	transcription elongation factor A (SII)-like 1	2.420
Hs.75106	CLU	clusterin (complement lysis inhibitor, SP-40,40, sulfated glycoprotein 2, testosterone-repressed prostate message 2, apolipoprotein J)	2.368
Hs.2265	SGNE1	secretory granule, neuroendocrine protein 1 (7B2 protein)	2.313
Hs.418650	FTH1	ferritin, heavy polypeptide 1	2.182
Hs.107003	HEI10	enhancer of invasion 10	2.069
Hs.268049	CGI-37	hypothetical protein HSPC031	1.966
Hs.169476	GAPD	glyceraldehyde-3-phosphate dehydrogenase	1.830
Hs.90436	SPAG7	sperm-associated antigen 7	1.776
		CTAAGACTTC	1.773
		CAAGCATCCC	1.647
Hs.2281	CHGB	chromogranin B (secretogranin 1)	1.635
Hs.180370	CFL1	cofilin 1 (non-muscle)	1.603
Hs.77495	UBXD2	UBX domain containing 2	1.603
Hs.76053	DDX5	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68 kDa)	1.585
Hs.419094	PRSS1	protease, serine, 1 (trypsin 1)	1.555
		TCCCGTACAT	1.539
Hs.323562	DKFZp564K142	hypothetical protein DKFZp564K142 similar to implantation-associated protein	1.533
Hs.426138	TMSB4X	thymosin, beta 4, X chromosome	1.526
Hs.75360	CPE	carboxypeptidase E	1.524
Hs.73818	UQCRH	ubiquinol-cytochrome c reductase hinge protein	1.492
		CACTACTCAC	1.444
		ACTAACACCC	1.409
		TCTCCATACC	1.385
Hs.142255	IAPP	islet amyloid polypeptide	1.351
Hs.379466	UBE2A	ubiquitin-conjugating enzyme E2A (RAD6 homolog)	1.346
Hs.14376	ACTG1	actin, gamma 1	1.341
Hs.404283	RAD23B	RAD23 homolog B (*S. cerevisiae*)	1.292
Hs.374523	GNAS	GNAS complex locus	1.282
Hs.181244	HLA-A	major histocompatibility complex, class I, A	1.240
Hs.181874	IFIT4	interferon-induced protein with tetratricopeptide repeats 4	1.231
Hs.75410	HSPA5	heat shock 70 kDa protein 5 (glucose-regulated protein, 78 kDa)	1.231
Hs.8709	CTRC	chymotrypsin C (caldecrin)	1.204
Hs.234726	SERPINA3	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	1.171
Hs.423	PAP	pancreatitis-associated protein	1.161



Molecular Function	Merged Islet Library (tags per million)	HEXO1 (tags per million)
binding	562 044	358 976
signal transducer activity	350 689	21 287
catalytic activity	246 689	1 053 119
transporter activity	101 832	75 243
transcription regulator activity	53 834	36 948
enzyme regulator activity	21 563	30 041
translation regulator activity	20 043	15 835
structural molecule activity	15 936	26 840
antioxidant activity	10 925	7468
chaperone activity	8781	4548
cell adhesion molecule activity	5104	3594
defense/immunity protein activity	4757	5559
chaperone regulator activity	3770	15 610
motor activity	2684	21 843
apoptosis regulator activity	1836	505
anticoagulant activity	426	112
cytoskeletal regulator activity	175	0
protein tagging activity	50	225
nutrient reservoir activity	0	112

Fig. 2. Distribution of the molecular functions in the merged islet and the exocrine libraries. The molecular functions of all identified transcripts were collected for the merged islet library and the exocrine library. The different molecular functions were regrouped into 19 major top-level classes. For each function found for a transcript, the tag count for the gene was added to the corresponding class (some transcripts bear several functions in the same or in different classes). The cumulative tag counts for each class are represented here in tags per million in two concentric pie charts for the merged islet library (*inner circle*) and the exocrine library (*outer circle*)

tein 1 (7B2) [23] and chromogranin B [24]. Interestingly, 12 tags from the most abundant 50 transcripts could not be clearly identified through SAGEmap, as they were either "no match" or the tags mapped to several genes. These could represent tags not previously depicted for known genes, or tags for completely novel genes.

Amongst the transcripts remaining after subtraction and using the criteria defined in the Methods, we were able to determine that there were respectively 1090 and 777 genes considered as islet enriched in these two libraries, whereas the remaining genes are composed of either exocrine transcripts or genes common to both tissues. The complete list of all transcripts considered here to be specific is tabulated by abundance in the Supplemental Material (respectively 3S and 4S of the Supplementary Materials).

GO functions for the exocrine (HEX01) and merged islet libraries. For all identified transcripts in the exocrine (HEXO1) and merged islet libraries, all available gene ontology functions were gathered and regrouped into the major top-level classes defined for the molecular functions by the Gene Ontology consortium (http://www.geneontology.org/). Within these two libraries, 1179 (merged islet library—representing 416 772 tags per million) and 590 (HEXO1—representing 470 829 tags per million) transcripts had gene ontology functions. These molecular functions represent the various tasks carried out by the proteins. The full range of tasks represented in these libraries represents hundreds of possibilities that have been regrouped according to the 19 major top level classes to simplify this diagram (Fig. 2). The three major categories represented here for both libraries are "Binding activities", representing several classes of mechanisms comprising ligands, protein binding, but also DNA binding activities, "Signal transducer activity" representing different members of the signal transduction pathways and "Catalytic activities", comprising various enzymes and signal transducers (kinases, phosphatases...). The hierarchy of the molecular functions can be consulted at the gene ontology web-site (http://www.geneontology.org/).

Relationship of gene expression in the islets and chromosomal location of linkage peaks for Type 2 diabetes. Genome scans of families with multiple members affected with diabetes have identified chromosomal regions likely to harbour genes that contribute to disease risk. Chromosome 1q, 12q and 20q have previously been identified as carrying four regions most reproducibly shown to harbour Type 2 diabetes mellitus genes [25]. Islet genes identified through SAGE and mapped to these regions are presented in Tables 5, 6 and 7 and ranked according to their chromosomal locations. The genes associated with the highest linkage peaks have been highlighted.

Table 5. Islet transcripts expressed in the Type 2 diabetes mellitus regions on Chromosome 1q. The transcripts identified as being expressed in the islets are organized according to their chromosome locations for chromosome 1q. Regions thought to harbour Type 2 diabetes mellitus gene have been highlighted for the transcript located between 1q12 and 1q23.2

UniGene ID	Symbol	Name	Tag per million	Cytoband
1q12–1q23.2				
Hs.265848	**PDE4DIP**	**phosphodiesterase 4D interacting protein (myomegalin)**	**320**	**1q12**
Hs.275243	**S100A6**	**S100 calcium binding protein A6 (calcyclin)**	**533**	**1q21**
Hs.400250	**S100A10**	**S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))**	**213**	**1q21**
Hs.15456	**PDZK1**	**PDZ domain containing 1**	**107**	**1q21**
Hs.6396	**JTB**	**jumping translocation breakpoint**	**107**	**1q21**
Hs.86386	**MCL1**	**myeloid cell leukemia sequence 1 (BCL2-related)**	**107**	**1q21**
Hs.417004	**S100A11**	**S100 calcium binding protein A11 (calgizzarin)**	**78**	**1q21**
Hs.85844	**TPM3**	**tropomyosin 3**	**568**	**1q21.2**
Hs.355906	**NICE-3**	**NICE-3 protein**	**252**	**1q21.2**
Hs.151536	**RAB13**	**RAB13, member RAS oncogene family**	**50**	**1q21.2**
Hs.333512	**LOC64182**	**similar to rat myomegalin**	**46**	**1q21.2**
Hs.50785	**SEC22L1**	**SEC22 vesicle trafficking protein-like 1 (*S. cerevisiae*)**	**160**	**1q21.2-q21.3**
Hs.12284	**F11R**	**F11 receptor**	**107**	**1q21.2-q21.3**
Hs.111680	**ENSA**	**endosulfine alpha**	**376**	**1q21.3**
Hs.89230	**KCNN3**	**potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3**	**259**	**1q21.3**
Hs.285976	**LASS2**	**LAG1 longevity assurance homolog 2 (*S. cerevisiae*)**	**107**	**1q21.3**
Hs.226499	**RUSC1**	**RUN and SH3 domain containing 1**	**160**	**1q21-q22**
Hs.334841	**SELENBP1**	**selenium binding protein 1**	**46**	**1q21-q22**
Hs.74564	**SSR2**	**signal sequence receptor, beta (translocon-associated protein beta)**	**78**	**1q21-q23**
Hs.406504	**TAGLN2**	**transgelin 2**	**683**	**1q21-q25**
Hs.15318	**HAX1**	**HS1 binding protein**	**373**	**1q22**
Hs.168670	**PXF**	**peroxisomal farnesylated protein**	**213**	**1q22**
Hs.8015	**USP21**	**ubiquitin specific protease 21**	**160**	**1q22**
Hs.18851	**YAP**	**YY1-associated protein**	**107**	**1q22**
Hs.75117	**ILF2**	**interleukin enhancer binding factor 2, 45 kDa**	**107**	**1q22**
Hs.110707	**H326**	**H326**	**107**	**1q22-q23**
Hs.78629	**ATP1B1**	**ATPase, Na+/K+ transporting, beta 1 polypeptide**	**102**	**1q22-q25**
Hs.173611	**NDUFS2**	**NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49 kDa (NADH-coenzyme Q reductase)**	**120**	**1q23**
Hs.372679	**FCGR3B**	**Fc fragment of IgG, low affinity IIIb, receptor for (CD16)**	**107**	**1q23**
Hs.424468	**MGST3**	**microsomal glutathione S-transferase 3**	**107**	**1q23**
Hs.1708	**CCT3**	**chaperonin containing TCP1, subunit 3 (gamma)**	**107**	**1q23**
Hs.177507	**HSPC155**	**hypothetical protein HSPC155**	**386**	**1q23.1**
Hs.97784	**na**	**Homo sapiens, similar to hypothetical protein, clone MGC:33651 IMAGE:4827863, mRNA, complete cds**	**143**	**1q23.1**
Hs.169681	**DEDD**	**death effector domain containing**	**41**	**1q23.1**
Hs.69559	**XTP2**	**HBxAg transactivated protein 2**	**107**	**1q23.3**
Hs.75887	**COPA**	**coatomer protein complex, subunit alpha**	**107**	**1q23-q25**
Hs.76285	DKFZP564B167	DKFZP564B167 protein	131	1q24
Hs.274479	NME7	non-metastatic cells 7, protein expressed in (nucleoside-diphosphate kinase)	107	1q24
Hs.77266	QSCN6	quiescin Q6	107	1q24
Hs.120	PRDX6	antioxidant protein 2	107	1q24.1
Hs.105737	COP1	constitutive photomorphogenic protein	107	1q24.2
Hs.12532	C1orf21	chromosome 1 open reading frame 21	160	1q25
Hs.169750	TPR	translocated promoter region (to activated MET oncogene)	46	1q25
Hs.54451	LAMC2	laminin, gamma 2	213	1q25-q31
Hs.211585	PFKFB2	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	160	1q31
Hs.554	SSA2	Sjogren syndrome antigen A2 (60 kDa, ribonucleoprotein autoantigen SS-A/Ro)	107	1q31
Hs.170171	GLUL	glutamate-ammonia ligase (glutamine synthase)	42	1q31
Hs.23585	KIAA1078	KIAA1078 protein	107	1q31.3
Hs.108080	CSRP1	cysteine and glycine-rich protein 1	358	1q32
Hs.75074	MAPKAPK2	mitogen-activated protein kinase-activated protein kinase 2	160	1q32
Hs.283667	RNPEP	arginyl aminopeptidase (aminopeptidase B)	107	1q32

Table 5 (continued)

UniGene ID	Symbol	Name	Tag per million	Cytoband
Hs.7309	na	Homo sapiens cDNA FLJ34019 fis, clone FCBBF2002898	46	1q32.1
Hs.155079	PPP2R5A	protein phosphatase 2, regulatory subunit B (B56), alpha isoform	107	1q32.2-q32.3
Hs.432132	G0S2	putative lymphocyte G0/G1 switch gene	107	1q32.2-q41
Hs.30318	FLJ10874	hypothetical protein FLJ10874	50	1q32.3
Hs.109494	SPUF	secreted protein of unknown function	46	1q32.3
Hs.3764	GUK1	guanylate kinase 1	126	1q32-q41
Hs.15087	C1orf16	chromosome 1 open reading frame 16	107	1q35
Hs.181307	H3F3A	H3 histone, family 3A	692	1q41
Hs.74870	HLX1	H2.0-like homeo box 1 (Drosophila)	46	1q41-q42.1
Hs.74571	ARF1	ADP-ribosylation factor 1	107	1q42
Hs.89649	EPHX1	epoxide hydrolase 1, microsomal (xenobiotic)	88	1q42.1
Hs.26244	FLJ10052	hypothetical protein FLJ10052	107	1q42.12
Hs.273186	CABC1	chaperone, ABC1 activity of bc1 complex like (*S. pombe*)	280	1q42.13
Hs.75975	SRP9	signal recognition particle 9 kDa	263	1q42.13
Hs.150763	ERO1LB	ERO1-like beta (*S. cerevisiae*)	420	1q42.2-q43
Hs.117183	na	hypothetical protein LOC200169	92	1q42.3
Hs.32976	GNG4	guanine nucleotide binding protein (G protein), gamma 4	69	1q42.3
Hs.356624	NID	nidogen (enactin)	46	1q43
Hs.300642	SDCCAG8	serologically-defined colon cancer antigen 8	107	1q44
Hs.103804	HNRPU	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	65	1q44
Hs.298573	KIAA1720	KIAA1720 protein	46	1q44

Discussion

This study uses SAGE analysis for human islet mRNA. From three human libraries, nearly 50 000 sequence tags were deposited in public databases (SAGEmap, http://www.ncbi.nlm.nih.gov/SAGE/). These data record the relative levels of expression of the most abundant human exocrine and endocrine pancreatic genes. Interestingly, this is the first exocrine SAGE library from normal tissue. A distinct advantage of submitting a tissue to SAGE analysis is that the digital readout of gene expression can be directly compared to those of several hundred libraries created from other human tissues such as brain, liver, skeletal muscle, etc. With software provided through SAGEmap, one can readily examine genes common to these tissues, as well as the relative level of enrichment of particular genes in specialized tissues by means of, for example, "virtual Northern blots". An integrated analysis of the RNA abundance in the islets and chromosomal locations can also allow combining gene expression and linkage analysis for a disease. The study of candidate factors in chromosomal regions associated with Type 2 diabetes can be filtered according to their expression in the islets.

Over a fifth of the transcripts in ficoll-gradient purified islet libraries came from exocrine mRNAs (Table 4). After digital subtraction of exocrine transcripts from both islet libraries, 2180 genes expressed in islets were identified. The analysis of the libraries, after subtraction, depicted the relative expression level of the most abundant transcripts for three islet donors for the gradient purified library and one donor for the gradient purified and subsequent hand selected islet library. As gene expression varies from individual to individual, these variations can affect the relative "ranks" of the transcripts identified. To assess this issue, a linear regression analysis between the two subtracted islet libraries was performed that revealed a high correlation factor (R^2=0. 9664) allowing us to merge the libraries providing values more representative in human adult islets as the results represent a composite from four donors.

The results of the SAGE analysis for islets shown in Table 4 are supported by the observation of well-recognized abundant islet gene products such as insulin, glucagon, islet amyloid polypeptide, and pancreatic polypeptide. These findings also concur with the findings of a large endocrine pancreas EST sequencing project [Endocrine Pancreas Consortium (EPCon)], http://www.cbil.upenn.edu/EPConDB/index.shtml. Human and mouse pancreas, islet, and insulinoma libraries were constructed and over 170 000 ESTs submitted to the public databases [26]. These ESTs were derived from a mixture of developmental stages of pancreas, not exclusively islet transcripts, and for these reasons cannot be compared directly to the two islet SAGE libraries presented here. Many of the most abundant genes in the current SAGE analysis such as insulin, transthyretin, glucagon, beta-2-microglobulin, carboxypeptidase-E or islet amyloid polypeptide, however, were also found to be amongst the most abundant transcripts within the EPCon libraries.

Table 6. Islet transcripts expressed in the Type 2 diabetes mellitus regions on Chromosome 12q. The transcripts identified as being expressed in the islets are organized according to their chromosome locations for chromosome 12q. Regions thought to harbour Type 2 diabetes mellitus gene have been highlighted for the transcript located between 12q13.12 and 12q15

UniGene ID	Symbol	Name	Tag per million	Cytoband
Hs.298275	SLC38A2	solute carrier family 38, member 2	41	12q
Hs.288856	PFDN5	prefoldin 5	154	12q12
Hs.82643	PTK9	PTK9 protein tyrosine kinase 9	88	12q12
Hs.43847	MADP-1	MADP-1 protein	46	12q12
Hs.433394	TUBA3	tubulin, alpha 3	160	12q12–12q14.3
Hs.74637	TEGT	testis enhanced gene transcript (BAX inhibitor 1)	411	12q12-q13
Hs.23881	KRT7	keratin 7	156	12q12-q13
Hs.433996	CD63	CD63 antigen (melanoma 1 antigen)	78	12q12-q13
Hs.130730	AQP2	aquaporin 2 (collecting duct)	42	12q12-q13
Hs.406578	TUBA6	tubulin alpha 6	160	12q12-q14
Hs.242463	KRT8	keratin 8	589	12q13
Hs.406013	KRT18	keratin 18	341	12q13
Hs.274313	IGFBP6	insulin-like growth factor binding protein 6	120	12q13
Hs.250712	CACNB3	calcium channel, voltage-dependent, beta 3 subunit	107	12q13
Hs.181015	STAT6	signal transducer and activator of transcription 6, interleukin-4 induced	107	12q13
Hs.77690	RAB5B	RAB5B, member RAS oncogene family	107	12q13
Hs.1119	NR4A1	nuclear receptor subfamily 4, group A, member 1	107	12q13
Hs.76228	OS-9	amplified in osteosarcoma	88	12q13
Hs.376844	HNRPA1	heterogeneous nuclear ribonucleoprotein A1	881	12q13.1
12q13.12–12q15				
Hs.334842	**K-ALPHA-1**	**tubulin, alpha, ubiquitous**	**738**	**12q13.12**
Hs.24048	**FKBP11**	**FK506 binding protein 11, 19 kDa**	**126**	**12q13.12**
Hs.152982	**SCR59**	**hypothetical protein FLJ13117**	**107**	**12q13.12**
Hs.388645	**MGC14288**	**hypothetical protein MGC14288**	**107**	**12q13.12**
Hs.26613	**LOC113251**	**c-Mpl binding protein**	**107**	**12q13.12**
Hs.268189	**FLJ20436**	**hypothetical protein FLJ20436**	**46**	**12q13.12**
Hs.432699	**ASB8**	**ankyrin repeat and SOCS box-containing 8**	**42**	**12q13.12**
Hs.63525	**PCBP2**	**poly(rC) binding protein 2**	**291**	**12q13.12-q13.13**
Hs.77385	**MYL6**	**myosin, light polypeptide 6, alkali, smooth muscle and non-muscle**	**808**	**12q13.13**
Hs.76780	**PPP1R1A**	**protein phosphatase 1, regulatory (inhibitor) subunit 1A**	**381**	**12q13.13**
Hs.288771	**DKFZP586A0522**	**DKFZP586A0522 protein**	**211**	**12q13.13**
Hs.75884	**HCCR1**	**cervical cancer 1 protooncogene**	**160**	**12q13.13**
Hs.278270	**TEBP**	**unactive progesterone receptor, 23 kD**	**143**	**12q13.13**
Hs.13144	**ORMDL2**	**ORM1-like 2 (*S. cerevisiae*)**	**107**	**12q13.13**
Hs.6147	**TENC1**	**tensin like C1 domain-containing phosphatase**	**107**	**12q13.13**
Hs.9911	**FLJ11773**	**hypothetical protein FLJ11773**	**107**	**12q13.13**
Hs.104555	**NPFF**	**neuropeptide FF-amide peptide precursor**	**46**	**12q13.13**
Hs.32374	**FLJ34766**	***Homo sapiens* cDNA FLJ37066 fis, clone BRACE2015132, weakly similar to *Drosophila melanogaster* Oregon R cytoplasmic basic protein (deltex) mRNA.**	**107**	**12q13.2**
Hs.65377	**LOC92979**	**hypothetical protein BC009489**	**46**	**12q13.2**
Hs.181271	**COPZ1**	**CGI-120 protein**	**259**	**12q13.2-q13.3**
Hs.50984	**SAS**	**sarcoma amplified sequence**	**107**	**12q13.3**
Hs.156764	**RAP1B**	**RAP1B, member of RAS oncogene family**	**107**	**12q14**
Hs.283670	**CGI-119**	**CGI-119 protein**	**78**	**12q14.1-q15**
Hs.124813	**MGC14817**	**hypothetical protein MGC14817**	**185**	**12q14.2**
Hs.234734	**LYZ**	**lysozyme (renal amyloidosis)**	**107**	**12q14.3**
Hs.8752	**TMEM4**	**transmembrane protein 4**	**213**	**12q15**
Hs.433676	**KIAA0546**	**KIAA0546 protein**	**206**	**12q15**
Hs.79914	LUM	lumican	46	12q21.3-q22
Hs.78546	ATP2B1	ATPase, Ca++ transporting, plasma membrane 1	3641	12q21-q23
Hs.77054	BTG1	B-cell translocation gene 1, anti-proliferative	248	12q22
Hs.81118	LTA4H	leukotriene A4 hydrolase	160	12q22
Hs.24135	VEZATIN	transmembrane protein vezatin	107	12q23.1
Hs.108301	NR2C1	nuclear receptor subfamily 2, group C, member 1	46	12q23.1

Table 6 (continued)

UniGene ID	Symbol	Name	Tag per million	Cytoband
12q23.3–12q24.23				
Hs.32916	**NACA**	**nascent-polypeptide-associated complex alpha polypeptide**	**530**	**12q23-q24.1**
Hs.992	**PLA2G1B**	**phospholipase A2, group IB (pancreas)**	**181**	**12q23-q24.1**
Hs.1526	**ATP2A2**	**ATPase, Ca++ transporting, cardiac muscle, slow twitch 2**	**46**	**12q23-q24.1**
Hs.191450	**B3GNT4**	**UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyl-transferase 4**	**107**	**12q24**
Hs.287994	**NCOR2**	**nuclear receptor co-repressor 2**	**78**	**12q24**
Hs.12106	**MMAB**	**methylmalonic aciduria (cobalamin deficiency) type B**	**32**	**12q24**
Hs.173824	**TDG**	**thymine-DNA glycosylase**	**107**	**12q24.1**
Hs.9908	**NIFU**	**nitrogen fixation cluster-like**	**42**	**12q24.1**
Hs.293750	**ARPC3**	**actin-related protein 2/3 complex, subunit 3, 21 kDa**	**158**	**12q24.11**
Hs.197642	**RPC2**	**RNA polymerase III subunit RPC2**	**107**	**12q24.11**
Hs.75841	**C12orf8**	**chromosome 12 open reading frame 8**	**65**	**12q24.13**
Hs.180714	**COX6A1**	**cytochrome c oxidase subunit VIa polypeptide 1**	**42**	**12q24.2**
Hs.70582	**DDX54**	**DEAD box helicase 97 KDa**	**46**	**12q24.21**
Hs.80423	**PBP**	**prostatic binding protein**	**555**	**12q24.23**
Hs.184227	**FBXO21**	**F-box only protein 21**	**107**	**12q24.23**
Hs.136644	**WSB2**	**likely ortholog of mouse WD-40-repeat-containing protein with a SOCS box 2**	**69**	**12q24.23**
Hs.5120	**DNCL1**	**dynein, cytoplasmic, light polypeptide 1**	**46**	**12q24.23**
Hs.82689	**TRA1**	**tumor rejection antigen (gp96) 1**	**912**	**12q24.2-q24.3**
Hs.31638	RSN	restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)	107	12q24.3
Hs.47061	ULK1	unc-51-like kinase 1 (C. elegans)	46	12q24.3
Hs.75914	RNP24	coated vesicle membrane protein	1023	12q24.31
Hs.432714	VPS33A	vacuolar protein sorting 33A (yeast)	569	12q24.31
Hs.61976	DKFZp761B128	hypothetical protein DKFZp761B128	213	12q24.31
Hs.19523	FLJ38663	hypothetical protein FLJ38663	195	12q24.31
Hs.94308	RAB35	RAB35, member RAS oncogene family	107	12q24.31
Hs.406367	15E1.2	hypothetical protein 15E1.2	107	12q24.31
Hs.103561	ARL6IP4	ADP-ribosylation-like factor 6 interacting protein 4	92	12q24.31
Hs.77870	FLJ12750	hypothetical protein FLJ12750	46	12q24.31
Hs.9450	ZNF84	zinc finger protein 84 (HPF2)	107	12q24.33
Hs.127270	KIAA1545	KIAA1545 protein	107	12q24.33

While SAGE analysis is unique in its ability to quantify gene expression in a given tissue, there are several limitations for the analysis of the data [6]. For example, SAGE generates tags from the most 3′-*Nla*III restriction sites, but only on those mRNAs that have the site. In addition, tag to gene mapping is not completely definitive, as some tags correspond to several genes. SAGEmap provides two lists allowing the mapping of tags to genes. The most complete list describes all possibilities of tag to gene map and the second list provides a list of the most common and reliable tag to gene mapping. To clarify the analysis, we decided to use the more conservative second list. While more reliable, this limits the information obtained from this type of analysis since only about 60% of the tags from the three libraries could be identified with these conservative criteria. This number could be increased as new human ESTs are entered into the public databases. As a result of the large number of ESTs recently contributed by the Endocrine Pancreas Consortium (http://www.cbil.upenn.edu/EPConDB/index.shtml), several tags have been mapped that had no match to known genes in Unigene.

The difficulties in collecting healthy human tissues limited the quantity of islets and RNA that was obtained to build pancreatic islet SAGE libraries. Even though these libraries are large enough to provide valuable information about the most abundant transcripts in the pancreas, they were not large enough to analyze the relative expression levels of low abundance transcripts. Oligonucleotide arrays provide relative intensities of transcripts that can allow comparison of the expression level of one gene from individual to individual or under various conditions such as presence or absence of disease, metabolic conditions, or nutritional state for example. Unfortunately, the relative intensity of different transcripts within a tissue is difficult to interpret on an array because each gene has different hybridization characteristics. Thus, while it is far more difficult to create a SAGE library com-

Table 7. Islet transcripts expressed in the Type 2 diabetes mellitus regions on Chromosome 20q. The transcripts identified as being expressed in the islets are organized according to their chromosome locations for chromosome 20q. Regions thought to harbour Type 2 diabetes mellitus gene have been highlighted for the transcript located between 20q11.21 and 20q13.13

UniGene ID	Symbol	Name	Tag per million	Cytoband
Hs.274411	SCAND1	SCAN domain containing 1	185	20q11.1-q11.23
20q11.21–20q13.13				
Hs.386538	**HM13**	**histocompatibility (minor) 13**	**133**	**20q11.21**
Hs.352579	**C20orf178**	**chromosome 20 open reading frame 178**	**107**	**20q11.22**
Hs.241205	**PXMP4**	**peroxisomal membrane protein 4, 24 kDa**	**78**	**20q11.22**
Hs.401703	**C20orf52**	**chromosome 20 open reading frame 52**	**224**	**20q11.23**
Hs.334489	**SLA2**	**Src-like-adaptor 2**	**107**	**20q11.23**
Hs.177425	**DAP4**	**KIAA0964 protein**	**107**	**20q11.23**
Hs.168073	**C20orf188**	**chromosome 20 open reading frame 188**	**35**	**20q11.23**
Hs.5300	**BLCAP**	**bladder cancer associated protein**	**46**	**20q11.2-q12**
Hs.169487	**MAFB**	**v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)**	**349**	**20q11.2-q13.1**
Hs.225977	**NCOA3**	**nuclear receptor coactivator 3**	**175**	**20q12**
Hs.252189	**SDC4**	**syndecan 4 (amphiglycan, ryudocan)**	**160**	**20q12**
Hs.3407	**PKIG**	**protein kinase (cAMP-dependent, catalytic) inhibitor gamma**	**160**	**20q12-q13.1**
Hs.406532	**RPN2**	**ribophorin II**	**107**	**20q12-q13.1**
Hs.182238	**YWHAB**	**GW128 protein**	**107**	**20q13.1**
Hs.272168	**TDE1**	**tumor differentially expressed 1**	**46**	**20q13.1–13.3**
Hs.10590	**ZNF313**	**zinc finger protein 313**	**320**	**20q13.13**
Hs.3657	**ADNP**	**activity-dependent neuroprotector**	**107**	**20q13.13**
Hs.374523	GNAS	GNAS complex locus	1282	20q13.2-q13.3
Hs.2642	EEF1A2	eukaryotic translation elongation factor 1 alpha 2	46	20q13.3
Hs.182281	C20orf43	chromosome 20 open reading frame 43	160	20q13.31
Hs.352413	C20orf108	chromosome 20 open reading frame 108	46	20q13.31
Hs.119286	FLJ90166	hypothetical protein FLJ90166	46	20q13.32
Hs.165563	DNAJC5	DnaJ (Hsp40) homolog, subfamily C, member 5	320	20q13.33
Hs.31334	C20orf14	chromosome 20 open reading frame 14	160	20q13.33
Hs.39850	URKL1	uridine kinase-like 1	107	20q13.33
Hs.233952	PSMA7	proteasome (prosome, macropain) subunit, alpha type, 7	42	20q13.33

pared to performing a microarray experiment, SAGE analysis was necessary to provide us with relative abundance levels of transcripts in pancreatic islets. Regarding the precision of quantitative estimates of transcripts by SAGE, while more accurate for abundant transcripts, non-abundant transcripts with few tags yield only relative levels of expression. There are, however, no methods currently available that provide sufficient precision for expression levels of low abundance transcripts. Recently, Lynx Therapeutics (Lynx Therapeutics, Hayward, Calif., USA) has proposed a method called Massively Parallel Signature Sequencing (MPSS) measuring transcript abundance through a digital approach in which over a million transcripts can be counted simultaneously and which uses longer tags (17 bases) than in typical SAGE libraries. This new method could provide more extensive measurement of the expression of islet transcripts and provide information about low abundance transcripts as well [27].

The mechanical and enzymatic treatment required to isolate the different pancreatic fractions could induce expression of stress proteins not associated with normal tissue. This stress might be reflected in the current libraries through the very high expression level of several factors such as reg-1α and -1β, Hsp 70, and clusterin. Reg-1α is known to be expressed in both exocrine cells [16] and regenerating islets [15, 28, 29, 30]. The abundance of reg in pancreatic juice (10 to 14% of total protein) suggests that it plays an important role in exocrine pancreatic function. Its expression level is known to rise drastically in acute pancreatitis. While its expression was increased in regenerating pancreas, suggesting one of its names, its actual role as a proliferation factor for islets is currently unclear [31]. Even though clusterin has been described to be expressed in both exocrine [32] and endocrine cells [33] during cell injuries, its expression in our libraries seemed restricted to the islets.

A number of interesting observations were made from the merged islet libraries (Table 4). The expression of many secreted endocrine factors like insulin, glucagon, chromogranin B, islet amyloid polypeptide and pancreatic polypeptide were prominent. Intriguingly, the fourth most abundant islet transcript tag had no match in the SAGEmap databases and could not be

identified with the current tag to gene mapping available through SAGEmap. As EST databases expand, additional tags adjacent to *Nla*III sites close to poly-A tails will be identified and contributed to SAGEmap allowing identification of these unknown transcripts. An mRNA encoding a protein involved in prohormone processing and previously shown to be expressed in islets [19, 20], proprotein convertase subtilisin, also known as proSAAS, was an abundant islet transcript. Another abundant transcript represents secretory granule neurendocrine protein 1 or the 7B2 protein. Both proSAAS and 7B2 are proteins involved in hormone processing and are expressed in the brain as well as neurendocrine cells. ProSAAS is an inhibitor of prohormone convertase 1 activity [19], whereas 7B2 is a specific chaperone for proprotein convertase-2 that keeps the enzyme transiently inactive in vivo [34]. Mice homozygous for a null mutation in the 7B2 gene had no demonstrable PC2 activity and displayed hypoglycaemia, hyperproinsulinaemia, and hypoglucagonaemia [35].

Other abundant mRNAs identified in this study include recently described secretagogin, a cytoplasmic protein with six putative EF finger hand calcium-binding motifs [36, 37]. Its expression in pancreas is specific to the islets, and it is thought to be involved in KCl-stimulated calcium flux and the regulation of cell proliferation. The current study highlights the potential importance of this newly described islet protein, whose function has been little studied. Other less abundant identified islet transcripts included Protein tyrosine phosphatase, receptor-type N, also know as IA2 or islet cell antigen 512. It was discovered through the screening of a human islet library for clones encoding proteins reactive with sera from patients with Type 1 diabetes mellitus [38]. It was reported that 48% of Type 1 diabetes mellitus patients had antibodies directed to this islet antigen. The p57 (KIP2) protein is a genomically imprinted inhibitor of cyclin/Cdk complexes with an N-terminal CDK inhibitory domain highly similar to p21 (CIP1). Its implication in both sporadic cancers and Beckwith-Wiedemann syndrome makes it a tumor suppressor candidate. Since islet regeneration is a promising area of research, the abundance of this CDK inhibitor suggests that it could play an important role in this process. Interestingly this factor is located in an imprinted domain on chromosome 11p15 comprising IGF2 and H19, which seems to share the same tissue specific expression and imprinting pattern [39].

An analysis of the molecular functions represented in the islets showed that the most represented function corresponded to “binding activity” (Fig. 2). Approximately 35% of the molecular functions of the transcripts in the merged islet libraries are classified within this category. The factors contributing to this include insulin, transthyretin and glucagon for instance. The next most abundant islet function was “catalytic activities”, accounted for mostly by hormone processing enzymes, followed by the signal transducer activity (11% in the islets compared to 1.1% only for the exocrine library), suggesting the importance of responses to external signaling for islet function. Not surprisingly, the most common molecular function in the exocrine tissue (greater than 67%) corresponds to catalytic activities.

In summary, the results of these studies of human exocrine and endocrine pancreas libraries now provide in SAGEmap (http://www.ncbi.nlm.nih.gov/SAGE/) and the Endocrine Pancreas Consortium (http://www.cbil.upenn.edu/EPConDB/) transcript maps cataloging the relative levels of expression of the most abundant islet genes. This information should serve a number of useful functions, including the monitoring of relative abundance of transcripts during islet neogenesis, and the classification of altered patterns of gene expression during various stages of islet beta-cell failure in the development of diabetes. This data can also be analyzed in parallel with any kind of platform assessing RNA abundance through the RNA Abundance Database platform used on the Endocrine Pancreas Consortium web-site.

Additionally, the tedious job of sifting through hundreds of genes in the analysis of linkage peak regions in the analysis of the genetic basis for Type 2 diabetes could be facilitated by an addition to the currently used candidate gene approach, where one of the criteria for selection of a candidate would be relatively high level of expression in human pancreatic islets. Similar analyses for chromosomal regions identified to harbour Type 1 diabetes mellitus could be conducted as well.

Acknowledgements. The authors would like to thank the Washington University Genome Sequencing Center’s EST Sequencing Group (Darwin) including Mr. J. Martin, Mr. T. Wylie, and Mr. M. Dante. Mr. G. Skolnick was responsible for preparing the manuscript. We would also like to acknowledge the Human Islet Core (B. Olack and T. Mohanakumar) of Diabetes Research and Training Center at Washington University for its support. This work was supported in part by National Institutes of Health grant DK99007 for a project entitled “Functional Genomics of the Endocrine Pancreas” that included collaborations with D. Melton (Harvard) and K. Kaestner and C. Stoeckert (Univ. of Pennsylvania).

References

1. Cavaghan MK, Ehrmann DA, Polonsky KS (2000) Interactions between insulin resistance and insulin secretion in the development of glucose intolerance. J Clin Invest 106:329–333
2. Neophytou PI, Muir EM, Hutton JC (1996) A subtractive cloning approach to the identification of mrnas specifically expressed in pancreatic beta-cells. Diabetes 45:127–133
3. Ferrer J, Schoor K, Wasson J, Kipnis D (1994) Identification of 5 human islet-specific genes by differential display of mRNA. Diabetes 43 (Suppl 1):6A

4. Minami K, Yano H, Miki T, et al. (2000) Insulin secretion and differential gene expression in glucose-responsive and -unresponsive MIN6 sublines. Am J Physiol Endocrinol Metab 279:E773–E781
5. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487
6. Velculescu VE, Vogelstein B, Kinzler KW (2000) Analysing uncharted transcriptomes with SAGE. Trends Genet 16:423–425
7. Yang X, Pratley RE, Tokraks S, Bogardus C, Permana PA (2002) Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians. Diabetologia 45:1584–1593
8. De Chaldee M, Gaillard MC, Bizat N, et al. (2003) Quantitative assessment of transcriptome differences between brain territories. Genome Res 13:1646–1653
9. Patino WD, Mian OY, Shizukuda Y, Hwang PM (2003) Current and future applications of SAGE to cardiovascular medicine. Trends Cardiovasc Med 13:163–168
10. Olack BJ, Swanson CJ, Howard TK, Mohanakumar T (1999) Improved method for the isolation and purification of human islets of langerhans using liberase enzyme blend. Hum Immunol 60:1303–1309
11. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185
12. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194
13. Diehn M, Sherlock G, Binkley G, et al. (2003) Source: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res 31:219–223
14. Ashburner M, Ball CA, Blake JA, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29
15. Sanchez D, Baeza N, Blouin R, et al. (2000) Overexpression of the reg gene in non-obese diabetic mouse pancreas during active diabetogenesis is restricted to exocrine tissue. J Histochem Cytochem 48:1401–1410
16. Graf R, Schiesser M, Scheele GA, et al. (2001) A family of 16-kDa pancreatic secretory stress proteins form highly organized fibrillar structures upon tryptic activation. J Biol Chem 276:21028–21038
17. Pin CL, Rukstalis JM, Johnson C, Konieczny SF (2001) The bhlh transcription factor mist1 is required to maintain exocrine pancreas cell organization and acinar cell identity. J Cell Biol 155:519–530
18. Itoh N, Hanafusa T, Miyagawa J, et al. (1992) Transthyretin (prealbumin) in the pancreas and sera of newly diagnosed type I (insulin-dependent) diabetic patients. J Clin Endocrinol Metab 74:1372–1377
19. Fricker LD, McKinzie AA, Sun J, et al. (2000) Identification and characterization of prosaas, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. J Neurosci 20:639–648
20. Feng Y, Reznik SE, Fricker LD (2001) Distribution of prosaas-derived peptides in rat neuroendocrine tissues. Neuroscience 105:469–478
21. De Vos P (2002) —to: B.M. Kim et al.: Clusterin expression during regeneration of pancreatic islet cells in streptozotocin-induced diabetic rats. Diabetologia 44:2192–2202. Diabetologia 45:931–932
22. Leiter EH (1997) Carboxypeptidase E and obesity in the mouse. J Endocrinol 155:211–214
23. Azzoni C, Yu JY, Baggi MT, et al. (1992) Studies on co-localization of 7B2 and pancreatic hormones in normal and tumoural islet cells. Virchows Arch A Pathol Anat Histopathol 421:457–466
24. Portela-Gomes GM, Stridsberg M (2002) Region-specific antibodies to chromogranin b display various immunostaining patterns in human endocrine pancreas. J Histochem Cytochem 50:1023–1030
25. Stern MP (2002) The search for type 2 diabetes susceptibility genes using whole-genome scans: an epidemiologist's perspective. Diabetes Metab Res Rev 18:106–113
26. Scearce LM, Brestelli JE, McWeeney SK, et al. (2002) Functional genomics of the endocrine pancreas: The pancreas clone set and pancChip, new resources for diabetes research. Diabetes 51:1997–2004
27. Brenner S, Johnson M, Bridgham J, et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634
28. Bartoli C, Gharib B, Giorgi D, Sansonetti A, Dagorn JC, Berge-Lefranc JL (1993) A gene homologous to the reg gene is expressed in the human pancreas. FEBS Lett 327:289–293
29. Terazono K, Yamamoto H, Takasawa S, et al. (1988) A novel gene activated in regenerating islets. J Biol Chem 263:2111–2114
30. Baeza N, Sanchez D, Christa L, Guy-Crotte O, Vialettes B, Figarella C (2001) Pancreatitis-associated protein (hip/pap) gene expression is upregulated in nod mice pancreas and localized in exocrine tissue during diabetes. Digestion 64:233–239
31. Graf R, Schiesser M, Lussi A, Went P, Scheele GA, Bimmler D (2002) Coordinate regulation of secretory stress proteins (psp/reg, pap i, pap ii, and pap iii) in the rat exocrine pancreas during experimental acute pancreatitis. J Surg Res 105:136–144
32. Xie MJ, Motoo Y, Su SB, Sawabu N (2001) Expression of clusterin in pancreatic acinar cell injuries in vivo and in vitro. Pancreas 22:126–134
33. Park IS, Che YZ, Bendayan M, Kang SW, Min BH (1999) Up-regulation of clusterin (sulfated glycoprotein-2) in pancreatic islet cells upon streptozotocin injection to rats. J Endocrinol 162:57–65
34. Mbikay M, Seidah NG, Chrétien M (2001) Neuroendocrine secretory protein 7B2: structure, expression and functions. Biochem J 357:329–342
35. Vishnuvardhan D, Connolly K, Cain B, Beinfeld MC (2000) PC2 and 7B2 null mice demonstrate that PC2 is essential for normal pro-CCK processing. Biochem Biophys Res Commun 273:188–191
36. Gartner W, Lang W, Leutmetzer F, Domanovits H, Waldhausl W, Wagner L (2001) Cerebral expression and serum detectability of secretagogin, a recently cloned EF-hand Ca(2+)-binding protein. Cereb Cortex 11:1161–1169
37. Wagner L, Oliyarnyk O, Gartner W, et al. (2000) Cloning and expression of secretagogin, a novel neuroendocrine- and pancreatic islet of Langerhans-specific Ca2+-binding protein. J Biol Chem 275:24740–24751
38. Rabin DU, Pleasic SM, Shapiro JA, et al. (1994) Islet cell antigen 512 is a diabetes-specific islet autoantigen related to protein tyrosine phosphatases. J Immunol 152:3183–3188
39. Hatada I, Mukai T (1995) Genomic imprinting of p57kip2, a cyclin-dependent kinase inhibitor, in mouse. Nat Genet 11:204–206