Fig.1

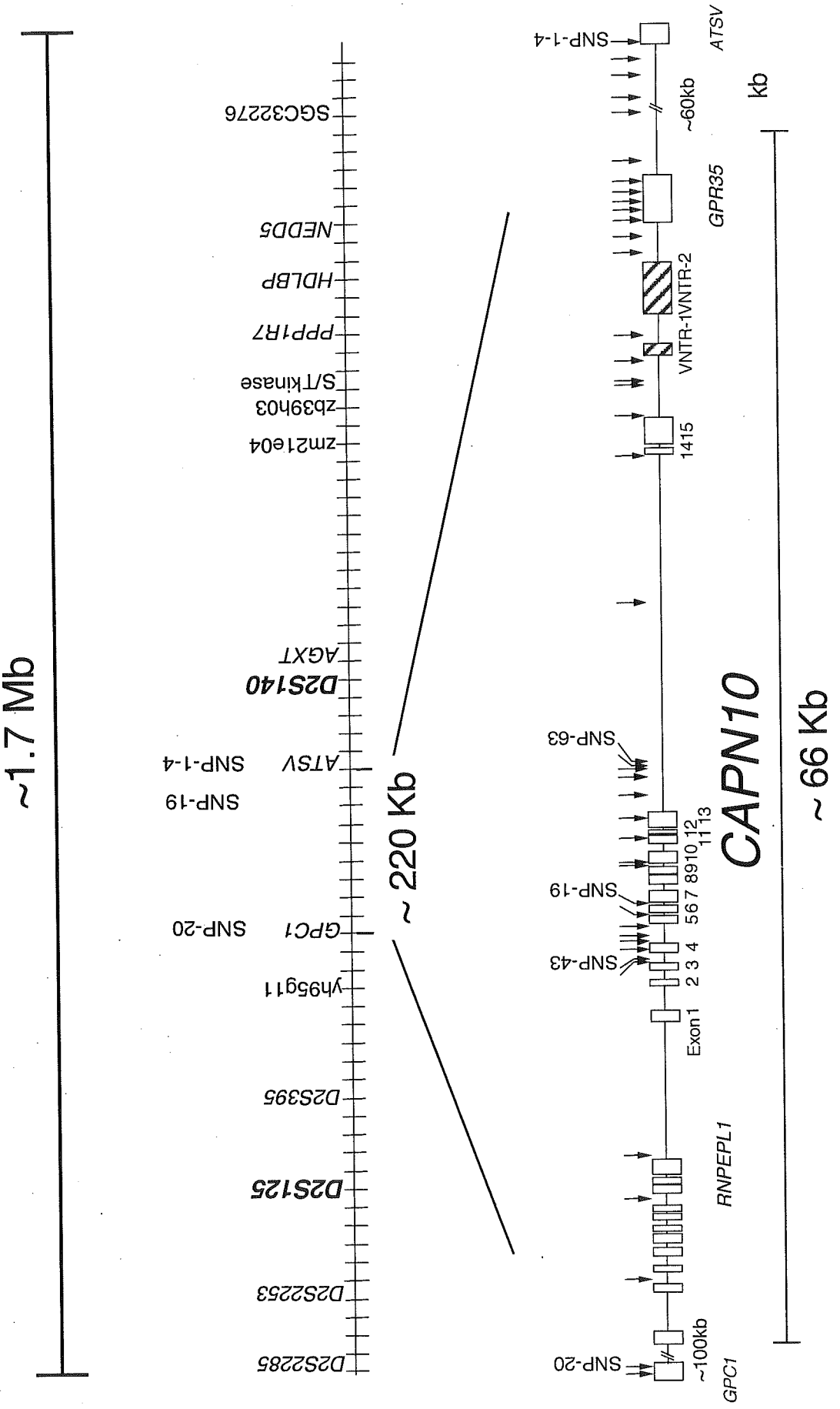
~1.7 Mb
SGC32276
NEDD5
HDLBP
PPP1R7
S/Tkinase
zb39h03
zm21e04
AGXT
D2S140
ATSV
SNP-1-4
SNP-19
~ 220 Kb
GPC1
SNP-20
yh95g11
D2S395
D2S125
D2S2253
D2S2285
SNP-1-4
ATSV
~60kb
GPR35
VNTR-1VNTR-2
1415
SNP-63
SNP-19
SNP-43
1 2 3 4 5 6 7 8 9 10 11 12 13
CAPN10
Exon 1
RNPEPL1
~100kb
SNP-20
GPC1
kb
~ 66 Kb

Fig. 2

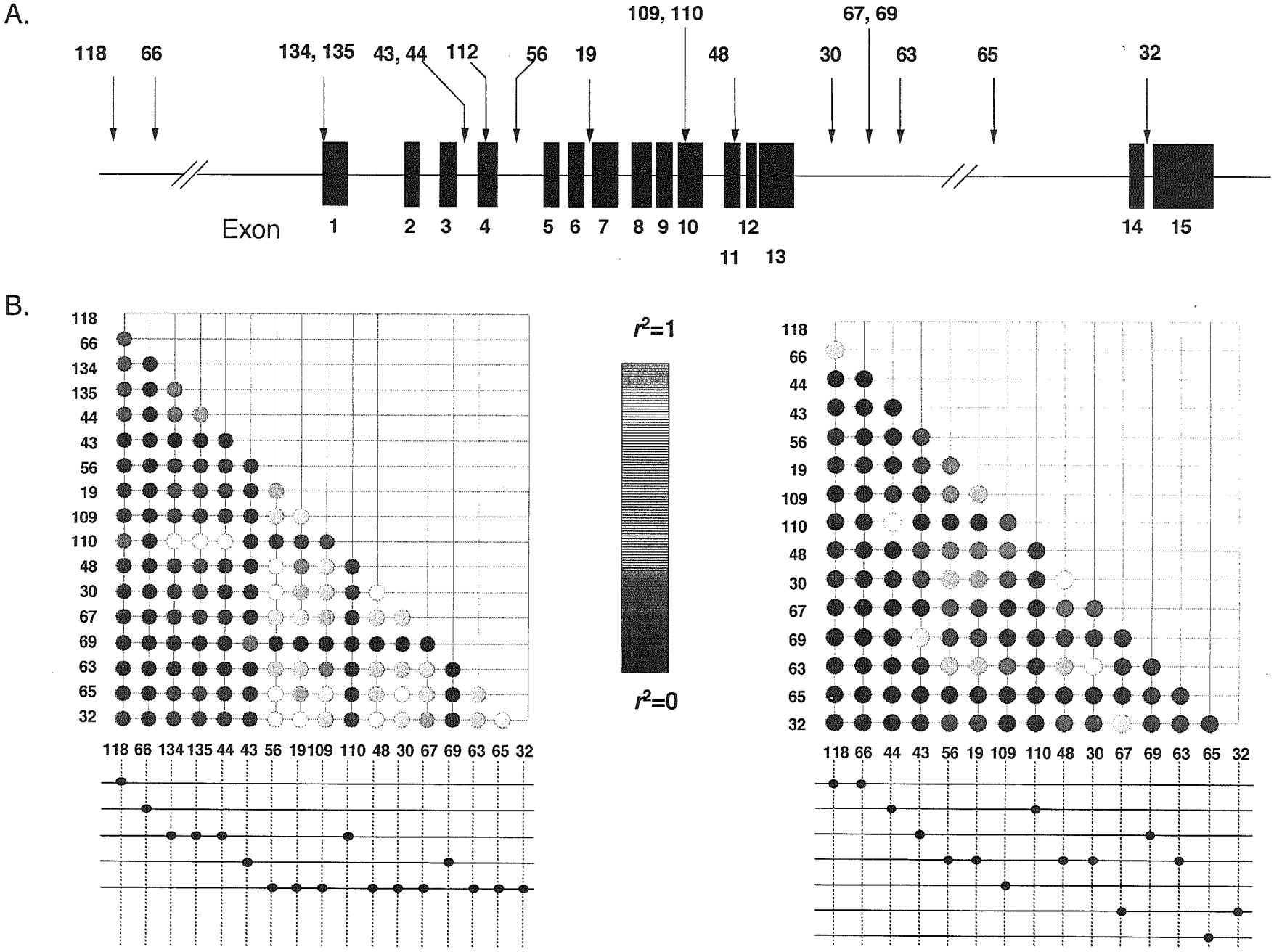
A.
118
66
134, 135
43, 44
112
56
19
109, 110
48
30
67, 69
63
65
32
Exon
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
B.
118
66
134
135
44
43
56
19
109
110
48
30
67
69
63
65
32
118 66 134 135 44 43 56 19 109 110 48 30 67 69 63 65 32
r^2=1
r^2=0
118
66
44
43
56
19
109
110
48
30
67
69
63
65
32
118 66 44 43 56 19 109 110 48 30 67 69 63 65 32

Fig. 3

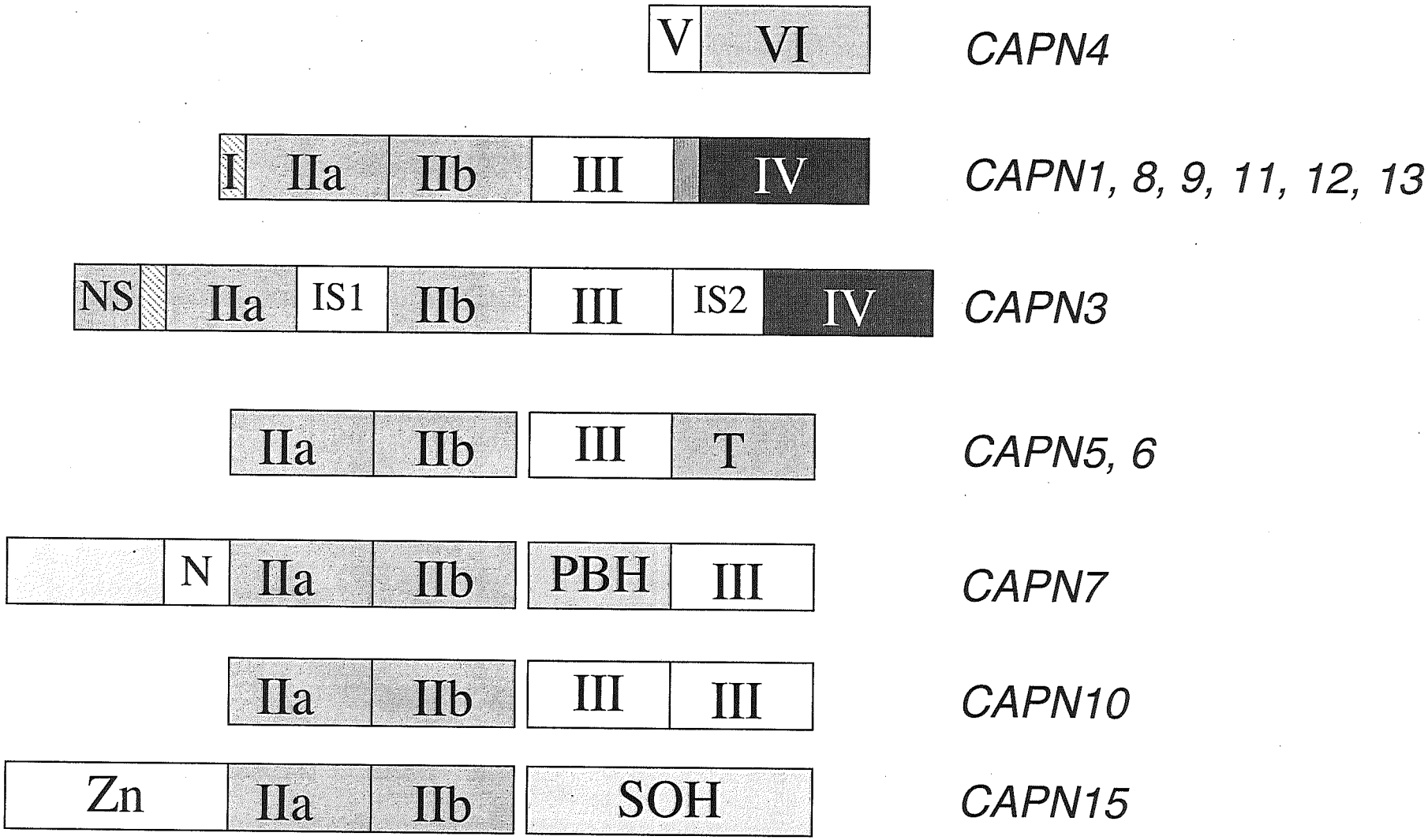
V
VI
CAPN4
I
IIa
IIb
III
IV
CAPN1, 8, 9, 11, 12, 13
NS
IIa
IS1
IIb
III
IS2
IV
CAPN3
IIa
IIb
III
T
CAPN5, 6
N
IIa
IIb
PBH
III
CAPN7
IIa
IIb
III
III
CAPN10
Zn
IIa
IIb
SOH
CAPN15

Table1

Population	Genotype	Odds Ratio	Phenotype
Mexican American	112/121	3.02(1.37-6.64)	
Pima Indian	SNP-43 G/G		Glucose utilization ↓
British/Irish Whites	111/111	2.04(1.22-3.39)	
	2111/2111	2.52(1.06-5.97)	
	2111/1111	2.36(1.19-4.66)	A.I.R ↓
	112/121		HOMA-R ↑
Samoans	112/121	1.42(0.68-2.98)	
Utah-Caucasian	111/221	1.48(1.06-1.91)	Insulin AUC ↑
	SNP-19 -63		HOMA-R ↑
African-American	SNP-43 G/G	1.38(1.04-1.83)	Insulin AUC ↑
	112/121	2.18(1.06-4.45)	PCOS Odds Ratio ↑
Spanish	SNP-44 CC,TC	2.57 (1.22-5.44)	PCOS Odds Ratio ↑
Polish	121/121	1.93(1.03-3.54)	
Finnish	1121/1121	1.93 (1.07-3.47)	Fasting Insulin HOMA-R FFA ↑
Japanese	112/121		FFA ↑
	121/121		Protective against T2DM

Gene expression profile in rat pancreatic islet and RINm5F cells

H Wang[1], Y Horikawa[1,2,3], L Jin[1], T Narita[1], S Yamada[1], N Shihara[1,2], K Tatemoto[4], M Muramatsu[3], T Mune[3] and J Takeda[1,2,3]

[1]Laboratory of Molecular Genetics, Department of Cell Biology, Institute for Molecular and Cellular Regulation, Gunma University, Gunma, Japan

[2]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), Kawaguchi, Japan

[3]Department of Diabetes and Endocrinology, Division of Molecule and Structure, Gifu University School of Medicine, Gifu, Japan

[4]Laboratory of Peptide and Protein Research, Department of Molecular Physiology, Institute for Molecular and Cellular Regulation, Gunma University, Gunma, Japan

(Requests for offprints should be addressed to Yukio Horikawa, Department of Diabetes and Endocrinology, Gifu University School of Medicine, 1-1 Yanagido, Gifu-city, Gifu 500-1194, Japan; Email: yhorikaw@cc.gifu-u.ac.jp)

Abstract

To clarify tissue-specificity of pancreatic β cells, comparison of mRNA expression in various conditions of the tissue of multiple organisms is important. Although the developed methodologies for mRNA monitoring such as microarray, rely on the growth of dbEST (database of expressed sequence tag), a large number of unknown genes in the genome, especially in the rat, have not been shown to be expressed. In this study, we have established the first database of ESTs from rat pancreatic islet and RINm5F cells. Two cDNA libraries were constructed using mRNAs from rat pancreatic islet and RINm5F cells to cover a wider spectrum of expressed genes. Over 40 000 clones were randomly selected from the two libraries and partially sequenced. The sequences obtained were subjected to BLAST database analyses. This large-scale sequencing generated 40 710 3′-ESTs. Clustering analysis and homology search of nucleotide and peptide databases using both 3′- and 5′-ESTs revealed 10 406 non-redundant transcripts representing 4078 known genes or homologs and 6328 unknown genes. To confirm actual expression, the unknown sequences were further subjected to dbEST search, resulting in the identification of 5432 significant matches to those from other sources. Interestingly, of the remaining sequences showing no match, 779 were found to be encoded by exon–intron organization in the corresponding genomic sequences, suggesting that these are newly found as actually expressed in this study. Since many genes are up- or down-regulated in differing conditions, applications of the expression profile should facilitate identification of the genes involved in cell-specific functions in normal and disease states.

Journal of Molecular Endocrinology (2005) **35**, 1–12

Introduction

Pancreatic islets play the critical role in the regulation of blood glucose by secreting hormones from endocrine cells that differentiate from common progenitor cells during fetal development (postnatal origin of β-cell replenishment remains controversial) (Edlund 2002, Bonner-Weir & Sharma 2002, Dor *et al.* 2004). Functional defects of β-cells lead to the development of diabetes mellitus. Since a number of genes are involved in the pathogenesis of impaired insulin secretion, it is important to characterize the expression profile of a set of genes that endow β-cells with the tissue-specific functions of insulin synthesis and secretion.

Recent genome projects demonstrated a similar number of 30 000–40 000 genes in human and mouse chromosomes, including ~27 000 protein-encoding transcripts for which there was strong corroborating evidence and ~10 000 computationally derived genes with weak supporting evidence (International Human Genome Sequencing Consortium 2001, Venter *et al.* 2001, Mouse Genome Sequencing Consortium 2002). Comparison with the transcriptome revealed almost all of the human genes known to be expressed to have orthologues in the mouse genome. The other putatively novel genes in the genome were detected using computer algorithms for transcript prediction. To estimate the accuracy of the power of new gene detection, the results of the gene annotation done by the two human genome efforts were previously compared (Hogenesch *et al.* 2001). Surprisingly, although a similar number of the total genes was demonstrated, there is little agreement regarding the new genes predicted by the two projects, suggesting that a significant fraction of tissue-restricted transcripts for novel genes remain undiscovered, possibly due to limitations in the computer prediction methods.

As expression analysis of the genes in multiple organisms becomes a major focus in the new era of biology, functional genomics will rely largely on the vast

Journal of Molecular Endocrinology (2005) **35**, 1–12
0952–5041/05/035–001

DOI: 10.1677/jme.1.01738
Online version via http://www.endocrinology-journals.org

sources of subsets of partial cDNA sequences from various tissues that have proven enormously valuable and are deposited as expressed sequence tags (ESTs) in the public databases. The Endocrine Pancreas Consortium has recently constructed human and mouse cDNA libraries from various conditions of endocrine pancreas and generated over 100 000 ESTs (Bernal-Mizrachi *et al.* 2003). We also have collected ~20 000 ESTs from human normal pancreatic islets and islet tumors, resulting in the identification of ~3000 new genes expressed in the islets (Takeda *et al.* 1993, Jin *et al.* 2003). Such systematic sequencing efforts complement each other and should improve the various methodologies including DNA microarray technology (Scearce *et al.* 2002) for monitoring differential gene expression in normal and disease states. In addition, the laboratory rat is an indispensable model organism of human diseases, providing a useful tool in experimental medicine and drug discovery. As various spontaneous diabetic rats such as the GK and OLETF rats and the experimental streptozotocin-induced diabetic rat are widely used in pancreatic islet studies, it is important to establish an additional source of rat expressed sequences. However, although ~26 millions of ESTs have so far been deposited in the database (dbEST release 031105), approximately 40% of which are derived from human and mouse, only 2·6% of the sequences are from rat, and none are from pancreatic islets except for the present deposition. In this study, toward elucidation of the entire transcriptome in rat pancreatic islets, we have made two cDNA libraries, one from rat normal pancreatic islets and the other from RINm5F tumor cells having undergone less differentiation (Gazdar *et al.* 1980, Philippe *et al.* 1987), and performed a large-scale collection of ESTs. Since a number of the genes are up- or down-regulated in different conditions, a collection of ESTs from these distinct cDNA sources should more effectively cover a wider spectrum of expressed genes, generating a larger pool of non-redundant sequences. In addition, since the insulin content of the less-differentiated RINm5F cells used in this study was previously found to be much lower than that of normal islets (Kayo *et al.* 1997), direct comparison of the expression profiles of the two cDNA libraries should facilitate the identification of the genes involved in insulin synthesis and secretion as well as in β-cell differentiation and tumorigenesis.

Materials and methods

Preparation of rat pancreatic islets

Pancreatic islets were prepared from male Wister rats by a collagenase digestion method as described previously (Ma *et al.* 1996). Briefly, under pentobarbital anesthesia, the pancreas was distended by an injection of 10 ml Hank's solution containing 0·3 mg/ml collagenase (type XI, Sigma-Aldrich, StLouis, MO, USA). Islets were separated by the Ficoll (Amersham, Piscataway, NJ, USA) density gradient method with four layers (27%, 23%, 20·5%, and 11% of Ficoll dissolved in Hank's solution). After centrifugation at 450 ***g*** for 15 min, pancreatic islets were concentrated at the interface between the 11% and 20·5% Ficoll layers. Islets were then harvested by a pick-up method under a stereomicroscope. Purity of the islets was estimated to be ~99% by counting the cells immunoreactive to insulin and glucagon antibodies after trypsin treatment of a fraction of the islets collected. The high purity also could be estimated by the frequency of cDNA for major exocrine molecules, such as α-amylase, in the entire islet ESTs identified, as described below.

Large-scale cDNA sequencing

Two unidirectional cDNA libraries were constructed in the Uni-ZAP XR vector (Stratagene, La Jolla, CA, USA) using mRNAs from rat normal pancreatic islets and RINm5F cells. A large set of plasmid DNAs for sequencing was prepared as described (Takeda *et al.* 1993, Jin *et al.* 2003). Briefly, the cDNA libraries were excised *en masse* from the λ phage into phagemid particles using the ExAssist phage system (Stratagene), and subsequently transfected into *E. coli* SOLR (Stratagene) for conversion to plasmid forms. Plasmid DNAs were extracted from *E. coli* colonies randomly selected from LB-Amp plates using the Biomek 2000 mini-prep system (Beckman, Fullerton, CA, USA). Single-pass DNA sequencing from the 3′-end of the inserts was performed using a BigDye Terminator Cycle Sequencing FS Ready Reaction Kit and DNA Sequencer model 3700 (Applied Biosystems, Foster City, CA, USA). Vector sequences were removed from the results using Assembly LIGN software (Oxford Molecular Group PLC, Oxford, UK). Quality assessment of the sequences obtained was performed using PE Sequencing Analysis 3·3 software (Applied Biosystems).

Database analysis of rat pancreatic islet and RINm5F ESTs

We compared a total of ~40 000 sequences from rat pancreatic islet and RINm5F cells with non-redundant nucleotide and peptide sequences extracted *in silico* from databases at the National Center for Biotechnology Information (NCBI). Before comparisons, interspersed repetitive sequences such as LINEs were unmasked and removed from the pool using software RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html). To assemble sequences sharing a stretch of nucleotide identity, the LaboServer system (World fusion, Tokyo, Japan) was applied to make contigs.

Representative sequences from each contig then were subjected to BLASTN analysis for sequence homology at nucleotide level against a merged database by the Kiroku program (World fusion). If a query sequence shared over 95% nucleotide identity and showed a score of more than 400 with any sequences in the database, they were grouped together. The clones without significant match to known sequences in the nucleotide database were re-sequenced from the other end to compare the sequences with those in the peptide database at NCBI using BLASTX program (Altschul *et al.* 1997), which conceptually translates the query sequence in all six reading frames for comparison. The ESTs identical or highly similar to functionally annotated genes were first classified into seven major categories according to the general functions of the proteins encoded, and then further classified into subcategories according to their specific functions.

Semi-quantitative RNA expression analysis

To ascertain the level of mRNA expression of the ESTs from rat normal islets and RINm-5F cells, real-time quantitative reverse transcriptase (RT)-PCR was carried out using an ABI PRISM 7900HT Sequence Detection System (Applied Biosystem). Total RNA was extracted from pooled islets isolated from normal rats and RINm5F cells, using an RNeasy Mini Preparation Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. TaqMan primers and probes were designed using Primer Express software purchased from Applied Biosystems. TaqMan reactions were performed in a reaction volume of 20 µl using components supplied in a TaqMan PCR reagent kit. Each reaction consisted of 10 µl TaqMan Universal Master Mix, 900 nM of each amplification primer, and 250 nM corresponding TaqMan probe. Each sample was run for an initial 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles at 95 °C for 15 s and at 60 °C for 1 min. Amplification data were collected by the 7900HT Sequence Detector and analyzed using Sequence Detection System software. The RNA concentration was determined from the threshold cycle at which fluorescence is first detected, cycle number being inversely related to RNA concentration.

In situ hybridization

ESTs showing marked differences in frequency between islet and RINm5F cells were subjected to analysis of mRNA distribution in the pancreas by *in situ* hybridization. Paraffin embedded blocks and sections of normal rat pancreas for *in situ* hybridization (ISH) were obtained from GENOSTAFF, Inc. (Tokyo, Japan). The pancreases of male Wistar rats 8 weeks old (CREA Tokyo, Japan, Inc.) were dissected after perfusion, fixed by Tissue Fixative (GENOSTAFF, Cat No.STF-01), and embedded in paraffin by the proprietary procedures.

ISH was performed with the Ventana HX system (Ventana Medical Systems, Inc., Tucson, AZ, USA). Entire EST inserts were amplified by PCR using ExTaq (TaKaRa, Kyoto, Japan) in a 50 µl reaction mixture using M13 forward and reverse primers. Amplification was performed as follows: 3 min at 94 °C for initial denaturation, 35 cycles of 94 °C denaturing for 30 s 60 °C annealing for 30 s, and 72 °C extension for 1 min, followed by a final extension at 72 °C for 10 min. Quality and quantity of the purified PCR product was confirmed by 1·2% agarose gel electrophoresis. Antisense and sense RNA probes were labeled using the T7/T3 digoxigenin RNA labeling kit (Roche Diagnostics, Indianapolis, IN, USA), according to the manufacturer's instructions. Sections were pre-treated and hybridized with a Ventana RiboMap kit (Ventana Medical Systems) on the automated Ventana HX system Discovery. Detection of hybrids was performed with a digoxigenin nucleic acid detection kit (Boehringer Mannheim, Germany) following the manufacturer's instructions. Sections were then dehydrated through an ethanol series (80, 90 and 100% ethanol, for 1 min each) and washed for 1 min in xylene before mounting in malinol mounting medium (Muto Pure Chemicals Ltd, Tokyo, Japan).

Results and discussion

The expression profile of genes in pancreatic islets of experimental animals will greatly complement human studies of functional genomics of the tissue and the genetics of its disease states. This study both establishes a first molecular inventory of rat pancreatic islets and reveals a number of novel genes, the expression of which has not previously been described. These should provide important insights into the entire transcriptome of endocrine pancreas and be an immensely valuable aid to the improvement of genomic annotation.

Large-scale collection of ESTs from rat pancreatic islet and RINm5F cells

A total of 40 710 clones randomly selected from the two cDNA libraries were partially sequenced from the 3′-end. Sequences containing less than 1% ambiguous bases longer than 200 bp were subjected to BLASTN database search. Contaminated genomic sequences, e.g. repetitive sequences, (1967 clones) and mitochondrial DNAs (4633 clones), were removed from the pool of sequences, resulting in a collection of 34 110 ESTs comprising 22 310 known and 11 800 unknown sequences. Our previous study of 1000 ESTs from human pancreatic islets, the purity of which was

Table 1 Redundancy of 3′-ESTs from rat pancreatic islet and RINm5F cells

	Islet	RINm5F	Islet & RINm5F
Redundancy			
>1000	1	0	1
101–1000	1	3	7
51–100	3	6	15
31–50	6	12	28
21–30	21	31	69
11–20	87	138	253
6–10	228	299	533
3–5	1039	1113	1878
2	1556	1593	2075
1	3088	3065	5547

The 40 710 ESTs were generated by sequencing cDNA clones from the 3′-ends. After clustering analysis, a total of 6030 and 6260 independent groups were obtained from pancreatic islet and RINm5F cells, respectively.

estimated to be ~90% by microscopic examination and protein analysis, identified 13, 12, 6, and 9 ESTs for major exocrine genes for α-amylase, elastase, pancreatic lipase, and trypsinogen respectively (Takeda *et al.* 1993). Since only 2, 8, 2, and 5 clones for these exocrine genes respectively, were found in the ~20 000 EST sequences of this study, the possibility of contamination from exocrine cells appears to be quite negligible (~0·1%). This estimation of purity is consistent with that of the protein analysis described above. Because large-scale sequencing based on random isolation of clones generates high redundancy, clustering analysis was performed to assemble the sequences into non-redundant sequence groups. A total of 6030 and 6260 independent groups were obtained from pancreatic islet and RINm5F cells respectively, and the pattern of redundancy was similar between the two sources (Table 1 & 2). Together, 10 406 non-redundant sequences comprising 4859 clusters of sequences and 5547 singletons were obtained representing 4078 known genes and 6328 unknown genes. Since only 1896 distinct genes (18%) were found to overlap in the normal islets and RINm5F cells, this large-scale sequencing using two distinct cDNA sources was quite effective in identifying a larger number of non-redundant sequences. Studies of the number of different mRNA sequences in a cell suggest that a typical higher eukaryotic cell synthesizes 10 000 to 20 000 different proteins (Alberts *et al.* 1994), so this approach covered at least 50% of the possible protein-coding genes. As in similar large-scale cDNA sequencing studies carried out in other tissues, about 50% of the clones obtained are derived from genes not functionally annotated. These unknown clones were re-sequenced from the 5′-end, and the 5512 clones sequenced successfully were also subjected to database search. As a result, 1404 sequences representing 502 distinct transcripts showed perfect identity to known genes, so the 3′-end sequences of these clones clearly are not contained in the cDNA sequences deposited in the nucleotide databases. These ESTs were then assigned to the known group. All representative EST sequences obtained from each cluster were deposited in the public database to be freely available to all researchers (DDBJ accession No. BP464981–BP504629).

Table 2 Summary of non-redundant 3′-ESTs and/or 5′-ESTs

	Known	Unknown	Total
Islet only	1165	2981	4146
RINm5F only	1569	2807	4376
Islet & RINm5F	1344	540	1884
Total	4078	6328	10 406

The 6030 and 6260 non-redundant sequences from pancreatic islet and RINm5F cells, respectively, were subjected to BLASTN database search at the National Center for Biotechnology Information.

Characterization of known genes in pancreatic islet and RINm5F cells

The ESTs showing identity or high similarity to known genes were classified into seven major categories on the basis of putative general functions of the protein encoded, as described previously (categories: cell division, cell signaling/communication, cell structure/motility, cell/organism defense, gene/protein expression, metabolism, and unclassified). In total, 3951 out of 4078 known genes were represented in the classified data set (online supplement). The largest category of genes was gene/protein expression (26·4%). Successively smaller categories were cell signaling and communication (19·0%), metabolism (16·8%), cell structure/motility (7·7%), cell/organism defense (7·5%), and cell division (5·6%). ESTs lacking sufficient information to be classified constituted the remainder, unclassified (16·9%). To further analyze the molecular complexity, each major category was subdivided according to the putative specific functions of the proteins (Table 3, also see online supplement). For example, the largest category, gene/protein expression, was subdivided into eight subgroups. Of these, transcription factor constituted the largest number of non-redundant genes (416 genes by 1209 ESTs). The transcription factors include PDX-1, BETA2/NeuroD, HNF-4α, Nkx-2·2, Nkx-6·1, and Isl-1 etc, all of which are important for pancreatic development and islet-specific functions, and the first three of which are the causal genes for monogenic forms of diabetes, MODY4, MODY6, and MODY1 (Fajans *et al.* 2001). The other genes for transcription factors also are plausible candidates for diabetogenes or genes responsible for β-cell specific functions.

In this study, 60·8% of the non-redundant ESTs did not match any of the known genes in the nucleotide

Table 3 Functional categories of proteins encoded by non-redundant ESTs

Functional category	Subcategory	Islet only	RINm5F only	Islet & RINm5F	Total
Cell division	General	18	13	21	52
	DNA synthesis	1	18	5	24
	Apoptosis	13	15	18	46
	Cell cycle	13	36	19	68
	Chromosomal structure	8	15	9	32
	Subtotal	53	97	72	222
Cell signaling	Cell adhesion	20	18	12	50
	Channel	15	21	19	55
	Effectors	23	24	24	71
	Hormone	31	28	25	84
	Intracellular transducers	50	53	62	165
	Metabolism	2	1	3	6
	Protein modification	56	43	55	154
	Receptor	60	57	51	168
	Subtotal	257	245	251	753
Cell structure	General	9	17	16	42
	Contractile protein	9	8	7	24
	Cytoskeletal	23	26	39	88
	Extracellular matrix	24	10	8	42
	Microtubule-associated	17	23	17	57
	Vesicular transport	15	19	19	53
	Subtotal	97	103	106	306
Cell defense	Homeostasis (general)	21	16	20	57
	DNA repair	11	17	11	39
	Carrier protein	22	24	22	68
	Stress response	9	12	14	35
	Immunology	47	23	27	97
	Subtotal	110	92	94	296
Gene expression	RNA polymerase	6	6	12	24
	RNA processing	23	70	62	155
	Transcription factor	136	173	107	416
	Targeting	37	49	65	151
	Protein turnover	34	28	47	109
	Ribosomal proteins	17	20	87	124
	tRNA synthesis	1	9	8	18
	Translation factor	9	12	29	50
	Subtotal	263	367	417	1047
Metabolism	General	4	4	5	13
	Amino acid	12	15	18	45
	Cofactors	1	3	1	5
	Energy	31	37	56	124
	Lipid	39	51	49	139
	Nucleotide	18	30	13	61
	Protein modification	7	25	20	52
	Sugar	19	81	30	130
	Transport	25	36	31	92
	Subtotal	156	282	223	661
Unclassified		197	300	169	666
Total No. of unique genes		1133	1486	1332	3951

The ESTs showing identity or high similarity to known genes were classified into seven major categories on the basis of putative general functions of the protein encoded and each major category was further subdivided according to the putative specific functions of the proteins.

database. To identify novel rat genes encoding proteins structurally related to the known proteins, we performed BLASTX similarity search in the peptide databases using 5512 distinct ESTs. Of these, 127 represent rat homologs of genes identified in other species or new members of structurally related families in rat, the cut off for significant similarity being P value of 10^7 and similarity of 50% (Table 4). Functional analyses of the proteins encoded by these ESTs should clarify their novel roles in pancreatic islets.

Table 4 Rat homologs of known genes and new members of gene families

Clone ID	Gene	Species	%SIM	*P* value	Islet	RINm5F
RBC00545	breast carcinoma amplified sequence 3 homolog	Homo sapiens	0·72	2E-21	1	1
RBC00858	retrovirus-related POL polyprotein	Mus musculus	0·74	4E-42	1	0
RBC01084	eukaryotic translation initiation factor 4 gamma 3	Homo sapiens	0·66	5E-44	1	0
RBC01744	protein C20orf149 homolog	Rattus norvegicus	0·74	1E-51	1	0
RBC03066	DNA transformation protein comF	Pseudomonas stutzeri	0·74	3E-63	1	0
RBC03516	hypothetical protein in acoE 3′region	Rhodobacter sphaeroides	0·85	7E-48	1	0
RBC04593	speckle-type POZ protein-like 1; POZ 56 protein	Rattus norvegicus	0·72	3E-32	1	0
RBC05170	hypothetical protein 3	Rattus norvegicus	0·78	7E-35	1	1
RBC05610	14 kDa phosphohistidine phosphatase	Rattus norvegicus	0·64	9E-28	3	0
RBC06005	transposase for insertion sequence element IS904	Pseudomonas putida	0·83	2E-49	1	0
RBC06162	aurora-A kinase interacting protein	Rattus norvegicus	0·79	2E-45	2	0
RBC06680	epsin 1	Rattus norvegicus	0·75	8E-38	1	0
RBC07890	probable Pol polyprotein	Rattus norvegicus	0·71	5E-43	1	0
RBC08283	G protein-coupled receptor 150	Mus musculus	0·81	4E-89	1	0
RBC08942	Egl nine homolog 2	Mus musculus	0·75	2E-07	2	0
RBC09487	hypothetical protein PP2447 homolog	Mus musculus	0·75	4E-19	1	1
RBC09605	putative NF-kappa-B activating protein	Rattus norvegicus	0·64	1E-43	1	0
RBC10109	polyposis locus protein 1	Rattus norvegicus	0·79	2E-34	1	0
RBC11646	immunoglobulin light chain variable region	Mus musculus	0·58	2E-27	1	0
RBC12591	chromosome 10 open reading frame 45	Homo sapiens	0·62	4E-41	2	0
RBC12830	pORF2	Mus musculus	0·64	1E-34	1	0
RBC12840	methyl-accepting chemotaxis protein	Pseudomonas syringae	0·64	9E-34	1	0
RBC13044	ribonuclease P protein subunit p29	Homo sapiens	0·64	2E-19	2	0
RBC13272	hypothetical protein KIAA0174	Rattus norvegicus	0·76	5E-13	3	0
RBC13530	histone deacetylase 4 (HD4)	Rattus norvegicus	0·73	2E-18	1	0
RBC14441	lactoylglutathione lyase	Rattus norvegicus	0·71	7E-49	2	0
RBC14577	lymphocyte antigen Ly-6D precursor	Rattus norvegicus	0·79	4E-38	2	0
RBC15029	protein C21orf5	Homo sapiens	0·85	7E-33	1	0
RBC15417	POL polyprotein	Trichosurus vulpecula	0·71	9E-25	2	0
RBC15535	amino-acid ABC transporter ATP-binding protein	Thermobifida fusca	0·83	3E-80	1	0
RBC15661	zinedin	Rattus norvegicus	0·76	1E-22	1	1
RBC15731	collagen alpha 1(X) chain	Mus musculus	0·66	3E-64	4	0
RBC17834	probable transcriptional regulatory protein ygiX	synthetic construct	0·72	2E-53	1	0
RBC18842	nectin 4	Mus musculus	0·75	8E-73	1	0
RBC19899	probable 3-hydroxybutyryl-CoA dehydrogenase	Brucella melitensis	0·81	3E-99	3	0
RBC20402	hypothetical protein CGI-143	Rattus norvegicus	0·64	4E-28	1	0
RIN00207	S-adenosylmethionine-dependent methyltransferase	Mycoplasma mobile	0·68	3E-54	0	2
RIN00228	heat inducible transcriptional repressor protein	Mycoplasma mobile	0·59	3E-32	0	1
RIN00249	cell cycle control protein cwf15	Mus musculus	0·56	4E-12	0	1
RIN00488	hypothetical protein C5D6·06c	Mus musculus	0·86	3E-85	0	1
RIN00804	zinc finger protein 35 (Zfp-35)	Rattus norvegicus	0·62	5E-22	0	1
RIN01128	tyrosine-protein kinase FLK	Rattus norvegicus	0·77	1E-42	0	1
RIN01130	trigger factor (TF)	Mycoplasma pulmonis	0·65	7E-11	0	1
RIN01171	WD-repeat protein CGI-48	Rattus norvegicus	0·6	9E-28	0	1
RIN01386	probable cation-transporting ATPase 1	Rattus norvegicus	0·79	3E-82	0	1
RIN01389	L-lactate dehydrogenase	Mesoplasma florum L1	0·58	2E-20	0	1
RIN01548	THAP domain protein 11	Rattus norvegicus	0·51	9E-22	0	1
RIN01784	hypothetical protein KIAA0233	Rattus norvegicus	0·86	5E-23	0	1
RIN01794	cell division protein ftsH homolog	Mycoplasma pulmonis	0·64	4E-32	0	1
RIN01840	hypothetical protein MG061	Mycoplasma pulmonis	0·55	1E-34	0	4
RIN01932	condensin complex subunit 2 (p105)	Mus musculus	0·88	3E-14	0	1
RIN02139	protein FAM3A precursor	Mus musculus	0·87	5E-36	0	1
RIN02177	ABC transporter ATP-binding protein	Mycoplasma gallisepticum	0·74	7E-60	0	3
RIN02445	translation initiation factor IF3	Mycoplasma fermentans	0·62	8E-12	0	2
RIN02483	splicing factor, arginine/serine-rich 2	Mus musculus	0·66	2E-59	0	2

Table 4 *Continued*

Clone ID	Gene	Species	%SIM	*P* value	Islet	RINm5F
RIN02532	ABC transporter permease protein MG188 homolog	Mycoplasma gallisepticum	0·65	7E-11	0	2
RIN02603	protein c20orf172 homolog	Mus musculus	0·89	2E-44	0	1
RIN03035	probable nicotinate-nucleotide adenylyltransferase	Mycoplasma mobile	0·65	3E-26	0	1
RIN03110	phospholipase A2 inhibitor gamma subunit B	Rattus norvegicus	0·84	2E-07	0	3
RIN03339	hypothetical lipoprotein MPN288	Mycoplasma gallisepticum	0·56	4E-10	0	2
RIN03427	putative Pol polyprotein	Rattus norvegicus	0·77	2E-19	0	2
RIN04062	hypothetical protein MG148	Ureaplasma parvum	0·69	3E-22	0	1
RIN04617	probable cation-transporting P-type ATPase	Ureaplasma parvum	0·55	1E-31	0	1
RIN05010	testis-specific protein PBS13	Rattus norvegicus	0·78	5E-88	0	1
RIN05122	hypothetical protein KIAA0036	Rattus norvegicus	0·86	6E-76	0	1
RIN05229	E3 ubiquitin-protein ligase Nedd-4	Cricetulus griseus	0·86	4E-48	0	1
RIN05576	TLM protein	Mus musculus	0·64	1E-26	2	1
RIN05605	RRS1 ribosome biogenesis regulator homolog	Mus musculus	0·67	3E-35	0	1
RIN05994	eukaryotic translation initiation factor 4E transporter	Rattus norvegicus	0·7	3E-13	0	1
RIN06079	brain protein 14	Mus musculus	0·61	8E-12	0	1
RIN06155	slingshot 3	Rattus norvegicus	0·79	6E-47	0	1
RIN06262	mitochondrial respiratory chain complexes assembly protein	Mycoplasma pulmonis	0·73	3E-37	0	1
RIN06288	SCO2 protein homolog, mitochondrial precursor	Mus musculus	0·76	4E-45	0	3
RIN06497	methionine aminopeptidase	Mycoplasma pulmonis	0·7	1E-37	0	1
RIN06650	retrovirus-related protease	Homo sapiens	0·54	2E-12	0	1
RIN07055	TGF-beta induced apoptosis protein 2	Rattus norvegicus	0·62	2E-51	0	1
RIN07707	methyl-CpG binding domain protein 6	Rattus norvegicus	0·61	1E-24	0	1
RIN07765	splicing factor 3 subunit 1	Mus musculus	0·55	2E-35	0	1
RIN07973	NADH-ubiquinone oxidoreductase chain 4L	Rattus norvegicus	0·72	7E-23	0	1
RIN08083	signal recognition particle protein	Mycoplasma mobile	0·72	1E-21	0	1
RIN08132	forkhead box protein K1	Homo sapiens	0·66	1E-26	0	1
RIN08211	meningioma-expressed antigen 6/11	Rattus norvegicus	0·71	2E-27	2	1
RIN08412	hypothetical 45·0 kDa protein in NOT1-MATAL2 region	Rattus norvegicus	0·85	3E-79	0	1
RIN09109	TED protein	Mus musculus	0·87	5E-85	0	3
RIN09205	SRR1-like protein	Homo sapiens	0·83	2E-42	0	4
RIN09341	30S ribosomal protein S6	Mycoplasma pulmonis	0·55	2E-17	0	7
RIN09690	valyl-tRNA synthetase	Mycoplasma pulmonis	0·87	5E-40	0	5
RIN10182	protein transport protein SEC61 gamma subunit	Rattus norvegicus	0·7	5E-20	0	4
RIN10530	hypothetical protein KIAA0893	Mus musculus	0·81	9E-93	1	3
RIN10621	splicing factor, arginine/serine-rich 8	Mus musculus	0·51	5E-15	0	1
RIN10814	G protein-coupled receptor family C group 5 member C	Rattus norvegicus	0·73	8E-43	0	2
RIN11008	tRNA-splicing endonuclease subunit SEN54	Rattus norvegicus	0·86	6E-98	0	3
RIN11027	protein yhgF	Homo sapiens	0·89	9E-93	1	1
RIN11030	p53-associated parkin-like cytoplasmic protein	Rattus norvegicus	0·84	9E-44	0	1
RIN11408	protein disulfide isomerase precursor	Rattus norvegicus	0·74	7E-35	0	1
RIN11741	polyhomeotic-like protein 1	Rattus norvegicus	0·82	3E-54	0	2
RIN11820	translation initiation factor IF-1	Leptospira interrogans	0·74	9E-12	0	1
RIN11856	fructose-bisphosphate aldolase	Mycoplasma mobile	0·8	6E-52	0	1
RIN11899	nuclear protein Hcc-1	Mus musculus	0·72	2E-67	0	2
RIN12027	60S ribosomal protein L22	Mus musculus	0·81	2E-39	0	1
RIN12420	putative ATP-dependent RNA helicase T26G10	Rattus norvegicus	0·8	4E-83	0	2
RIN12461	synaptic vesicle membrane protein VAT-1 homolog	Rattus norvegicus	0·73	8E-65	0	3
RIN12488	glutaminyl-peptide cyclotransferase	Rattus norvegicus	0·78	7E-90	0	2
RIN12608	probable RNA-dependent helicase p72	Rattus norvegicus	0·64	8E-62	0	2
RIN12715	oligoendopeptidase F homolog	Mycoplasma pulmonis	0·68	8E-58	0	1
RIN12733	PSL10 protein	Mus musculus	0·77	7E-80	0	1
RIN12794	hypothetical protein ORF-1137	Rattus norvegicus	0·72	2E-32	0	1
RIN12806	protein HI1455	Mycoplasma pulmonis	0·74	7E-50	0	3

Table 4 *Continued*

Clone ID	Gene	Species	%SIM	*P* value	Islet	RINm5F
RIN12897	dimethyladenosine transferase	Mycoplasma gallisepticum	0·67	5E-36	0	2
RIN12933	zinc finger protein 23	Mus musculus	0·79	4E-63	1	2
RIN13293	serine hydroxymethyltransferase	Rattus norvegicus	0·65	4E-93	0	1
RIN13306	peripherin	Mus musculus	0·7	1E-42	0	1
RIN13714	seryl-tRNA synthetase	Mycoplasma pulmonis	0·82	3E-62	0	1
RIN13766	UBA/UBX 33·3 kDa protein	Rattus norvegicus	0·65	1E-54	0	1
RIN13865	splicing factor 1	Mus musculus	0·73	6E-23	0	2
RIN14137	condensin subunit 1	Mus musculus	0·62	2E-33	0	1
RIN14394	Csr1	Cricetulus griseus	0·82	8E-12	0	1
RIN14494	hypothetical protein KIAA0117	Rattus norvegicus	0·77	7E-81	0	1
RIN14758	general transcription factor 3C polypeptide 5	Rattus norvegicus	0·73	1E-48	1	1
RIN15622	Crumbs protein homolog 1	Mus musculus	0·6	7E-08	0	1
RIN16005	hypothetical protein C19A8·09	Rattus norvegicus	0·89	4E-11	0	2
RIN16184	single-strand binding protein 1	Rattus norvegicus	0·67	6E-13	1	1
RIN16330	zinc finger protein 510	Mus musculus	0·83	2E-19	0	2
RIN16398	60S ribosomal protein L7	Rattus norvegicus	0·83	1E-32	1	3
RIN16446	single-strand binding protein	Rattus norvegicus	0·87	3E-13	0	1
RIN16588	putative adenosylhomocysteinase 2	Homo sapiens	0·89	2E-07	0	2
RIN16592	thymidine phosphorylase	Mycoplasma pulmonis	0·66	3E-20	0	1

The unknown clones after BLASTN search in the nucleotide databases were re-sequenced from the 5′-end and then subjected to BLASTX search in the peptide databases. The cut off used for significant similarity was *P* value of 10^{-7} and similarity of 50%. The clone IDs RBC and RIN show ESTs from pancreatic islet and RINm5F cells, respectively.

To identify genes that had not been determined to be expressed, the sequences showing no significant match to any of the annotated genes were further compared with dbEST entries from other tissues, revealing 896 expressed genes that have not appeared in the database. These 896 putatively novel genes were analyzed in the context of recently determined rat genomic sequences (Rat Genome Sequencing Project Consortium 2004). Of these genes, 779 were encoded by exon-intron organization in the corresponding genomic sequences. However, since the transcripts of many of these genes were barely detectable by *in situ* hybridization, they may be expressed at low levels, at least in adult islets. Because chemiluminescence-based *in situ* detection of mRNAs is not sufficiently sensitive, a large-scale RT-PCR analysis is presently in progress in our laboratory to elucidate the tissue distribution. The expression of 71 of the other 117 sequences was uncertain due to their ambiguous genomic structure. The remainder could be derived from possible pseudogenes or retroposons, as the corresponding sequences in the genome are flanked by AT-rich sequences that were recognizable by oligo-dT priming in the process of cDNA synthesis.

Characterization of differentially expressed genes

The immunoreactive insulin (IRI) content of rat normal β-cells has been reported to be ~8000 pmol/10^6 cells, while the RINm5F cell line used in this study has been estimated to contain a very low level of IRI (0·43 pmol/10^6 cells) and a much lower number of secretory granules (Kayo *et al.* 1997). Accordingly, the expression levels of the genes involved in insulin synthesis and secretion in RINm5F cells should markedly differ from those of the normal β-cells. In addition, since the RINm5F cells were derived from radiation-induced tumor cells and exhibited a decrement of well-differentiation (Gazdar *et al.* 1980, Philippe *et al.* 1987, Kayo *et al.* 1997), the expression levels of the genes involved in cell differentiation and tumorigenesis may also be altered. The relative frequencies of ESTs have been shown to reflect the average level of expression of the corresponding mRNAs in the tissues examined (Lee *et al.* 1995). As pancreatic islet cells are mostly β-cells, the expression profile of insulin-related genes in the two cDNA sources (of same size) can be compared to identify differentially expressed genes. The EST frequencies for most of the house-keeping genes were similar in the two cDNA libraries, suggesting that such comparison of EST frequencies is reasonable. Over 2-fold differences in frequency between the two libraries were found in 204 genes (higher EST>10 times). The direction of change in mRNA levels in these ESTs observed by comparison of the EST frequency and the TaqMan semi-quantitative analysis was quite parallel, while the magnitude of the change was not correlated. The representative results for some of the ESTs (>15 times) are shown in Fig. 1. Previously, similar comparative analysis using ~6000 ESTs from two different conditions of PC-12 cells was performed (Lee *et al.* 1995). The study found the ratio of EST frequencies between the two cDNA sources to be

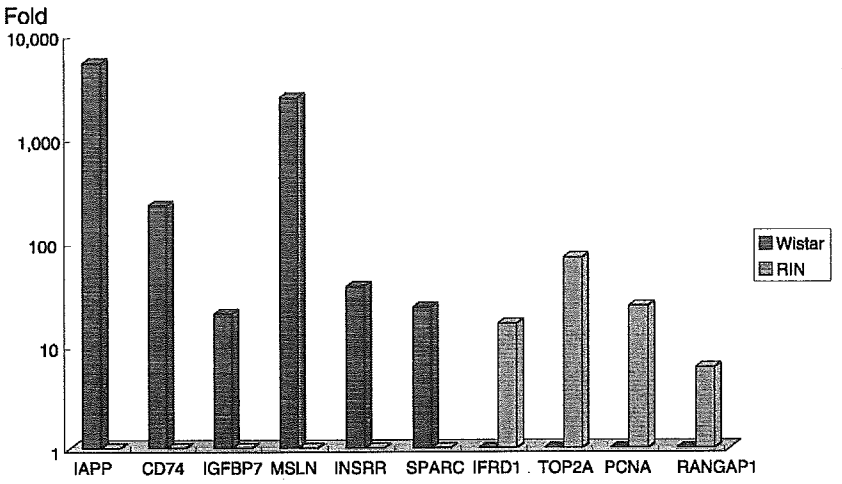

Figure 1 Expression profiles of the representative genes abundant either in islet (from Wistar rats) or in RINm5F cells. The expression levels of the genes were estimated by real-time quantitative RT-PCR using the TaqMan system. The mRNA for major α-globin could not be efficiently detected by the TaqMan system in its standard range for unknown reasons.

correlated with the Northern blot analysis, except for the low-frequency ESTs. Thus, the genes of interest that are expressed at least at moderate levels also should be examined by semi-quantitative analysis before further analysis.

We focus on the seven genes exhibiting high expression (>15 times) in islet but no expression in RINm5F cells (Table 5), since they may well play a specific role in insulin synthesis and secretion. Expression of all of the genes examined was confirmed in pancreatic β cells by *in situ* hybridization, although the expression was not restricted to β cells (Fig. 2A). The expression of the high EST clones (>15 times) in RINm5F cells and none in islets was found only outside the islets by *in situ* hybridization (Fig. 2B). These patterns of expression are consistent with a previous assumption that over 90% of the genes are house-keeping and are expressed at various levels in many tissues. Receptors of the insulin/insulin-like growth factor (IGF) family have been implicated both in the regulation of pancreatic β-cell growth and insulin secretion. IRR, an orphan receptor of the insulin receptor subfamily, is expressed at a considerably higher level in pancreatic β cells (Hirayama *et al.* 1999), and a decrease in the mRNA level was found in diabetic GK rats (N Shihara, Y Horikawa, J Takeda, unpublished observations). However, since glucose-stimulated insulin secretion and embryonic β cell development have been shown to occur normally in mice lacking IRR (Kitamura *et al.* 2001), decreased expression of IRR alone may be insufficient for the development of diabetes. To understand the functional properties of IRR in pancreatic β cells, it is important to identify its possible ligand and functional partners. Mesothelin, produced by mesothelial cells, has been suggested to play a role in cellular adhesion in ovarian cancer cells (Scholler *et al.* 1999). Since this EST was not found in cultured RINm5F cells, methothelin may well be unimportant in single-cell growth without cell adhesion and not directly involved in endocrine

Table 5 Genes abundantly identified either in islet or in RINm5F cells

Gene	Islet	RINm5F
Islet amyloid polypeptide (amylin)	321	0
CD74	28	0
Follistatin-like protein (mac25)	24	0
Mesothelin	23	0
Major alpha-globin	23	0
Insulin receptor-related receptor (IRR)	18	0
Osteonectin	17	0
Interferon-related developmental regulator 1	0	53
DNA topoisomerase II alpha	0	43
Proliferating cell nuclear antigen	0	26
Glycosyl-phosphatidyl-inositol-anchored protein homolog	0	24
Casein kinase 1 gamma 2 isoform	0	22
Ran-GTPase activating protein 1	0	21
Heat shock 70kD protein 5	0	20
High mobility group protein 17	0	17
Phosphoglycerate mutase type B subunit	0	17
Synaptic regulatory protein RIM2beta	0	17
Heat shock protein 60	0	17
HLA-B associated transcript 2	0	16
Secretory carrier membrane protein 3	0	15

Genes with high (>15 times) expression in either islet and RINm5F and none of the other, are listed.

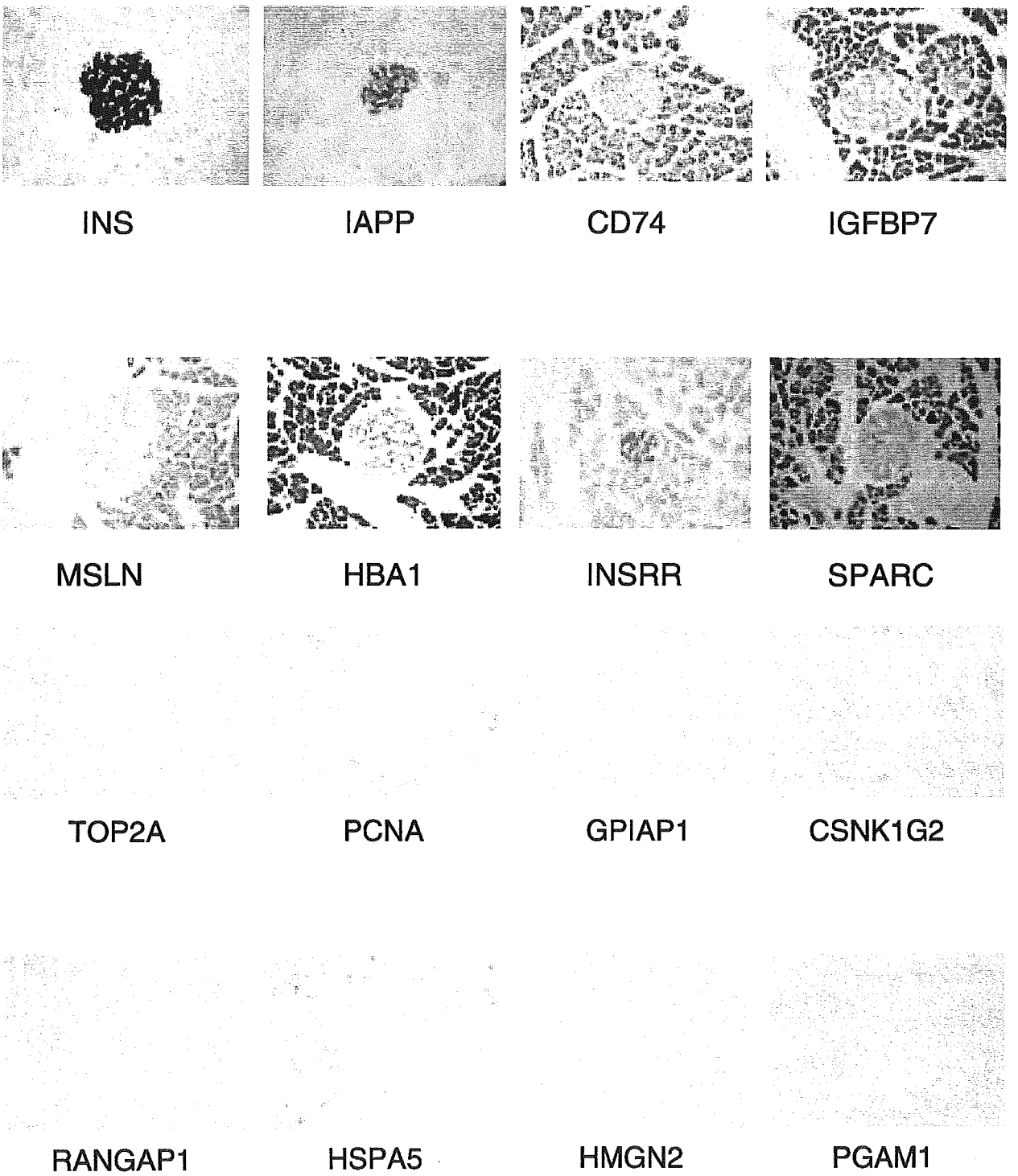

Figure 2 (A) *In situ* hybridization of genes with high expression in islet and none in RINm5F cells (A). The representative results of seven highly expressed genes (> 15 times) and the insulin gene are shown. (B) *In situ* hybridization of genes with high expression in RINm5F cells and none in islet.

tumorigenesis. CD74, which is transitorily associated with class II histocompatibility antigens during intracellular transport (Claesson *et al.* 1983), also was found to be highly expressed in human islet tumor in our previous study (Jin *et al.* 2003). Since this islet tumor exhibited features of moderately differentiated islet cells, despite markedly reduced insulin secretion, in contrast to RINm5F cells, CD74 might be related to the degree of cell differentiation. Osteonectin is reported to be abundantly expressed in adipose tissue. Although the molecule has not been reported to be expressed in islets, *in situ* hybridization showed that its expression is moderate in islet cells and diffuse throughout the pancreas of normal Wistar rat. It has been reported that absence of osteonectin leads to an increase in the size of individual adipocytes as well as in the number of adipocytes per fat pad (Bradshaw *et al.* 2003). Thus, osteonectin might be related to β cell mass and growth rather than insulin synthesis and secretion. As these genes might contribute to the development of islet cell specificity, their functional significance should be examined in various conditions of pancreatic islets.

Applications of rat ESTs for islet studies

In this study, we describe a collection of 40 710 rat pancreatic islet-related ESTs representing 10 406 different transcripts. This is the first report describing a systematic collection of rat expressed genes from pancreatic islets and a β-cell line. Since DNA microarray technology relies largely on the rapid growth of the EST databases, these newly identified expressed genes should facilitate analysis of differential gene expression in pancreatic islets under various conditions. At present, only the PanChip microarray, which was prepared using 3400 cDNA sequences from mouse whole pancreas, is available as a tissue-specific microarray for islet studies (Scearce *et al.* 2002). Accordingly, the establishment of islet-specific DNA microarrays for the rat should be especially important in the analysis of the transcriptome of diabetic rats such as GK and OLETF, and is presently underway in our laboratory. Another advantage of the large-scale collection of EST clones is that the cDNA fragments obtained can be used as hybridizing probes for Northern blotting or *in situ* hybridization to analyze the size and number of alternatively spliced transcripts and their local tissue distribution. However, it is possible that a small fraction of the ESTs obtained might be contaminated from other cell types such as endothelial cells or blood. Indeed, our preliminary trial of non-isotopic *in situ* hybridization using rat ESTs was found to be quite effective for analysis of mRNA expression in pancreatic islets. A large-scale *in situ* hybridization of rat islet mRNAs is also presently in progress in our laboratory. Functional analysis of a wide spectrum of islet-specific genes and genes highly abundant or less abundant in islets identified by this approach might clarify the molecular mechanisms underlying the differentiation of islet cells, tumorigenesis, and the pathogenesis of diabetes, as well as lead to new therapies for the improvement and regeneration of β-cell function through manipulation of gene expression and gene products.

In addition, as the genome sequence analysis of the Brown Norway rat recently has been completed (Rat Genome Sequencing Project Consortium 2004), the results of this study should be helpful in annotating the genes actually expressed in the rat genome and thus provide further insight into mammalian evolution of genes involved in tissue-specificity of endocrine pancreas.

Acknowledgements

We thank S Oike, R Kawakami, Y Yaginuma, I Uda, Y Ibe, and T Takahashi for excellent assistance. This study was supported by Grant-in-Aid for Scientific Research and for Scientific Research on Priority Areas (C) "Medical Genome Science" from the Japanese Ministry of Science, Education, Sports, Culture and Technology and by a Health and Labor Science Research Grant for Research on Human Genome and Tissue Engineering from the Japanese Ministry of Health, Labor and Welfare, the Naito Foundation and the Yamanouchi Foundation. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References

Alberts B, Bray D, Lewis J, Raff M, Roberts K & Watson JD 1994 The cell nucleus. In: *Molecular Biology of the Cell.* pp 335–400. New York: Garland.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25** 3389–3402.

Bernal-Mizrachi E, Cras-Meneur C, Ohsugi M & Permutt MA 2003 Gene expression profiling in islet biology and diabetes research. *Diabetes Metabolism Research Review* **19** 32–42.

Bonner-Weir S & Sharma A 2002 Pancreatic stem cells. *Journal of Pathology* **197** 519–526.

Bradshaw AD, Graves DC, Motamed K & Sage EH 2003 SPARC-null mice exhibit increased adiposity without significant differences in overall body weight. *PNAS* **100** 6045–6050.

Claesson L, Larhammar D, Rask L & Peterson PA 1983 cDNA clone for the human invariant gamma chain of class II histocompatibility antigens and its implications for the protein structure. *PNAS* **80** 7395–7399.

Dor Y, Brown J, Martinez OI & Melton DA 2004 Adult pancreatic β-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429** 41–46.

Edlund H 2002 Pancreatic organogenesis – Developmental mechanisms and implications for therapy. *Nature Genetics Review* **3** 524–532.

Fajans SS, Bell GI & Polonsky KS 2001 Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *New England Journal of Medicine* **345** 971–980.

Gazdar AF, Chick WL, Oie HK, Sims HL, King DL, Weir GC & Lauris V 1980 Continuous, clonal, insulin- and somatostatin-secreting cell lines established from a transplantable rat islet cell tumor. *PNAS* **77** 3519–3523.

Hirayama I, Tamemoto H, Yokota H, Kubo SK, Wang J, Kuwano H, Nagamachi Y, Takeuchi T & Izumi T 1999 Insulin receptor-related receptor is expressed in pancreatic beta-cells and stimulates tyrosine phosphorylation of insulin receptor substrate-1 and -2. *Diabetes* **48** 1237–1244.

Hogenesch JB, Ching KA, Batalov S, Su AI, Walker JR, Zhou Y, Kay SA, Schultz PG & Cooke MP 2001 A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* **106** 413–415.

International Human Genome Sequencing Consortium 2001 Initial sequencing and analysis of the human genome. *Nature* **409** 860–921.

Jin L, Wang H, Narita T, Kikuno R, Ohara O, Shihara N, Nishigori T, Horikawa Y & Takeda J 2003 Expression profile of mRNAs from human pancreatic islet tumors. *Journal of Molecular Endocrinology* **31** 519–528.

Kayo T, Sawada Y, Suda M, Konda Y, Izumi T, Tanaka S, Shibata H & Takeuchi T 1997 Proprotein-processing endoprotease furin controls growth of pancreatic β-cells. *Diabetes* **46** 1296–1304.

Kitamura T, Kido Y, Nef S, Merenmies J, Parada LF & Accili D 2001 Preserved pancreatic β-cell development and function in mice lacking the insulin receptor-related receptor. *Molecular and Cellular Biology* **21** 5624–5630.

Lee NH, Weinstock KG, kirkness EF, Earle-Hughes JA, Fuldner RA, Marmaros S, Glodek A, Gocayne JD, Adams MD, Kerlavage AR, Fraser CM & Venter JC 1995 Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *PNAS* **92** 8303–8307.

Ma H-T, Kato M & Tatemoto K 1996 Effects of pancreastatin and somatostatin on secretagogues-induced rise in intracellular free calcium in single rat pancreatic islet cells. *Regulatory Peptide* **61** 143–148.

Mouse Genome Sequencing Consortium 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* **420** 520–562.

Philippe J, Chick WL & Habener JF 1987 Multipotential phenotypic expression of genes encoding peptide hormones in rat insulinoma cell lines. *Journal of Clinical Investigation* **79** 351–358.

Rat Genome Sequencing Project Consortium 2004 Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428** 493–521.

Scearce LM, Brestelli JE, McWeeney SK, Lee CS, Mazzarelli J, Pinney DF, Pizarro A, Stoeckert CJ Jr, Clifton SW, Permutt MA, Brown J, Melton DA & Kaestner KH 2002 Functional genomics of the endocrine pancreas. The pancreas clone set and PanChip, new resources for diabetes research. *Diabetes* **51** 1997–2004

Scholler N, Fu N, Yang Y, Ye Z, Goodman GE, Hellstrom KE & Hellstrom I 1999 Soluble member(s) of the mesothelin/megakaryocyte potentiating factor family are detectable in sera from patients with ovarian carcinoma. *PNAS* **96** 11531–11536.

Takeda J, Yano H, Eng S & Bell GI 1993 A molecular inventory of human pancreatic islets: sequence analysis of 1000 cDNA clones. *Human Molecular Genetics* **2** 1793–1798.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.* 2001 The sequence of the human genome. *Science* **291** 1304–1351.

Received 17 March 2005
Accepted 29 April 2005
Made available online as an Accepted Preprint 16 May 2005

0021-972X/05/$15.00/0
Printed in U.S.A.

The Journal of Clinical Endocrinology & Metabolism 90(10):5841–5847

doi: 10.1210/jc.2005-0991

Genetic Variation in the Hypoxia-Inducible Factor-1α Gene Is Associated with Type 2 Diabetes in Japanese

Norihiro Yamada, Yukio Horikawa, Naohisa Oda, Katsumi Iizuka, Nobuyuki Shihara, Shoji Kishi, and Jun Takeda

Department of Ophthalmology (N.Y., S.K.), Gunma University Graduate School of Medicine, 371-8511 Gunma, Japan; Laboratory of Medical Genomics, Biosignal Genome Resource Center (Y.H., N.S., J.T.), Institute for Molecular and Cellular Regulation, Gunma University, 371-8512 Gunma, Japan; Core Research for Evolutional Science and Technology (Y.H., K.I., N.S., J.T.), Japan Science and Technology Corporation, 332-0012 Kawaguchi, Japan; Department of Diabetes and Endocrinology (Y.H., J.T.), Gifu University School of Medicine, Gifu 501-1194, Japan; and Department of Internal Medicine (N.O.), Fujita Health University School of Medicine, 470-1192 Aichi, Japan

Context and Objective: Vascular endothelial growth factor plays a critical role both in neovascularization of proliferative diabetic retinopathy and in angiogenesis of islets in the pancreatic developmental stage in determining β-cell mass and properties. Vascular endothelial growth factor mRNA levels increase as a result of increased transcriptional activation, mediated predominantly by hypoxia-inducible factor-1 α (HIF-1α) in response to hypoxia.

Design and Patients: In this study, we examined all regions of the HIF-1α to detect single-nucleotide polymorphisms (SNPs), evaluated the pattern of linkage disequilibrium to analyze haplotypes, and performed association studies in Japanese type 2 diabetes patients with or without retinopathy.

Results: A total of 35 SNPs were found in the gene, 27 of which were reported previously and eight of which were novel. Three of the 35 SNPs were located in coding regions, one in exon 2 (S28Y), and the others in exon 12 (P582S, A588T). The P582S HIF-1α mutation was associated with type 2 diabetes ($P = 0.0028$) by a consistently higher level of transcriptional activity than wild type, especially under hypoxic condition ($P = 0.012$), but no association with retinopathy was detected.

Conclusion: This is the first report that HIF-1α is associated with the occurrence of type 2 diabetes and suggests that the P582S HIF-1α mutation should be assessed in larger studies as a risk factor for type 2 diabetes. (***J Clin Endocrinol Metab* 90: 5841–5847, 2005**)

DIABETES LEADS TO specific microvascular complications of retinopathy, nephropathy, and neuropathy as well as increased risk of atherosclerosis, which may reflect underlying endothelial dysfunction. The risk of developing these complications is increased by poor glycemic control, but the relevance of genetic background is clearly established (1–3).

Diabetic retinopathy is a major cause of new-onset blindness among diabetic adults and is characterized by increased vascular permeability, tissue ischemia, and neovascularization. Neovascularization of the retina carries a high risk of blindness as a result of vitreous hemorrhage, fibrosis, and tractional retinal detachment. Vascular endothelial growth factor (VEGF) can stimulate angiogenesis, enhance collateral vessel formation, and increase permeability of the microvasculature (4). In diabetic proliferative retinopathy, VEGF plays a critical role in neovascularization and breakdown of the blood-retinal barrier characterized by hyperpermeability of retinal vessels. VEGF levels have been found to be markedly elevated in vitreous and aqueous fluids in the eyes of patients with proliferative diabetic retinopathy (PDR) (5). Furthermore, VEGF has been reported as a susceptibility gene for type 2 diabetes mellitus (T2DM) as well as diabetic retinopathy (6). VEGF is known to be a key factor in angiogenesis of islets in the pancreatic developmental stage in determining β-cell mass and properties (7). In response to hypoxia, VEGF mRNA levels are increased by increased transcriptional activation, whereas overall protein synthesis is inhibited. This increase is mediated predominantly by hypoxia-inducible factor-1 (HIF-1) binding to a hypoxia response element located 1 kb upstream of the transcriptional start site of VEGF (8, 9).

HIF-1, a transcription factor found in mammalian cells cultured under reduced oxygen tension, plays an essential role in cellular and systemic homeostatic responses to hypoxia. HIF-1 acts as a heterodimer composed of a 120-kDa HIF-1α subunit complexed with a 91- to 94-kDa HIF-1β subunit (10). Although HIF-1β is ubiquitously expressed and maintained at constant cellular levels, the HIF-1α protein level and transcriptional activity are tightly regulated in response to oxygen levels. Thus, HIF-1 activity is controlled by the oxygen-regulated expression of the HIF-1α subunit (11). Under nonhypoxic conditions, HIF-1α is hydroxylated on proline residues by a family of oxygen-dependent prolyl hydroxylases. The hydroxylated prolines independently mediate high-affinity binding to the von Hippel-Lindau (VHL) protein, a component of the E3 ubiquitin-protein ligase com-

First Published Online July 26, 2005

Abbreviations: BMI, Body mass index; CI, confidence interval; cSNP, coding SNP; HbA_{1c}, glycosylated hemoglobin; HDL, high-density lipoprotein; HIF-1α, hypoxia-inducible factor-1α; HT, hypertension; LD, linkage disequilibrium; OR, odds ratio; PDR, proliferative diabetic retinopathy; SNP, single-nucleotide polymorphism; T-chol, total cholesterol; T2DM, type 2 diabetes mellitus; TG, triglyceride; TK, thymidine kinase; VEGF, vascular endothelial growth factor; VHL, von Hippel-Lindau.

JCEM is published monthly by The Endocrine Society (http://www.endo-society.org), the foremost professional society serving the endocrine community.

plex that ubiquitinates HIF-1α, thereby targeting it for degradation. The critical proline residues for VHL binding when hydroxylated are P402 and P564, both of which are located in the oxygen-dependent degradation domain (12, 13)

The HIF-1α gene is located at chromosome 14q21-q24, where the susceptibility locus to T2DM was localized in Finns (14). The predicted 826-amino-acid HIF-1α contains a basic helix-loop-helix-Per-Arnt-Sim (PAS) domain at its N terminus and a transactivation domain and transcriptional inhibitory domain at its C terminus (15). HIF-1 target genes include those for energy metabolism, iron homeostasis, angiogenesis, and cell proliferation and viability (16). Thus, HIF-1α activates many genes in the glycolysis system under conditions of hypoxia differently from insulin (17). In addition, the involvement of HIF-1 in the pathophysiology of human disease including myocardial ischemia, cerebral ischemia, retinal ischemia, pulmonary hypertension, preeclampsia, intrauterine growth retardation, and cancer has been reported (16). However, the correlations between HIF-1 and T2DM including complications remain to be elucidated.

In this study, we examined all regions of the HIF-1α gene in Japanese subjects to detect single-nucleotide polymorphisms (SNPs), evaluated the pattern of linkage disequilibrium (LD) to compose haplotypes in the gene, and performed association studies in T2DM patients. We identified a susceptibility coding SNP (cSNP) (P582S) and haplotype in the HIF-1α gene for T2DM. This is the first report suggesting that HIF-1α is associated with the occurrence of T2DM.

Subjects and Methods

Subjects

A total of 440 patients with T2DM [245 males and 195 females; age at testing, 60.5 ± 11.4 yr; duration, 11.4 ± 9.1 yr; body mass index (BMI), 23.9 ± 4.3 kg/m^2; maximum BMI, 27.5 ± 4.6 kg/m^2; glycosylated hemoglobin (HbA_{1c}), 7.9 ± 3.8%; total cholesterol (T-chol), 0.52 ± 0.11 mmol/liter; high-density lipoprotein (HDL), 0.13 ± 0.04 mmol/liter; triglyceride (TG), 0.39 ± 0.25 mmol/liter) and 572 controls (231 males and 342 females; age at testing, 67.3 ± 6.5 yr; BMI, 23.0 ± 2.9 kg/m^2; HbA_{1c}, 5.0 ± 0.4%) were examined for an association study. The diagnosis of T2DM was based on medical records or 75-g oral glucose tolerance test according to the criteria of the Japan Diabetes Society (18). One hundred eighty-two T2DM patients with retinopathy [80 males and 102 females; PDR, n = 117; pre-PDR, n = 29; and simple diabetic retinopathy, n = 36; age at testing, 61.9 ± 11.2 yr; duration, 13.0 ± 9.7 yr; BMI, 23.7 ± 3.9 kg/m^2; maximum BMI, 27.9 ± 6.6 kg/m^2; HbA_{1c}, 7.6 ± 1.9%; T-chol, 0.52 ± 0.14 mmol/liter; HDL, 0.12 ± 0.04 mmol/liter; TG, 0.41 ± 0.20 mmol/liter; hypertension (HT) drug use, 94 (−)/88(+); insulin therapy, 105(−)/77(+)] and 125 without retinopathy (47 males and 78 females; age at testing, 62.9 ± 11.6 yr; duration, 11.0 ± 8.6 yr; BMI, 23.7 ± 4.8 kg/m^2; maximum BMI, 26.1 ± 6.6 kg/m^2; HbA_{1c}, 6.8 ± 1.4%; T-chol, 0.50 ± 0.10 mmol/liter; HDL, 0.10 ± 0.04 mmol/liter; TG, 0.36 ± 0.22 mmol/liter; HT drug use, 99(−)/26(+); insulin therapy, 89(−)/36(+)]were examined by an association study with special reference to retinopathy. An additional 88 patients diagnosed with PDR by an ophthalmologist were included in the patients with retinopathy. Control subjects were recruited on the following criteria: 60 or more years of age, no past history of diagnosis of diabetes mellitus, HbA_{1c} less than 5.6%, and no familial history of diabetes mellitus in second-degree relatives. The study was approved by the Ethics Committee of Gunma University and Fujita Health University School of Medicine upon written, informed consent of each subject.

SNP identification in the HIF-1α gene

Genomic DNA was extracted from samples of whole blood using QIAamp DNA blood kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Sixteen of the random control samples (32 alleles) were used to detect SNPs in the HIF-1α gene. Primers for PCR experiments were designed by Primer 3 (available from http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) on the basis of the genomic contig sequence (GenBank accession number NT_026437) of the HIF-1α region. The mixture for PCR was 20 μl in 10 ng template DNA, 0.5 nM of each dNTP, 2.5 pmol of each forward and reverse primer, 0.5 U Ex*Taq* polymerase (Takara, Kyoto, Japan), and 2 μl of 10× PCR buffer. The reaction conditions were an initial denaturation step of 95 C for 3 min and a subsequent 40 cycles of reaction at 94 C for 30 sec, 52–63 C for 30 sec, and 72 C for 1 min, and a final extension step of 72 C for 10 min. A 3-μl aliquot from each reaction was assayed on a 1% agarose gel to confirm the product, and the remainder was purified using MultiScreen Filtration System (Millipore, Billerica, MA) with Sephadex G-75 (Amersham Biosciences, Piscataway, NJ). Each PCR product was subjected to cycle sequencing with BigDye Terminator cycle sequencing FS (Applied Biosystems, Foster City, CA) using each forward and reverse primer. Reaction products were purified by ethanol precipitation and sequenced by ABI PRISM 3100 or 3700 sequencer. Results were processed with Autoassembler version 2.1 (Applied Biosystems) to compare sequences.

Mutation screening and genotyping of frequent polymorphisms in the HIF-1α gene

We examined all of the coding regions of the HIF-1α gene in 96 of the 440 T2DM patients (59 males and 37 females; age, 58.6 ± 12.1 yr; age at diagnosis, 44.2 ± 12.7 yr; duration, 14.4 ± 9.2 yr; BMI, 23.7 ± 3.8 kg/m^2; maximum BMI, 28.9 ± 4.5 kg/m^2; HbA_{1c}, 7.3 ± 1.5%) and 96 of the 576 control subjects (35 males and 61 females; age, 67.6 ± 5.8 yr; BMI, 22.9 ± 2.8 kg/m^2; HbA_{1c}, 4.9 ± 0.3%). Twenty-four frequent SNPs were examined in the 440 T2DM patients and 572 controls by direct sequencing or TaqMan assay (Applied Biosystems). PCR was performed in a total volume of 5 μl, which contained 2.5 ng DNA, 1× TaqMan Universal PCR Master Mix, with each primer at a concentration of 900 nM and each probe at a concentration of 200 nM. Thermal cycling conditions were as follows: 50 C for 2 min and 95 C for 10 min to activate the amperase uracil-*N*-glycosylase and AmpliTaq Gold enzyme, respectively, followed by 40 cycles of 92 C for 15 sec and 52–63 C for 1 min. The fluorescence level was measured with an ABI PRISM 7900HT sequence detector (Applied Biosystems), resulting in clear identification of three genotypes.

Estimation of haplotype frequencies and evaluation of pattern of LD in the HIF-1α gene

Haplotypes and haplogenotypes were inferred by the expectation-maximization method by Haploview (http://www.broad.mit.edu/personal/jcbarret/haploview) and PHASE 2.1.1 (http://www.stat.washington.edu/stephens), respectively. The coefficients for LD, D', and r^2 value were estimated by GOLD software (available from http://www.well.ox.ac.uk/asthma/GOLD).

Cloning of human HIF-1α and variants

A cDNA identical to HIF-1α was retrieved from a human islet cDNA library and subcloned in pENTR/D-TOPO (Invitrogen, Carlsbad, CA) after amplification with Pfu (Stratagene, La Jolla, CA) and the primer set 5′-CACCATGGAGGGCGCCGGCGGCGCGAAC-3′ and 5′-TCAGTTAACTTGATCCAAAGCTCTG-3′ and transferred for expression to pcDNA3.2-DEST (Invitrogen). The P582S mutation was introduced by QuikChange site-directed mutagenesis kit (Stratagene) with pENTR/D-TOPO wild-type HIF-1α as template and confirmed by sequencing.

Functional analysis of the identified HIF-1α mutation

The reporter construct VEGF promoter-pGL3 was prepared by cloning the human VEGF gene promoter (−1180 to −+338) into the firefly luciferase reporter vector pGL3-Basic (Promega, Madison, WI) (17). HEK293 cells were maintained in DMEM supplemented with 10% fetal calf serum and transfected using ExGen 500 (Fermentas, St. Leon-Rot, Germany) with various amounts of the reporter and test plasmid constructs. Transcriptional activity was normalized with a cotransfected control thymidine kinase (TK)-regulated Renilla luciferase vector,

pRL-TK (Promega). The transactivation activity of wild-type and mutant proteins was measured using the Dual-Luciferase Reporter Assay System (Promega). After 6 h transfection, the cells were incubated 24 h under normoxic (21% O_2) or hypoxic (1% O_2) conditions using Anaero Pack (Mitsubishi Gas Chemical, Tokyo, Japan) before analysis of reporter gene activity.

Statistical analyses

Statistical difference in allele frequencies between T2DM and control groups or between T2DM with and without retinopathy groups was assessed by χ^2 test, and other categorical clinical variables were compared using t test or logistic regression analysis adjusted for relevant covariates. Statistical analysis was performed with StatView 5.0 software (SAS Institute, Inc., Cary, NC).

Results

Identification of polymorphisms in the HIF-1α gene

Sixteen individuals were examined for sequence variations in 38 kb of the 78-kb region of the HIF-1α gene including all 15 exons (NT_026437, nucleotides 61222006–61300408). A total of 35 SNPs were found in the gene; the locations of these SNPs are shown in Fig. 1 in relation to the genomic structure of the HIF-1α gene. All SNPs were in Hardy-Weinberg equilibrium, and 27 were reported in the Institute of Medical Science-Japan Science and Technology Agency SNP database (http://snp.ims.u-tokyo.ac.jp/index.html) or in the National Center for Biotechnology Information SNP database (http://www.ncbi.nlm.nih.gov/SNP/). Eight SNPs (SNP-6, SNP-12, SNP-18, SNP-19, SNP-31, SNP-32, SNP-33, and SNP-34) were novel and had not been reported. Three of the 35 SNPs were located in coding regions, one in exon 2 (S28Y), and the other SNPs in exon 12 (P582S, A588T) (Table 1).

Evaluation of pattern of LD in the HIF-1α gene

Twenty-three SNPs were used to define haplotypes and evaluate the pattern of LD in the 440 T2DM patients and 572 control subjects. As shown in Fig. 2, one LD block appears in this region in both groups. The five SNPs at position SNP-22 (g.25200), SNP-23 (g.25299), SNP-24 (g.27009), SNP-26 (g.34942), SNP-27 (g.35074), and the three SNPs at position SNP-14 (g.45483), SNP-16 (g.46572), and SNP-17 (g.46820) are in complete LD.

Mutation screening and association study of genetic variation of the HIF-1α gene in T2DM patients

All exons were examined in 96 T2DM patients and 96 control subjects. We found a total of three cSNPs (S28Y, P582S, and A588T), of which S28Y is novel. We then performed an association study using possible pairwise haplotypes in T2DM patients and controls. Ten SNPs (SNP-30, -2, -3, -4, -25, -7, -18, -20, -13, and -28) were used to define haplotypes. The other SNPs were excluded because of the rarity of minor alleles. We found that a haplotype comprising SNP-25 and SNP-13, which are in strong LD, represents significant susceptibility to T2DM at a P value of 10^{-11} based on a two by four χ^2 test, and also after multiple adjustment. The 1-1/2-2 haplogenotype comprising these two SNPs was associated with significantly decreased risk of T2DM [T2DM, 25.5%; control, 32.0%; odds ratio (OR) = 0.73; 95% confidence interval (CI), 0.55–0.96; 1 − β = 53%). The 2-2/2-2 haplogenotype also was associated with decreased risk (T2DM, 2.8%; control, 5.1%; OR = 0.51; 95% CI, 0.27–1.08), but the difference was not significant because of the small sample. As shown in Table 2, the P582S mutant allele was found with significantly less frequency in T2DM patients than in control subjects ($P = 0.0028$; 1 − β = 52%) and also was significant after adjustment for sex, age, and BMI by logistic regression analysis ($P = 0.0048$). Interestingly, the P582S mutant allele is completely assigned on the 2-2 haplotype of SNP-25 and SNP-13, which was observed more frequently in control subjects with strong statistical significance (Fig. 3). Accordingly, the mutant allele S582 is a representative SNP of the 2-2 haplotype, which contributes to decreased risk of T2DM. Indeed, if a dominant model is assumed, the S582 HIF-1α mutation is associated with decreased risk of T2DM (OR = 0.57; 95% CI, 0.37–0.88; $P = 0.010$).

Similar haplotypes were also identified at significant levels in patients with and without retinopathy, but none reached statistical significance after multiple adjustment (data not shown). We also performed an association study using single polymorphisms in patients with and without retinopathy, resulting in identification of two significant SNPs (SNP-15 and -29 at P values of 0.033 and 0.045, respectively), although these were also not statistically significant after multiple adjustment and adjustment for sex, age, duration, HbA_{1c}, BMI, T-chol, TG, presence of medicine for HT, and presence of insulin therapy by logistic regression analysis (data not shown).

Hypoxia-dependent transactivation of polymorphic HIF-1α

The transcriptional activity of the S582 HIF-1α mutant was then compared with that of wild type under normoxic or hypoxic conditions. The HIF-1α vectors were transfected into HEK293 cells with a firefly luciferase reporter gene regulated by human VEGF gene promoter and with a control TK-regulated Renilla luciferase vector. The mutant S582 HIF-1α showed a hypoxia-dependent increase in transcription ac-

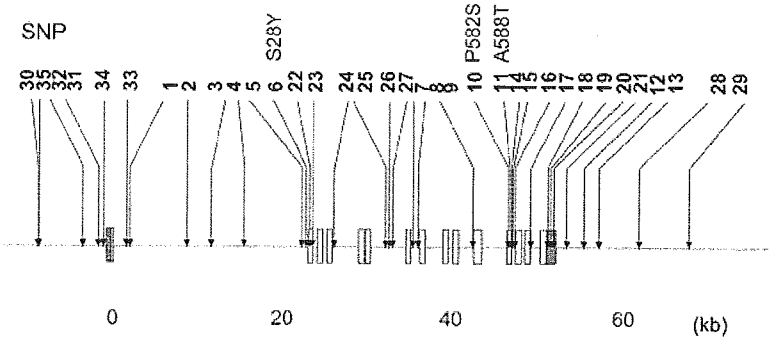

FIG. 1. Polymorphisms of *HIF1A* identified in this study. Nucleotide indicates the location of the SNP relative to the A of ATG of the initiator Met of *HIF1A* (GenBank no. NT_026437).

TABLE 1. Polymorphisms identified in *HIF1A* region in this study

SNP	Position genome	JSNP ID	dbSNP ID	Amino acid change	Variation	Location	Frequency of minor allele
30	−8393		rs7400961		C→G	5′ flanking	0.19
35	−8088		rs12717492		G→A	5′ flanking	0.19
32[a]	−7058				A→G	5′ flanking	0.06
31[a]	−4800				T→C	5′ flanking	0.09
34[a]	−1169				ACT/−	5′ flanking	0.01
33[a]	−907				T→C	5′ flanking	0.01
1	2506	IMS-JST057041	rs2301104		G→C	Intron 1	0.09
2	2598	IMS-JST057042	rs2301105		C→T	Intron 1	0.16
3	8904	IMS-JST023401	rs1951795		C→A	Intron 1	0.23
4	23170	IMS-JST140142	rs3783752		G→A	Intron 1	0.22
5	23956	IMS-JST035040	rs2284999		T→C	Intron 1	0.19
6[a]	24625			S28Y	C→A	Exon 2	0.01
22	25200		rs10137588		G→T	Intron 2	0.19
23	25299		rs10148514		T→C	Intron 2	0.19
24	27009		rs4899056		C→T	Intron 4	0.19
25	34776		rs12434438		A→G	Intron 6	0.25
26	34942	IMS-JST057045	rs2301108		G→A	Intron 6	0.19
27	35074	IMS-JST057046	rs2301109		A→G	Intron 6	0.19
7	37679	IMS-JST057048	rs2301111		C→G	Intron 7	0.22
8	37996	IMS-JST057049	rs966824		C→T	Intron 7	0.18
9	44026	IMS-JST057051	rs2301113		A→C	Intron 10	0.45
10	45035		rs11549465	P582S	C→T	Exon 12	0.06
11	45053		rs11549467	A588T	G→A	Exon 12	0.04
14	45483		rs4902080		C→T	Intron 12	0.21
15	45952		rs4902081		C→G	Intron 12	0.46
16	46572		rs8020184		T→C	Intron 12	0.20
17	46820	IMS-JST140141	rs3783751		G→C	Intron 12	0.21
18[a]	48445				G→T	Intron 14	0.23
19[a]	50292				C→T	3′ UTR	0.05
20	51326		rs2057482		C→T	3′ UTR	0.17
21	52853		rs994740		C→T	3′ UTR	0.16
12[a]	53485				C→G	3′ flanking	0.25
13	56703		rs1319462		A→G	3′ flanking	0.20
28	62194		rs7143626		G→A	3′ flanking	0.50
29	68040		rs2165601		T→C	3′ flanking	0.16

Nucleotide indicates the location of the SNP relative to the A of ATG of the initiator Met of *HIF1A* (GenBank no. NT_026437). Frequencies of minor alleles of SNPs in this table are observed in random control samples. JSNP, Japanese Single Nucleotide Polymorphism; dbsnp, SNPs deposited in the NCBI.

[a] Novel polymorphism.

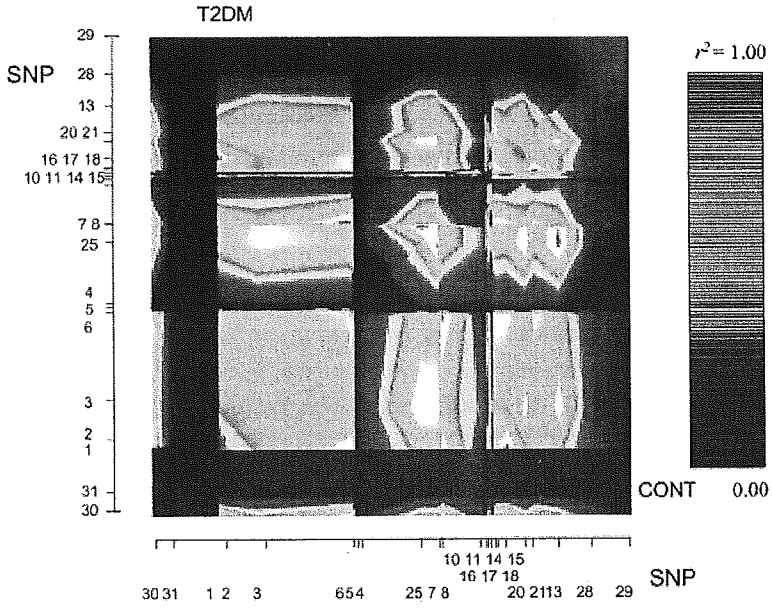

FIG. 2. Pairwise LD in *HIF1A* evaluated by r^2. Pairwise LD was determined using 276 marker pairs. Color gradations from *red* (perfect LD, *i.e.* $r^2 = 1$) to *blue* (no LD, *i.e.* $r^2 = 0$) reflect the degree of the observed LD. The *upper triangle* shows LD pattern estimated with 440 T2DM patients, and the *lower triangle* shows that with 576 controls (CONT).