J Hum Genet (2005) 50:92–98
DOI 10.1007/s10038-004-0225-5

ORIGINAL ARTICLE

Naoko Iwasaki · Yukio Horikawa · Takafumi Tsuchiya
Yutaka Kitamura · Takahiro Nakamura
Yukio Tanizawa · Yoshitomo Oka · Kazuo Hara
Takashi Kadowaki · Takuya Awata · Masashi Honda
Katsuko Yamashita · Naohisa Oda · Li Yu
Norihiro Yamada · Makiko Ogata · Naoyuki Kamatani
Yasuhiko Iwamoto · Laura del Bosque-Plata
M. Geoffrey Hayes · Nancy J. Cox · Graeme I. Bell

Genetic variants in the calpain-10 gene and the development of type 2 diabetes in the Japanese population

Received: 14 September 2004 / Accepted: 10 December 2004 / Published online: 5 February 2005

Abstract Variation in the gene encoding the cysteine protease calpain-10 has been linked and associated with risk of type 2 diabetes. We have examined the effect of three polymorphisms in the calpain-10 gene (SNP-43, Indel-19, and SNP-63) on the development of type 2 diabetes in the Japanese population in a pooled analysis of 927 patients and 929 controls. We observed that SNP-43, Indel-19, and SNP-63 either individually or as a haplotype were not associated with altered risk of type 2 diabetes with the exception of the rare 111/221 haplogenotype (odds ratio (OR) = 3.53, $P = 0.02$). However, stratification based on the median age-at-diagnosis in the pooled study population (< 50 and ≥50 years) revealed that allele 2 of Indel-19 and the 121 haplotype were associated with reduced risk in patients with later age-at-diagnosis (age-at-diagnosis ≥50 years OR = 0.82 and 0.80, respectively; $P = 0.04$ and 0.02). Thus, variation in the calpain-10 gene may affect risk of type 2 diabetes in Japanese, especially in older individuals.

Keywords Association study · Age-at-diagnosis · Calpain-10 · Genetics · Polymorphism · Type 2 diabetes

Naoko Iwasaki, Yukio Horikawa and Takafumi Tsuchiya contributed equally to this work.

N. Iwasaki (✉) · M. Ogata · Y. Iwamoto
Diabetes Center, Tokyo Women's Medical University,
8-1 Kawada-cho, Shinjuku-ku,
Tokyo 162-8666, Japan
E-mail: niwasaki@dmc.twmu.ac.jp
Tel.: +81-3-33538111
Fax: +81-3-33581941

Y. Horikawa · L. Yu · N. Yamada
Laboratory of Molecular Genetics,
Department of Cell Biology,
Institute for Molecular and Cellular Regulation,
Gunma University, Maebashi, Japan

T. Tsuchiya · L. del Bosque-Plata · M. G. Hayes · N. J. Cox · G. I. Bell
Departments of Biochemistry and Molecular Biology,
Human Genetics and Medicine,
The University of Chicago,
Chicago, Illinois, USA

Y. Kitamura · T. Nakamura · N. Kamatani
Department of Statistical Genetics,
Institute of Rheumatology,
Tokyo Women's Medical University,
Tokyo, Japan

Y. Tanizawa
Division of Molecular Analysis of Human Disorders,
Department of Bio-Signal Analysis,
Yamaguchi University Graduate School of Medicine,
Ube, Japan

Y. Oka
Division of Molecular Metabolism and Diabetes,
Department of Internal Medicine,
Tohoku University, Sendai, Japan

K. Hara · T. Kadowaki
Department of Metabolic diseases,
Graduate School of Medicine and Faculty of Medicine,
University of Tokyo, Tokyo, Japan

T. Awata
Division of Endocrinology and Diabetes, Department of Medicine,
Saitama Medical School, Saitama, Japan

M. Honda
Shiseikai Daini Hospital, Tokyo, Japan

K. Yamashita
Seijin Igaku Medical Clinic, Tokyo, Japan

N. Oda
Department of Internal Medicine,
Fujita Health University School of Medicine,
Aichi, Japan

Introduction

The results of association and linkage studies indicate that multiple genes are involved in determining susceptibility to type 2 diabetes in Japanese with each gene making a modest contribution to overall risk (Seino et al. 2001; Mori et al. 2001, 2002; Iwasaki et al. 2003). The gene encoding the cysteine protease calpain-10 (CAPN10) was first found to be associated with risk of type 2 diabetes in studies carried out in Mexican Americans (Horikawa et al. 2000). Two recent meta-analyses and a large association study have confirmed that single nucleotide polymorphisms (SNP)-43 and SNP-44 are associated with a 1.19- and 1.17-fold increased risk, respectively, of type 2 diabetes (Weedon et al. 2003; Song et al. 2004). SNP-43 may be a functional polymorphism affecting transcriptional regulation of the calpain-10 gene (Horikawa et al. 2000; Baier et al. 2000). However, Indel-19 and SNP-63 are just tagging SNPs, and their effect on transcriptional regulation or other functions of calpain-10 are unknown. The effect of the core CAPN10 polymorphisms SNP-43, Indel-19, and SNP-63 on risk of type 2 diabetes in Japanese has been examined in three small studies (Daimon et al. 2002; Horikawa et al. 2003; Shima et al. 2003). The results suggest that variation in CAPN10 is not a major risk factor. However, these studies were not able to quantify the effect of CAPN10 on risk because of the small number of cases and controls in the individual studies. Here, we reexamine the role of the CAPN10 in the development of type 2 diabetes in the Japanese population.

Material and methods

Subjects

All subjects were Japanese. We studied three groups of cases and controls. The first group (study 1) included 205 unrelated subjects with type 2 diabetes recruited from the outpatient clinic in the Diabetes Center, Tokyo Women's Medical University and 208 unrelated normoglycemic subjects recruited from the Seijin Igaku Medical Clinic of Tokyo Women's Medical University using the following inclusion criteria: age >60 years, HbA1c $<5.6\%$, and no family history of diabetes. The second group (study 2) consisted of 281 unrelated normal glucose-tolerant (by oral glucose tolerance testing) subjects who were recruited at four outpatient clinics: Diabetes Center, Tokyo Women's Medical University ($n=50$); Third Department of Internal Medicine, Yamaguchi University ($n=121$); Department of Internal Medicine, University of Tokyo ($n=30$); and Shiseikai Daini Hospital, Tokyo ($n=80$). The third group of subjects (study 3) comprised 454 patients with type 2 diabetes and 192 nondiabetic controls who were recruited from Gunma University Hospital and affiliated hospitals and Fujita Health University School of Medicine. The genetic studies were approved by the institutional review board of each participating institution. Informed consent was obtained from all participants.

The pooled analyses included the datasets above as well as the data from three published studies: Daimon et al. 2002, SNP-43; Horikawa et al. 2003, SNP-43, Indel-19 and SNP-63; and Shima et al. 2003, SNP-43, Indel-19, and SNP-63.

Linkage disequilibrium (LD)

We examined the structure of the linkage disequilibrium (LD) in the CAPN10 region using the software package GOLD [graphical overview of linkage disequilibrium (Abecasis and Cookson 2000)] and a common set of 14 SNPs having a minor allele frequency ≥ 0.10 in diabetic ($n=30$) and nondiabetic subjects ($n=30$).

Genotyping

Genomic DNA was prepared from peripheral blood lymphocytes by standard procedures. We typed three polymorphisms in CAPN10: SNP-43, CAPN10-g.4852G>A (rs3792267); insertion/deletion (Indel)-19, CAPN10-g.7920 (32 bp-repeats) (rs3842570); and CAPN10-g.16378C>T (rs5030952) as described previously (Horikawa et al. 2003) or using TaqMan-based assays with custom probes/primers (Applied Biosystems, Foster City, CA, USA). Additional SNPs used for studies of LD in the CAPN10 region were genotyped using TaqMan technology. Previous studies have shown that the three core polymorphisms lead to four common haplotypes described as 111, 112, 121, and 221 (allele 1 or 2 at SNP-43, Indel-19, and SNP-63, respectively). The haplogenotypes were assigned by inspection of the genotypes at SNP-43, Indel-19, and SNP-63.

Statistical analyses

Polymorphisms were tested for deviation from Hardy-Weinberg equilibrium, heterogeneity in allele and genotype among studies and differences in allele, genotype, haplotype, and haplogenotype between groups using a chi-squared test. All P values are two sided.

Results

Haplotype structure across the CAPN10 region

The analysis of LD in the region of CAPN10 revealed a single region of strong LD (Fig. 1).

CAPN10 and SNP-43, and Indel-19 and SNP-63 are contained within this single LD block, which does not

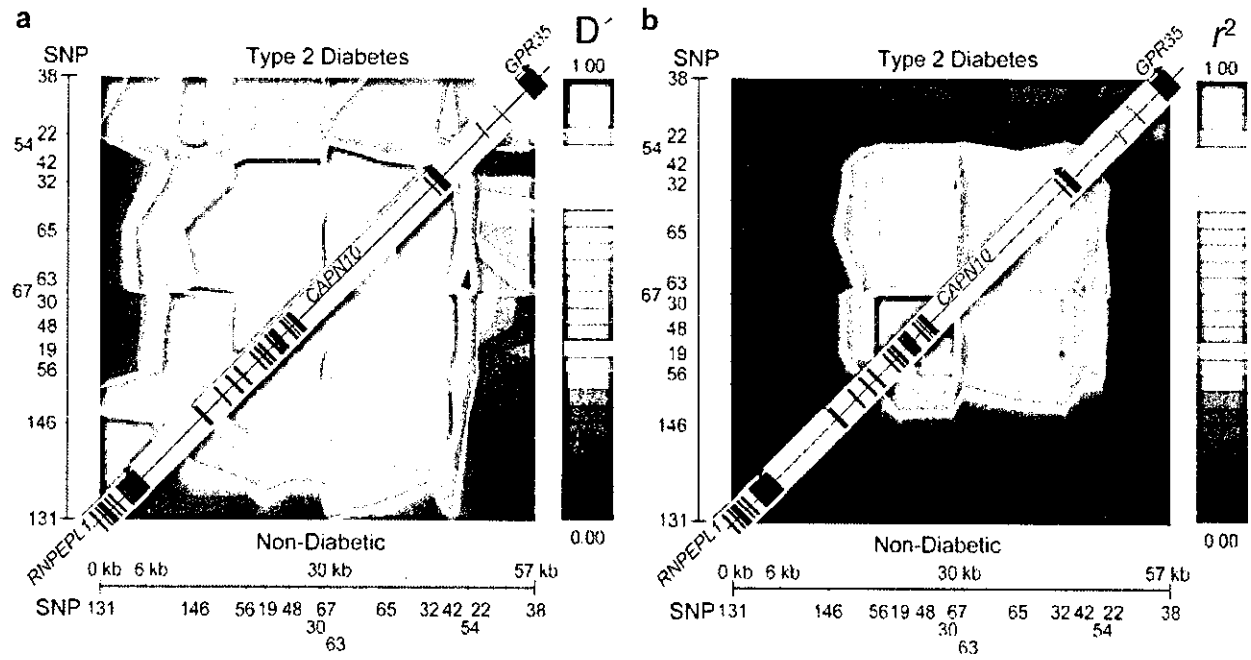

Fig. 1 Linkage disequilibrium (LD) in the CAPN10 region visualized using GOLD. The red and orange regions denote strong LD as defined using D' and r^2. The exons of CAPN10 and the adjacent genes RNPEPL1 and GPR35 are shown as *filled boxes* along the diagonal. The two variable number of tandem repeats (*VNTRs*) between CAPN10 and GPR35 are shown as *open boxes*. The SNPs used in this analysis are described in Horikawa et al. (2000). SNP-43 was not included in this analysis because the minor allele frequency was < 0.10 in the Japanese population

include the flanking RNPEPL1 and GPR35 genes. Thus, association of polymorphisms in this block with type 2 diabetes or a type-2-diabetes-related trait in the Japanese population implies that it is a variation in CAPN10 itself and not an adjacent locus that is responsible for the effect. The associated variant may or may not be causal depending on the LD with other variants in the block.

Genetic variation in CAPN10 and type 2 diabetes

We typed SNP-43, Indel-19, and SNP-63 in the three study groups described above (Table 1). There was no significant difference in the frequency of SNP-43 or Indel-19 between cases and controls (Table 2), which is in agreement with previous studies in Japanese (Daimon et al. 2002; Horikawa et al. 2003). However, we observed a significant difference in SNP-63 allele frequency between cases and controls in the subjects from study 1 (0.66 and 0.73, respectively, $P=0.04$) but not in the subjects from study 3 (Table 2). In order to gain a better understanding of what role SNP-63 may play in the progression of type 2 diabetes, we examined the effect of SNP-63 genotype on various clinical and metabolic characteristics assessed by a standard 75-g oral glucose tolerance test in a group of 281 normal glucose-tolerant subjects from study 2 (Table 3). No significant effects of SNP-63 genotype on phenotype were observed except for area-under-the-curve plasma glucose level from 0 min to 120 min ($P=0.03$).

We then carried out a pooled analysis using data from all known studies carried out in the Japanese population (Daimon et al. 2002; Horikawa et al. 2003; Shima et al. 2003). The 281 nondiabetic subjects from study 2 were excluded from the primary analyses because their mean age-at-study in this group was significantly younger than other the control groups. The pooled study population included 927 patients and 929 controls although Indel-19 and SNP-63 were not typed in all subjects. There was no significant difference in SNP-43, Indel-19, or SNP-63 genotype or allele frequencies between the type 2 diabetic and control groups in the overall analysis (Table 4). There was also no significant difference in SNP-43/Indel-19/SNP-63 haplotype (Table 5) or haplogenotype frequency (Table 6) except for the rare 111/221 combination, which was associated with significantly increased risk of type 2 diabetes (OR = 3.53, $P=0.02$).

Genetic variation in CAPN10 may modify risk of type 2 diabetes in older patients

Since age is a risk factor for type 2 diabetes, we split the type 2 diabetic group based on median age-at-diagnosis, which was 50 years in the pooled sample, and repeated the comparisons but using only those cases for whom age-at-diagnosis was available: Patients with age-at-diagnosis < 50 years included 118 patients from study 1, 103 from study 3, and 68 from Horikawa et al. (2003); and patients with age-at-diag-

Table 1 Clinical characteristics of study populations. Data are mean ± SD. *NA* data not available, *Ctrl* control, *T2D* type 2 diabetes

Trait	Study population											
	Study 1		Study 2		Study 3		Horikawa et al. (2003)		Shima et al. (2003)		Daimon et al. (2002)	
Subjects	Ctrl[a]	T2D	Ctrl	T2D	Ctrl[a]	T2D	Ctrl[a]	T2D	Ctrl	T2D	Ctrl	T2D
n	208	205	281	...	192	454	172	177	276	10	81	81
Gender (F/M)	68/140	84/121	135/146	...	70/122	206/248	90/82	63/114	NA	NA	46/35	45/36
Age-at-study (years)	67.9 ± 5.5	59.1 ± 13.0	44.4 ± 15.9	...	67.8 ± 5.6	59.9 ± 11.7	68.0 ± 5.7	62.0 ± 11.0	NA	NA	62.3 ± 8.2	65.1 ± 10.7
Age-at-diagnosis (years)	...	45.9 ± 12.7	...	...	...	51.0 ± 12.0[b]	...	49.8 ± 11.4	...	NA	...	NA
BMI	23.1 ± 2.4	23.5 ± 3.6	22.3 ± 3.0	...	23.1 ± 2.8	24.2 ± 4.2	22.8 ± 3.3	23.9 ± 3.3	NA	NA	23.8 ± 3.6	25.6 ± 3.9
HbA_{1c} (%)	5.0 ± 0.3	8.2 ± 2.0	4.9 ± 0.4	...	4.9 ± 0.3	7.9 ± 1.9	5.0 ± 0.4	6.7 ± 1.0	NA	NA	5.3 ± 0.3	6.7 ± 1.1
Fasting glucose (mg/dl)	95.0 ± 8.7	160.5 ± 48.5	92.8 ± 9.4	...	...	NA	NA	NA	NA	NA	NA	NA
Treatment (diet/oral agents/insulin)	...	42/81/82	...	...	...	NA	...	46/70/61	...	NA	...	NA

[a] All subjects were > 60 years old
[b] Data are available for only 246 subjects

nosis ≥50 years included 87 patients from study 1, 143 from study 3, and 85 from Horikawa et al. (2003). There was no significant difference in SNP-43, Indel-19, and SNP-63 allele or genotype frequencies between the type 2 diabetic group with age-at-diagnosis < 50 years and the controls (Table 4). The SNP-43 and SNP-63 frequencies were also not different between the type 2 diabetic group with age-at-diagnosis ≥50 years and the controls. However, there was a small but significant difference in Indel-19 allele frequency (Table 4). The 3R allele at Indel-19 (allele 2 in the haplotype) was associated with lower risk of type 2 diabetes (OR = 0.82, $P = 0.04$).

The 121 haplotype was associated with significantly decreased risk (OR = 0.80, $P = 0.02$) of type 2 diabetes in the group of patients with age-at-diagnosis ≥50 years (Table 5). The 111 haplotype had the highest risk (OR = 1.33, $P = 0.046$) in the older group of patients. This effect of the 111 haplotype likely reflects the contribution of SNP-44 to type 2 diabetes risk (Weedon et al. 2003) since 88% of the 111 haplotypes in Japanese carry the at-risk C-allele at SNP-44. The rare 111/221 haplogenotype was associated with increased risk of type 2 diabetes irrespective of age-at-diagnosis (Table 6). The 121/121 haplogenotype had a protective effect against type 2 diabetes that approached significance in patients with age-at-diagnosis ≥50 years (OR = 0.76, $P = 0.06$). Individuals with haplotypes 111/111, 111/112, and 111/221 had the highest risk of type 2 diabetes (OR = 1.83, 1.25, and 4.12, respectively) although the increase in risk was not significant because of the small numbers of individuals studied. If the nondiabetic subjects in study 2 (Table 1) are included in the pooled control group, the results are similar, including the effects of the 121/121 haplogenotype on risk in patients with age-at-diagnosis ≥50 years (OR = 0.75, $P = 0.04$).

Discussion

The results suggest that genetic variation in CAPN10 may affect risk of type 2 diabetes in the Japanese population, especially in older individuals. Interestingly, the common 121 haplotype appears to be protective in Japanese, suggesting the overall effect of CAPN10 in this population is to reduce the risk of diabetes rather than increase it. It is important to note, though, that the statistical significance of the comparison is marginal ($P = 0.01$–0.04), and none of the comparisons would be significant if corrected for multiple testing. Thus, the results presented here need to be confirmed through studies of a much larger dataset. However, if our results are correct, they suggest an interaction between genetic (CAPN10) and nongenetic (age) factors to modify risk of type 2 diabetes. In this regard, recent studies have shown that calpain-10 is part of a novel apoptotic pathway in insulin-secreting pancreatic beta cells and thus may

Table 2 Genotype and allele frequencies of CAPN10 polymorphisms in Japanese. The number of subjects of each genotype are indicated. All genotypic distributions are in Hardy-Weinberg equilibrium. *NA* not available. *Ctrl* control, *T2D* type 2 diabetes

Marker	Subjects	Genotype	This study 1	This study 2	This study 3	Horikawa et al. (2003)	Shima et al. (2003)	Daimon et al. (2002)	*P* for heterogeneity
SNP-43	Ctrl	G/G	188	251	165	154	252	76	
		G/A	20	29	24	18	24	5	
		A/A	0	1	0	0	0	0	0.75
		Allele frequency	G: 0.95	G: 0.94	G: 0.94	G: 0.95	G: 0.96	G: 0.97	
			A: 0.05	A: 0.06	A: 0.06	A: 0.05	A: 0.04	A: 0.03	0.62
	T2D	G/G	184	NA	389	158	8	76	
		G/A	21	NA	57	19	2	5	
		A/A	0	NA	1	0	0	0	0.75
		Allele frequency	G: 0.95	NA	G: 0.93	G: 0.95	G: 0.90	G: 0.97	
			A: 0.05		A: 0.07	A: 0.05	A: 0.10	A: 0.03	0.37
Indel-19[a]	Ctrl	2R/2R	27	42	35	23	42	NA	
		2R/3R	99	126	78	78	126	NA	
		3R/3R	82	113	73	71	108	NA	0.90
		Allele frequency	2R: 0.37	2R: 0.37	2R: 0.40	2R: 0.36	2R: 0.38	NA	
			3R: 0.63	3R: 0.63	3R: 0.60	3R: 0.64	3R: 0.62		0.86
	T2D	2R/2R	32	NA	63	28	1	NA	
		2R/3R	104	NA	209	82	3	NA	
		3R/3R	69	NA	176	67	6	NA	0.62
		Allele frequency	2R: 0.41	NA	2R: 0.37	2R: 0.39	2R: 0.25	NA	
			3R: 0.59		3R: 0.63	3R: 0.61	3R: 0.75		0.34
SNP-63	Ctrl	C/C	111	151	99	90	151	NA	
		C/T	81	106	65	70	103	NA	
		T/T	16	24	18	12	22	NA	0.99
		Allele frequency	C: 0.73	C: 0.73	C: 0.72	C: 0.73	C: 0.73	NA	
			T: 0.27	T: 0.27	T: 0.28	T: 0.27	T: 0.27		0.99
	T2D	C/C	90	NA	255	93	6	NA	
		C/T	92	NA	165	74	3	NA	
		T/T	23	NA	30	10	1	NA	0.09
		Allele frequency	C: 0.66	NA	C: 0.75	C: 0.73	C: 0.75	NA	
			T: 0.34		T: 0.25	T: 0.27	T: 0.25		0.03

[a] Indel-19 is a diallelic insertion/deletion polymorphism with alleles of two repeats (2R) or three repeats (3R) of a 32-bp sequence

Table 3 Clinical and metabolic characteristics of normal glucose tolerant subjects (study 2) by SNP-63 genotype. Data are mean ± SD. Subjects underwent a standard 75-g oral glucose tolerance test with glucose and insulin determined at 0′, 30′, 60′, and 120′. The number of individuals in each group for determination of insulinogenic index and HOMA are noted in parentheses. *BMI* body mass index, *AUC* area under the curve

Trait	Genotype			*P*[a]
	C/C	C/T	T/T	
n	150	105	25	
Gender (M/F)	72/78	59/46	14/11	0.40
Age (years)	45.0 ± 14.9	43.8 ± 16.7	44.4 ± 18.7	0.72
BMI (kg/m2)	22.3 ± 2.5	22.2 ± 3.6	22.6 ± 3.0	0.70
HbA1c (%)	4.9 ± 0.4	4.9 ± 0.4	5.0 ± 0.4	0.79
Plasma glucose (mg/dl)				
0 min	93.1 ± 9.6	92.2 ± 9.1	94.1 ± 8.8	0.51
120 min	104.1 ± 18.5	106.1 ± 18.8	108.2 ± 14.5	0.30
AUC 0-120′	14,510.6 ± 2706.2	15,233.4 ± 2937.1	15,043.2 ± 3056.1	0.03
Plasma insulin (μU/ml)				
0 min	6.4 ± 2.2	6.4 ± 2.6	6.5 ± 4.4	0.81
120 min	30.6 ± 16.9	32.0 ± 16.8	30.1 ± 20.2	0.74
AUC 0-120′	3,972.4 ± 2030.8	4,267.4 ± 1816.9	4,049.2 ± 1969.1	0.44
Insulinogenic index	0.89 ± 0.89 (147)	0.96 ± 3.62 (101)	1.14 ± 1.85 (24)	0.08
HOMA	1.43 ± 0.53 (146)	1.41 ± 0.54 (101)	1.30 ± 0.63 (23)	0.35

[a] *P* value by ANCOVA with institution, gender, and genotype as independent factors and age and BMI as covariates

affect the response of the beta cell to aging or its ability to compensate in response to an increasing demand for insulin (Johnson et al. 2004).

The observation that individuals homozygous for the 121 haplotype may be at increased risk of type 2 diabetes in some European populations (Orho-

Table 4 SNP-43, Indel-19, and SNP-63 and type 2 diabetes in Japanese—a pooled analysis. The number of subjects of each genotype are indicated. All genotypic distributions are in Hardy-Weinberg equilibrium. The cases were divided into two groups based on the median age-at-diagnosis in the pooled sample—50 years. Note that age-at-diagnosis was not available for all subjects

Marker	Genotype	Overall			Age-at-diagnosis < 50 years		Age-at-diagnosis ≥50 years	
		Ctrl	T2D	*P*	T2D	*P*	T2D	*P*
SNP-43	G/G	833	811		255		276	
	G/A	91	104		28	0.98	32	0.78
	A/A	0	1	0.35	0		0	
	Allele frequency	G: 0.95	G: 0.94		G: 0.95		G: 0.95	
		A: 0.05	A: 0.06	0.25	A: 0.05	0.98	A: 0.05	0.79
Indel-19[a]	2R/2R	127	124		36		58	
	2R/3R	381	396		143		151	0.11
	3R/3R	334	316	0.67	107	0.33	105	
	Allele frequency	2R: 0.38	2R: 0.39		2R: 0.38		2R: 0.43	
		3R: 0.62	3R: 0.61	0.69	3R: 0.62	0.91	3R: 0.57	0.04
SNP-63	C/C	451	443		149		151	
	C/T	319	333		125		129	
	T/T	68	62	0.72	14	0.09	28	0.35
	Allele frequency	C: 0.73	C: 0.73		C: 0.73		C: 0.71	
		T: 0.27	T: 0.27	0.94	T: 0.27	0.79	T: 0.29	0.17

[a] Indel-19: 2R, 2 repeats of 32-bp sequence; 3R, 3 repeats

Table 5 CAPN10 haplotype frequency and risk of type 2 diabetes in Japanese—a pooled analysis. The haplotypes are those defined by SNP-43, Indel-19, and SNP-63, and the specific alleles are: SNP-43, allele 1, G and allele 2, A; Indel-19, allele 1, 2 repeats of 32-bp sequence, and allele 2, 3 repeats; and SNP-63, allele 1, C, and allele 2, T. The cases were divided into two groups based on the median age-at-diagnosis in the pooled sample—50 years. Note that age-at-diagnosis was not available for all subjects. *Ctrl* control, *T2D* type 2 diabetes

Haplotype	Overall				Age-at-diagnosis < 50 years			Age-at-diagnosis ≥50 years		
	Ctrl (n=825)	T2D (n=827)	OR (95% CI)[a]	*P*	T2D (n=277)	OR (95% CI)[a]	*P*	T2D (n=305)	OR (95% CI)[a]	*P*
111	0.106	0.113	1.07 (0.86–1.34)	0.52	0.117	1.12 (0.83–1.52)	0.46	0.136	1.33 (1.00–1.75)	0.05
121	0.572	0.553	0.93 (0.81–1.07)	0.29	0.567	0.98 (0.81–1.19)	0.85	0.515	0.80 (0.66–0.96)	0.02
112	0.270	0.274	1.02 (0.88–1.19)	0.79	0.265	0.98 (0.78–1.21)	0.82	0.297	1.14 (0.93–1.40)	0.21
221	0.052	0.059	1.15 (0.85–1.54)	0.37	0.051	0.97 (0.62–1.50)	0.88	0.052	1.01 (0.66–1.53)	0.97

[a] The OR and 95% CI of each haplotype relative to other haplotypes as a group are shown

Table 6 CAPN10 haplogenotype and risk of type 2 diabetes in Japanese—a pooled analysis. The haplotypes are those defined by SNP-43, Indel-19, and SNP-63, and the specific alleles are indicated in the legend to Table 5. The number of individuals with each haplogenotype is indicated

Haplogenotype	Overall				Age-at-diagnosis < 50 years			Age-at-diagnosis ≥ 50 years		
	Ctrl	T2D	OR (95% CI)[a]	*P*	T2D	OR (95% CI)[a]	*P*	T2D	OR (95% CI)[a]	*P*
111/111	15	18	1.20 (0.60–2.40)	0.60	5	0.99 (0.36–2.76)	0.99	10	1.83 (0.82–4.07)	0.14
111/121	97	93	0.95 (0.70–1.29)	0.74	31	0.95 (0.62–1.45)	0.80	37	1.04 (0.69–1.55)	0.86
111/112	44	44	1.00 (0.65–1.53)	0.99	18	1.23 (0.70–2.17)	0.47	20	1.25 (0.72–2.15)	0.43
111/221	4	14	3.53 (1.24–10.1)	0.02	6	4.54 (1.42–14.5)	0.01	6	4.12 (1.27–13.3)	0.02
112/112	66	62	0.93 (0.65–1.34)	0.70	13	0.57 (0.31–1.04)	0.06	27	1.12 (0.70–1.78)	0.64
112/121	247	254	1.04 (0.84–1.28)	0.73	90	1.13 (0.84–1.51)	0.43	97	1.09 (0.82–1.45)	0.55
112/221	23	32	1.40 (0.82–2.41)	0.22	13	1.72 (0.86–3.41)	0.12	10	1.18 (0.56–2.51)	0.66
121/121	270	259	0.94 (0.76–1.15)	0.54	92	1.02 (0.77–1.37)	0.88	82	0.76 (0.56–1.01)	0.06
121/221	59	50	0.84 (0.57–1.23)	0.37	9	0.44 (0.22–0.87)	0.02	16	0.72 (0.41–1.27)	0.25
221/221	0	1	-	-	0			0		

[a] The OR and 95% CI of each haplogenotype relative to the other haplotype combinations as a group are shown

Melander et al. 2002; Malecki et al. 2002) but at decreased risk in older Japanese raises the possibility that additional genetic variation may distinguish high- and low-risk subtypes of the 121 haplotype. Transpopulation mapping may be a useful strategy for identifying this variation.

Acknowledgements The authors thank Ms. A. Nogami, Ms. M. Y. Sagisaka, and Mr. M. Ikeda for their skillful technical assistance. This study was supported by Grants-in-Aid for Scientific Research C (10671084, 10470234) and for Scientific Research on Priority Areas Medical Genome Science from the Japan Ministry of Science, Education, Sports, Culture and Technology (12204102, 13204082, 14013059, 15012250), Novo Nordisk Foundation, the Naito Foundation (to N.I.) and Grants-in-Aid for Scientific Research B (13470223, 13557091) (to Y.H.), and U.S. Public Health Service (Grants DK-20595, −47486 and −55889). G.I.B. is an Investigator of the Howard Hughes Medical Institute.

References

Abecasis GR, Cookson WO (2000) GOLD–graphical overview of linkage disequilibrium. Bioinformatics 16:182–183

Baier LJ, Permana PA, Yang X, Pratley RE, Hanson RL, Shen GQ, Mott D, Knowler WC, Cox NJ, Horikawa Y, Oda N, Bell GI, Bogardus C (2000) A 1-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. J Clin Invest 106:R69-R73

Daimon M, Oizumi T, Saitoh T, Kameda W, Yamaguchi H, Ohnuma H, Igarashi M, Manaka H, Kato T (2002) Calpain-10 gene polymorphisms are related, not to type 2 diabetes, but to increased serum cholesterol in Japanese. Diabetes Res Clin Pract 56:147–152

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26:163–175

Horikawa Y, Oda N, Yu L, Imamura S, Fujiwara K, Makino M, Seino Y, Itoh M, Takeda J (2003) Genetic variations in calpain-10 gene are not a major factor in the occurrence of type 2 diabetes in Japanese. J Clin Endocrinol Metab 88:244–247

Iwasaki N, Cox NJ, Wang YQ, Schwarz PE, Bell GI, Honda M, Imura M, Ogata M, Saito M, Kamatani N, Iwamoto Y (2003) Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. Diabetes 52:209–213

Johnson JD, Han Z, Otani K, Ye H, Zhang H, Wu H, Horikawa Y, Misler S, Bell GI, Polonsky KS (2004) RyR2 and calpain-10 delineate a novel apoptosis pathway in pancreatic islets. J Biol Chem 279:24794–24802

Malecki MT, Moczulski DK, Klupa T, Wanic K, Cyganek K, Frey J, Sieradzki J (2002) Homozygous combination of calpain 10 gene haplotypes is associated with type 2 diabetes mellitus in a Polish population. Eur J Endocrinol 146:695–699

Mori H, Ikegami H, Kawaguchi Y, Seino S, Yokoi N, Takeda J, Inoue I, Seino Y, Yasuda K, Hanafusa T, Yamagata K, Awata T, Kadowaki T, Hara K, Yamada N, Gotoda T, Iwasaki N, Iwamoto Y, Sanke T, Nanjo K, Oka Y, Matsutani A, Maeda E, Kasuga M (2001) The Pro12 → Ala substitution in PPAR-γ is associated with resistance to development of diabetes in the general population: possible involvement in impairment of insulin secretion in individuals with type 2 diabetes. Diabetes 50:891–894

Mori Y, Otabe S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P (2002) Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate Loci on 7p and 11p. Diabetes 51:1247–1255

Orho-Melander M, Klannemark M, Svensson MK, Ridderstrale M, Lindgren CM, Groop L (2002) Variants in the calpain-10 gene predispose to insulin resistance and elevated free fatty acid levels. Diabetes 51:2658–2664

Seino S on behalf of the Study Group of Comprehensive Analysis of Genetic Factors in Diabetes Mellitus (2001) S20G mutation of the amylin gene is associated with Type II diabetes in Japanese. Diabetologia 44:906–909

Shima Y, Nakanishi K, Odawara M, Kobayashi T, Ohta H (2003) Association of the SNP-19 genotype 22 in the calpain-10 gene with elevated body mass index and hemoglobin A1c levels in Japanese. Clin Chim Acta 336:89–96

Song Y, Niu T, Manson JE, Kwiatkowski DJ, Liu S (2004) Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. Am J Hum Genet 74:208–222

Weedon MN, Schwarz PE, Horikawa Y, Iwasaki N, Illig T, Holle R, Rathmann W, Selisko T, Schulze J, Owen KR, Evans J, Del Bosque-Plata L, Hitman G, Walker M, Levy JC, Sampson M, Bell GI, McCarthy MI, Hattersley AT, Frayling TM (2003) Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. Am J Hum Genet 73:1208–1212

DNA RESEARCH 11, 315–323 (2004)

Construction of a Multi-Functional cDNA Library Specific for Mouse Pancreatic Islets and Its Application to Microarray

Motoi NISHIMURA,[1,†] Norihide YOKOI,[2] Takashi MIKI[2,3] Yukio HORIKAWA[3,4] Hirokazu YOSHIOKA,[5] Jun TAKEDA,[3,6] Osamu OHARA,[7,8] and Susumu SEINO[1,2,*]

Department of Cellular and Molecular Medicine, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan,[1] Division of Cellular and Molecular Medicine, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan,[2] Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), 4-1-8 Honcho, Kawaguchi 332-0012, Japan,[3] Institute for Molecular and Cellular Regulation, Gunma University, 3-39-15 Showa-machi, Maebashi 371-8512, Japan,[4] Kakengeneqs Corporation, 439-1 Matsuhidai, Matsudo 270-2214, Japan,[5] Department of Endocrinology, Diabetes and Rheumatology, Division of Bioregulatory Medicine, Gifu University School of Medicine, 1-1 Yanagido, Gifu 501-1194, Japan,[6] Department of Human Gene Research, Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan,[7] and Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan[8]

(Received 19 January 2004: revised 24 August 2004)

Abstract

We have constructed a high-quality and multi-applicable cDNA library specific for mouse pancreatic islets. This is the first pancreatic islet cDNA library created using a recombination-based method, which can readily be converted into other applications including yeast two-hybrid and mammalian expression libraries. Based on sequence data of the library, we constructed a sequence database specific for mouse pancreatic islets. Among the 8882 non-redundant clones, 5799 were classified into specific functional categories using a classification system designed by the Gene Ontology Consortium, 10% of which were "molecular function unknown" genes. We also developed cDNA microarray membranes with 8108 non-redundant clones. Analyses of expression profiles of three different cell lines and of MIN6 cells with or without overexpression of transcription factor NeuroD1 established the usefulness and applicability of our microarrays. The mouse pancreatic islet cDNA library, sequence database, set of clones, and microarrays developed in this study should be useful resources for studies of pancreatic islets and related diseases including diabetes mellitus.

Key words: cDNA library; microarray; pancreatic islet; recombination-based method; sequence database

1. Introduction

Pancreatic islets are small inner parts of pancreata, and play the central role in glucose homeostasis.[1] The majority of islet cells are β-cells, which secrete the hypoglycemic hormone insulin; other islet cells (α-cells, δ-cells and PP cells) also secrete hormones affecting glucose homeostasis. Impairment of pancreatic β-cell function readily causes disorders of glucose homeostasis such as diabetes.

Studies of knockout mice[2–6] and a subtype of human diabetes[7–12] have revealed that impairment of the transcription network in pancreatic β-cells is involved in the pathogenesis of diabetes. In these models, diabetes occurs due to disruptions or mutations in transcription factors that regulate gene expression in pancreatic β-cells. Thus, construction of a gene expression database for pancreatic islets and comprehensive analysis of expression profiles in the pancreatic islets might provide valuable information on the molecular mechanisms of the disease.

To investigate the molecules involved in normal function of pancreatic islets, several groups have constructed cDNA libraries from human pancreatic islets and sequenced clones from the libraries (In the GenBank database (www.ncbi.nlm.nih.gov/genbank/index.html), nucleotide sequences of 4559 clones were deposited by

Communicated by Michio Oishi

* To whom correspondence should be addressed. Kobe University Tel. +81-78-382-5360, Fax. +81-78-382-5370, E-mail: seino@med.kobe-u.ac.jp

† Present address: Division of Cellular and Molecular Medicine, Kobe University Graduate School of Medicine, Kobe, Japan

Table 1. Characteristics of the mouse pancreatic islet cDNA library.

Starting material	About 5,000 pancreatic islets from C57BL/6 male mice
Complexity without amplification	3.12×10^7

Fraction by insert size[1)]	1	2	3	4	5	6
Proportion of chimeric clones (%)[2)]	4.23	2.76	4.60	6.45	1.84	1.10
Proportion of clones with full ORFs (%)[3)]	64	56	48	76	72	52

1) Insert sizes of the fractions are: 1, >8 kbp; 2, 6–8 kbp; 3, 4–6 kbp; 4, 3–4 kbp; 5, 1.5–3 kbp; 6, <1.5 kbp.
2) Proportions of chimeric clones are calculated on the analysis of 3264 clones sequenced from both ends.
3) Proportions of clones with full ORFs are estimated from 25 randomly selected clones from each fraction.

Bell and colleagues, Chicago University, and those of 2055 clones by the I.M.A.G.E. consortium). As mice are widely used as experimental animals in studies of pancreatic islets,[1] a cDNA library specific for mouse pancreatic islets should be useful. Recently, a recombination-based method for cDNA library construction has been developed.[13] With size fractionation of the cDNA inserts, this method provides several improvements.[13,14] The characteristics of the library are a high degree of complexity, an abundance of full-length sequences, and multiple applications. Libraries constructed by this method can readily be converted into other libraries including yeast two-hybrid and mammalian expression libraries.[15]

In the present study, we constructed a recombination-based cDNA library specific for mouse pancreatic islets. Based on the sequence data of the library, we built a mouse pancreatic islet sequence database, and also developed cDNA microarray membranes.

2. Materials and Methods

2.1. *Construction of a cDNA library specific for mouse pancreatic islets*

All animal procedures were approved by the Chiba University Animal Care Committee. Pancreatic islets were isolated from 8- to 10-week-old male C57BL/6 mice by hand-picking under the microscope and collagenase digestion method as described previously.[16] Poly(A)$^+$ RNA was obtained from the islets using RNeasy Mini kit (Qiagen GmbH, Hilden, Germany) and μMACS mRNA isolation kit (Miltenyi Biotech, Bergisch Gladbach, Germany). Two micrograms of poly(A)$^+$ RNA was isolated from approximately 5000 islets, from which cDNA synthesis was carried out as described previously.[14] The synthesized cDNA was size-fractionated by agarose gel electrophoresis into three segments (<1.5, 1.5–3, and >3 kbps, respectively), and subjected to recombination reaction with BP Clonase (Invitrogen, Carlsbad, CA) and plasmid attP pSP73.[14] Recombined plasmids were introduced into *Escherichia coli* cells (ElectroMAX DH10B cells, Invitrogen) by electroporation. The numbers of transformants from each fraction were counted to estimate the complexity of the library without any amplification (Table 1).

After transformants were grown, the retrieved plasmids from the largest size fraction (>3 kbps) were further size-fractionated into four segments on agarose gel. Thus, the extracted plasmids were finally size-separated into six segments (<1.5, 1.5–3, 3–4, 4–6, 6–8, and >8 kbps) and subjected to recombination reaction with LR clonase (Invitrogen) and *BamHI*-linearized plasmid attR pBC.[14] Recombined plasmids were again introduced into *E. coli* cells (ElectroMAX DH10B cells, Invitrogen) by electroporation. The direction of the cDNA inserts was T7-attB1-cDNA insert-attB2-T3.

2.2. *Plasmid preparation and DNA sequencing*

Plasmids for sequencing the cDNA inserts and spotting on microarray membranes were prepared with MAGNIA robot (TOYOBO, Osaka, Japan). With ABI3700 sequencers (Applied Biosciences, Mountain View, CA), 21,018 plasmid clones were successfully sequenced with T3 sequencing primer, and 3,264 clones were sequenced with T7 sequencing primer after elimination of redundant clones.

2.3. *Construction of a sequence database*

The sequence data obtained using T3 or T7 primers were subjected to BLAST search against the GenBank/DDBJ/EMBL nucleotide sequence database (www.ncbi.nlm.nih.gov/genbank/index.html) with a threshold E-value of $1.0e^{-50}$. The sequences were also subjected to BLAST search against the Ensembl mouse genome database (www.ensembl.org) to determine their locations on mouse chromosomes. All the sequence data were submitted to the DDBJ/EMBL/GenBank nucleotide sequence database (Accession nos. BP753069-BP777127), and are available from the ftp site at ftp://ftp.kazusa.or.jp/pub/pancreas/. The summary of the homology searches and the genome mapping data of the analyzed cDNAs are also available (Supplemental Information 1, http://www.dna-res.kazusa.or.jp/

11/5/01/supplemental/information1.html).

2.4. *cDNA microarray analysis*

The method of cDNA microarray analysis was described previously.[17] A total of 8108 cDNA clones were spotted on the microarray membranes. For single microarray analysis, 10 μg of total RNA from each sample was reverse-transcribed in the presence of SuperScriptII Reverse Transcriptase (Invitrogen), [^{35}S]dCTP, oligo(dT)$_{12-18}$, and oligo(dT)$_{25}$. The signal intensities in each analysis were normalized against the housekeeping gene Gapd, and the average signal intensities were calculated from more than two independent experiments.

2.5. *Cell culture and recombinant adenoviruses*

Mouse pancreatic islet β-cell line MIN6 cells were cultured as described previously.[18] Mouse ES cell line R1 cells[19] were cultured on mitomycin C-treated embryonic fibroblasts in a gelatinized dish with complete ES medium: high-glucose Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 20% fetal calf serum (FCS) (Invitrogen), 2 mM L-glutamine (Invitrogen), 1×nonessential amino acids (Invitrogen), 1 μM 2-mercaptoethanol (Invitrogen), 100 mg/ml streptomycin sulfate, 60.5 mg/l penicillin G (Invitrogen), and 1000 U/ml leukemia inhibitory factor (Chemicon, Temecula, CA). Mouse pituitary cell line AtT-20 cells[20] were grown in Ham's F10 medium (Invitrogen) with 5% FBS (Sigma, St. Louis, MO), and 15% heat-inactivated horse serum (Invitrogen).

Neurogenic differentiation 1 (NeuroD1/BETA2) recombinant adenovirus (Ad.CMVNeuroD1) and LacZ recombinant adenovirus (Ad.CMVLacZ) were constructed to express human NeuroD1 mRNA or *E. coli* LacZ (β-galactosidase) mRNA under cytomegalovirus (CMV) promoter, using an adenovirus expression vector kit (Takara Bio Inc., Otsu, Japan). MIN6 cells were infected with Ad.CMVNeuroD1 or Ad.CMVLacZ at a range of 5, 10, 30 and 50 multiplicity of infections (MOI), and cultured for another 2 days. Expression of NeuroD1 was confirmed by immunoblot analysis using antibodies specific for NeuroD1/BETA2 (Santa Cruz Biotechnology, Santa Cruz, CA) (Ishizuka et al. unpublished data), and expression of LacZ was confirmed using a galactosidase reporter gene staining detection kit (Sigma).

3. Results and Discussion

3.1. *Construction of a mouse pancreatic islet cDNA library based on a recombination-based method*

To construct a high quality cDNA library specific for normal mouse pancreatic islets, we applied a recombination-based method for library construction. Using 2 μg of poly(A)$^+$ RNA extracted from approximately 5000 mouse pancreatic islets, we successfully constructed a cDNA library having a complexity of more than 3.12×10^7 without amplification (Table 1). The library consists of six size fractions, which minimizes the size-bias effect on the population of cDNAs and results in improved variety.[14] The library is also characterized by an abundance of full-length cDNA clones and fewer chimeric clones. Since each cDNA insert has specific sequences for recombination (compatible with Invitrogen GATEWAY^{TM} system) at both ends, the library can readily be converted into other libraries such as mammalian expression and yeast two-hybrid libraries. This is the first pancreatic islet cDNA library created based on a recombination-based method, and should be a useful tool for pancreatic islet studies.

3.2. *Construction of a cDNA sequence database specific for mouse pancreatic islets*

We then sequenced a portion of the cDNA library to construct a sequence database specific for mouse pancreatic islets. We collected 4608 plasmid clones from each fraction, resulting in a total of 27,648 clones, and successfully sequenced 21,018 clones from the 3′ region of the cDNAs. This raw sequence data contained about 13.7% of insulin 1 or 2 (Ins1/2) sequences and only 0.03% amylase sequences (representative gene of exocrine pancreas), indicating that the library was constructed primarily from pancreatic islets.

After eliminating redundant clones, we obtained a total of 8882 non-redundant clones (Supplemental Information 2, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information2.html). Among them, we also sequenced 3264 clones from the 5′ region of the cDNAs, and detected only 114 chimeric clones by comparison with the Ensembl mouse genome database (www.ensemble.org). By analyzing 150 randomly selected clones, we found that about 60% of the clones contained full open reading frames (ORFs) (Table 1). We collected these results and built a mouse pancreatic islet sequence database.

We compared 8882 non-redundant sequence data with the "all genes" database (www.allgenes.org), and found that 261 had no match in the database. Of 8621 sequences, 5799 were classified into specific functional categories using a classification system designed by the Gene Ontology (GO) Consortium (www.geneontology.org) (Fig. 1, Table 2 and Supplemental Information 3, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information3.html), while 2822 have no GO annotation, suggesting that these were derived from non-coding RNAs or rare transcripts. Of the 5799 sequences having GO annotation, about 42% belong to the largest category "binding," which includes nucleic acid binding and peptide/protein binding. The second category "catalytic activity" comprises about 25%, and the third, "molecular function un-

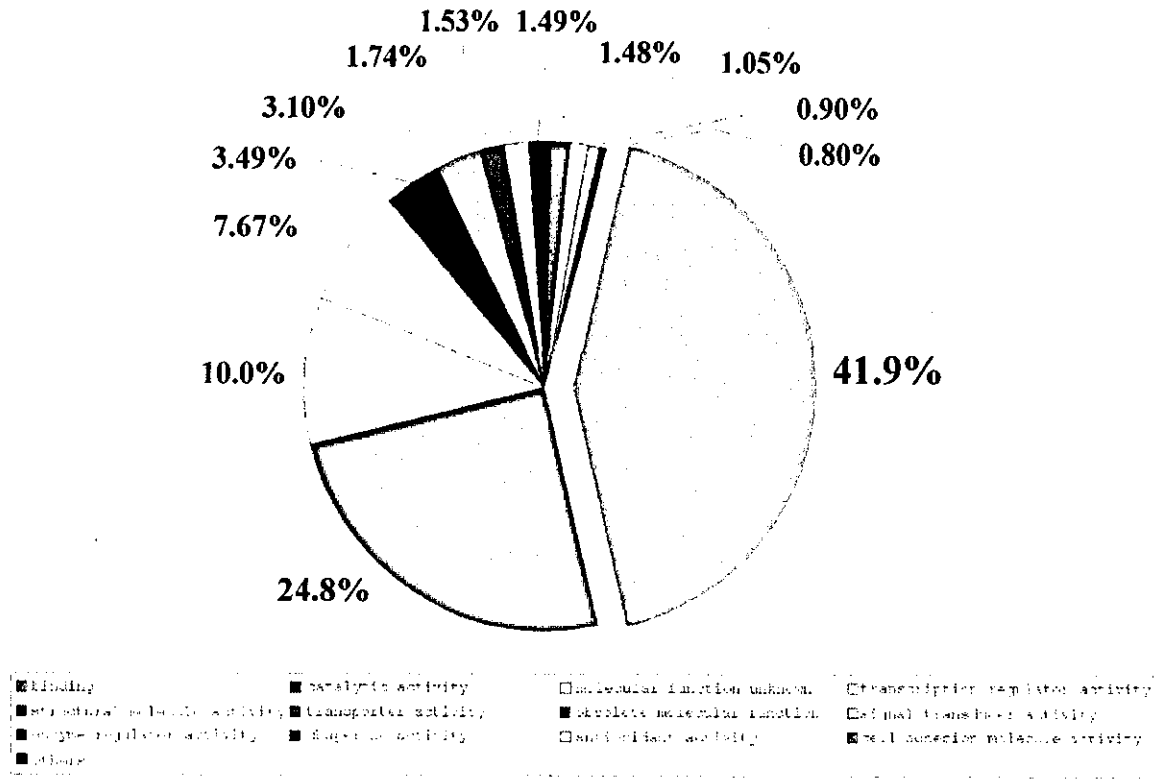

Figure 1. Functional classification of the mouse pancreatic islet cDNA clones. Out of 8882 non-redundant clones, 5799 were classified into specific functional categories using a classification system designed by the Gene Ontology Consortium. Representative genes for each functional category are listed in Table 2. A complete list is available in Supplemental Information 3, http://www.dna-res.kazusa.or.jp/11/5/03/supplemental/information3.html.

known," about 10%.

Although the distributions of GO annotations for mRNAs expressed in mouse whole pancreas (including exocrine and endocrine glands) have previously been reported,[21,22] those for pancreatic islets have not been available. Our present data are the first to show the distributions of GO annotations for mRNAs from mouse pancreatic islets. Possibly due to differences in materials or the method of library construction, our data exhibit some differences with previous reports.[21,22] For example, we found that "antioxidant" genes showed 1.1% distribution rather than 0.2%,[21] and "signal transducer" genes showed 1.6% rather than 10%.[22] In contrast, we detected "catalytic activity" genes at 24.8% distribution, which was similar to their "enzyme" classifications of 26.9%[21] and 31%.[22] We found 10% "molecular function unknown" genes, while they report only 3%.[22] Thus, gene expression profiles of pancreatic islets could provide valuable information on the molecules involved in normal function of the endocrine pancreas.

3.3. Application to microarray analyses

We then applied our cDNA library to microarray analyses. We selected a total of 8108 cDNA clones from 8882 non-redundant clones and spotted them on microarray membranes with 12 control clones (Supplemental Information 2, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information2.html). To estimate the reproducibility of our microarray analyses, we compared membranes of different lots hybridized with the same probe. Scattered plot analysis showed very similar signal intensities from the two different membranes, indicating that our microarray analyses are reproducible (Supplemental Information 4A, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information4.html).

To evaluate the usefulness and applicability of our microarray membranes, we performed the following two experiments. First, we compared the expression profiles of three cell lines: MIN6 cells (mouse insulin-secreting pancreatic islet β-cell line), mouse embryonic stem (ES) cells, and AtT-20 cells (mouse pituitary endocrine cell line), representing endocrine pancreatic islet cells, undifferentiated cells,

Table 2. A list of functional classifications of the mouse pancreatic islet cDNA clones.

GO function group	(subgroup)	annotation	GENBANK accession No.(gb)	gene index (gi)	unigene cluster (ug)
Binding	Nucleic acid binding	Mm#S2205667 Mus musculus, H2B histone family, member A, clone MGC:19269 IMAGE:3988862, mRNA, complete cds	gb=BC011440	gi=15030325	ug=Mm.21579
		Mm#S937714 Mus musculus H3 histone, family 3A (H3f3a), mRNA	gb=NM_008210	gi=6680158	ug=Mm.89136
		Mm#S937247 Mus musculus poly A binding protein, cytoplasmic 1 (Pabpc1), mRNA	gb=NM_008774	gi=6679196	ug=Mm.2642
		Mm#S2608854 Mus musculus dicer-like protein (Dicer1) mRNA, complete cds	gb=AF430845	gi=20385912	ug=Mm.31523
		Mm#S978931 Mus musculus polyrC) binding protein 2 (Pcbp2), mRNA	gb=NM_011042	gi=6997238	ug=Mm.111
	Peptide/Protein binding	Mm#S1985421 Mus musculus Ran binding protein 5 mRNA, partial cds	gb=AF294327	gi=12034715	ug=Mm.151329
		Mm#S2204674 Mus musculus, TAF binding protein, clone MGC:13789 IMAGE:4223922, mRNA, complete cds	gb=BC015074	gi=15929261	ug=Mm.14097
		Mm#S939589 Mus musculus syntaxin binding protein 2 (Stxbp2), mRNA	gb=NM_011503	gi=6755687	ug=Mm.7247
		Mm#S937219 Mus musculus nuclear receptor-binding SET-domain protein 1 (Nsd1), mRNA	gb=NM_008739	gi=6679137	ug=Mm.12964
		Mm#S1660484 Mus musculus piccolo (presynaptic cytomatrix protein) (Pclo), mRNA	gb=NM_011005	gi=15277339	ug=Mm.40886
	Others binding	Mm#S939404 Mus musculus retinol binding protein 1, cellular (Rbp1), mRNA	gb=NM_011254	gi=6755299	ug=Mm.2450
	(lipid, carbohydrate, etc.)	Mm#S938340 Mus musculus apolipoprotein E (Apoe), mRNA	gb=NM_009696	gi=6753101	ug=Mm.138866
		Mm#S937544 Mus musculus FK506 binding protein 2 (13 kDa) (Fkbp2), mRNA	gb=NM_008020	gi=6679804	ug=Mm.4234
		Mm#S979142 Mus musculus calcium binding protein 1 (Cabp1), mRNA	gb=NM_013879	gi=7304938	ug=Mm.38551
		Mm#S1660122 Mus musculus lectin, galactose binding, soluble 12 (Lgals12), mRNA	gb=NM_019516	gi=9506756	ug=Mm.87546
Catalytic activity	hydrolase activity	Mm#S937421 Mus musculus protein tyrosine phosphatase, non-receptor type 2 (Ptpn2), mRNA	gb=NM_008977	gi=6679552	ug=Mm.985
		Mm#S937616 Mus musculus acid beta glucosidase (Gba), mRNA	gb=NM_008094	gi=6679954	ug=Mm.5631
		Mm#S979159 Mus musculus carboxypeptidase E (Cpe), mRNA	gb=NM_013491	gi=7304972	ug=Mm.31395
		Mm#S979244 Mus musculus proprotein convertase subtilisin kexin type 1 (Pcsk1), mRNA	gb=NM_013628	gi=7305370	ug=Mm.1333
		Mm#S937263 Mus musculus proprotein convertase subtilisin kexin type 2 (Pcsk2), mRNA	gb=NM_008792	gi=6679228	ug=Mm.1247
	ligase activity	Mm#S1659886 Mus musculus cytidine 5'-triphosphate synthase 2 (Ctps2), mRNA	gb=NM_018737	gi=9055197	ug=Mm.2065
		Mm#S2003048 Mus musculus propionyl Coenzyme A carboxylase, beta polypeptide (Pccb), mRNA	gb=NM_025835	gi=13385300	ug=Mm.21079
		Mm#S2003116 Mus musculus phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole, succinocarboxamide synthetase (Paics), mRNA	gb=NM_025939	gi=13385433	ug=Mm.182931
		Mm#S253-4128 Mus musculus, glutamate-cysteine ligase, catalytic subunit, clone MGC:30487 IMAGE:4195425, mRNA, complete cds	gb=BC019374	gi=18043914	ug=Mm.4368
		Mm#S2581407 Mus musculus, similar to alanyl-tRNA synthetase (H. sapiens), clone MGC:37368 IMAGE:4976984, mRNA, complete cds	gb=BC026611	gi=20072364	ug=Mm.24174
	oxidoreductase activity	Mm#S1660039 Mus musculus NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1 (Ndufa1), mRNA	gb=NM_019443	gi=9506910	ug=Mm.34869
		Mm#S1660493 Mus musculus pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde synthetase) (Pycs), mRNA	gb=NM_019698	gi=9790060	ug=Mm.29751
		Mm#S2002811 Mus musculus sterol-C4-methyl oxidase-like (Sc4mol), mRNA	gb=NM_025436	gi=13384935	ug=Mm.30119
		Mm#S2455829 Mus musculus cytochrome c oxidase, subunit VIc (Cox6c), mRNA	gb=NM_053071	gi=16716342	ug=Mm.548
		Mm#S936925 Mus musculus stearoyl-Coenzyme A desaturase 2 (Scd2), mRNA	gb=NM_009128	gi=6677862	ug=Mm.193066
	transferase activity	Mm#S1437353 Mus musculus histone acetyltransferase (Mort-pending), mRNA	gb=NM_017479	gi=8567375	ug=Mm.30998
		Mm#S1660198 Mus musculus mixa methyltransferase (Mtxa-pending), mRNA	gb=NM_019721	gi=9790096	ug=Mm.76983
		Mm#S1985402 Mus musculus prenylcysteine carboxylmethyltransferase mRNA, partial cds	gb=AF209926	gi=12082482	ug=Mm.29356
		Mm#S2455982 Mus musculus, transglutaminase 2, C polypeptide, clone MGC:6152 IMAGE:3256943, mRNA, complete cds	gb=BC016492	gi=16741319	ug=Mm.18843
		Mm#S2003650 Mus musculus glycogen synthase 1, muscle (Gys1), mRNA	gb=NM_030678	gi=13507598	ug=Mm.185247
molecular function unknown		Mm#S1971846 Mus musculus adult male hippocampus cDNA, RIKEN full-length enriched library, clone:2900059A22:DiGeorge syndromic chromosome region 6, full insert sequence	gb=AK019346	gi=12859501	ug=Mm.27155
		Mm#S1973121 Mus musculus 8 days embryo whole body cDNA, RIKEN full-length enriched library, clone:5730433K22:unclassifiable, full insert sequence	gb=AK017611	gi=12856941	ug=Mm.156400

Table 2. Continued.

GO function (subgroup) group	annotation	GENBANK accession No (gb)	gene index (gi)	unigene cluster (ug)
	Mm#S2611167 Mus musculus RIKEN cDNA 2900316D05 gene (2900316D05Rik), mRNA	gb=NM_028381	gi=21312170	ug=Mm.22796
	Mm#S2533711 Mus musculus, Similar to hypothetical protein FLJ10276, clone IMAGE:5344303, mRNA, partial cds	gb=BC021480	gi=18204810	ug=Mm.17918
	Mm#S2049845 Mus musculus, Similar to hypothetical protein, clone MGC:7703 IMAGE:3497634, mRNA, complete cds	gb=BC008544	gi=14250242	ug=Mm.58660
transcription regulator activity	Mm#S937155 Mus musculus myelin transcription factor 1 (Myt1), mRNA	gb=NM_008665	gi=6678097	ug=Mm.20088
	Mm#S2635265 Mus musculus NK6 transcription factor related, locus 1 (Drosophila) (Nkx6-1), mRNA	gb=NM_144955	gi=21450628	ug=Mm.193072
	Mm#S979343 Mus musculus paired box gene 6 (Pax6), mRNA	gb=NM_013627	gi=7305368	ug=Mm.3608
	Mm#S2050432 Mus musculus sterol regulatory element binding protein 1 (Srebp1) mRNA, partial cds	gb=AF374266	gi=14161490	ug=Mm.214958
	Mm#S937735 Mus musculus hairy and enhancer of split 1, (Drosophila) (Hes1), mRNA	gb=NM_008235	gi=6680204	ug=Mm.4451
structural molecule activity	Mm#S121804 Tubulin, alpha 2, mRNA	gb=M13446	gi=202209	ug=Mm.197515
	Mm#S2577772 Mus musculus dystonin (Dst), mRNA	gb=NM_134448	gi=19882220	ug=Mm.25326
	Mm#S937847 Mus musculus integral membrane glycoprotein (Itmg) mRNA	gb=NM_008377	gi=6680444	ug=Mm.944
	Mm#S938318 Mus musculus actin, gamma, cytoplasmic (Actg), mRNA	gb=NM_009609	gi=6752953	ug=Mm.196173
	Mm#S979073 Mus musculus ribosomal protein L13a (Rpl13a), mRNA	gb=NM_009438	gi=7110730	ug=Mm.13020
transporter activity	Mm#S979102 Mus musculus ATP-binding cassette, sub-family A (ABC1), member 1 (Abca1), mRNA	gb=NM_013454	gi=7304848	ug=Mm.369
	Mm#S979103 Mus musculus ATP-binding cassette, sub-family C (CFTR/MRP), member 5a (Abcc5a), mRNA	gb=NM_013790	gi=7304850	ug=Mm.20845
	Mm#S1659675 Mus musculus solute carrier family 1 (glutamate neutral amino acid transporter), member 4 (Slc1a4), mRNA	gb=NM_018861	gi=9256943	ug=Mm.6379
	Mm#S2037281 Mus musculus solute carrier family 2 (facilitated glucose transporter), member 2 (Slc2a2), mRNA	gb=NM_031197	gi=13654261	ug=Mm.18443
	Mm#S1660140 Mus musculus solute carrier family 2 (facilitated glucose transporter), member 5 (Slc2a5), mRNA	gb=NM_019741	gi=9789966	ug=Mm.34150
obsolete molecular function	Mm#S1660236 Mus musculus SEC23B (S. cerevisiae) (Sec23b), mRNA	gb=NM_019787	gi=9790212	ug=Mm.28784
	Mm#S1437100 Mus musculus erythrocyte protein band 7.2 (Epb7.2), mRNA	gb=NM_013515	gi=7710017	ug=Mm.4441
	Mm#S936925 Mus musculus unc5 homolog (C. elegans) 3 (Unc5h3), mRNA	gb=NM_009472	gi=6678544	ug=Mm.24430
	Mm#S937693 Mus musculus mesoderm specific transcript (Mest), mRNA	gb=NM_008590	gi=6678865	ug=Mm.1089
	Mm#S93857 Mus musculus COP9 (constitutive photomorphogenic) homolog, subunit 4 (Arabidopsis thaliana) (Cops4), mRNA	gb=NM_012001	gi=6753489	ug=Mm.957
signal transducer activity	Mm#S939854 Mus musculus growth hormone receptor (Ghr), mRNA	gb=NM_010284	gi=6857794	ug=Mm.3986
	Mm#S1659947 Mus musculus G protein-coupled receptor 56 (Gpr56), mRNA	gb=NM_018882	gi=9256530	ug=Mm.13799
	Mm#S2050287 Mus musculus, vitamin D receptor, clone MGC:12147 IMAGE:3710866, mRNA, complete cds	gb=BC006716	gi=13879474	ug=Mm.44170
	Mm#S2000654 Mus musculus notch 1 protein mRNA, complete cds	gb=AF508809	gi=20805940	ug=Mm.31255
	Mm#S979471 Mus musculus prostaglandin E receptor 1 (subtype EP1), 42kD (Ptger1), mRNA	gb=NM_013641	gi=7363446	ug=Mm.213000
enzyme regulator activity	Mm#S939211 Mus musculus neutral sphingomyelinase (N-SMase) activation associated factor (Nsmaf), mRNA	gb=NM_010945	gi=6754807	ug=Mm.3056
	Mm#S2581974 Mus musculus guanine nucleotide exchange factor (Lanp) mRNA, complete cds	gb=AF467766	gi=19387125	ug=Mm.103659
	Mm#S939644 Mus musculus tissue inhibitor of metalloproteinase 3 (Timp3), mRNA	gb=NM_011595	gi=6755792	ug=Mm.4871
	Mm#S937702 Mus musculus Rho guanine nucleotide exchange factor (GEF) 1 (Arhgef1), mRNA	gb=NM_008488	gi=6678967	ug=Mm.3181
	Mm#S939368 Mus musculus RAS protein-specific guanine nucleotide-releasing factor 1 (Rasgrf1), mRNA	gb=NM_011245	gi=6755287	ug=Mm.29921
chaperone activity	Mm#S1097753 Mus musculus heat shock 70kD protein 5 (glucose-regulated protein, 78kD) (Hspa5), mRNA	gb=NM_022310	gi=11612488	ug=Mm.918
	Mm#S978930 Mus musculus osmotic stress protein 94 kDa (Osp94), mRNA	gb=NM_011620	gi=6997236	ug=Mm.4150
	Mm#S2017376 Mus musculus Dnaj (Hsp40) homolog, subfamily A, member 3 (Dnaja3), mRNA	gb=NM_023646	gi=13994154	ug=Mm.28330
	Mm#S939134 Mus musculus Dnaj (Hsp40) homolog, subfamily B, member 6 (Dnajb6), mRNA	gb=NM_011847	gi=6754735	ug=Mm.2761
	Mm#S1660317 Mus musculus prefoldin 5 (Pfdn5), mRNA	gb=NM_020031	gi=9910215	ug=Mm.181847

Table 2. Continued.

GO function (subgroup) group	annotation	GENBANK accession No (gb)	gene index (gi)	unigene cluster (ug)
antioxidant activity	Mm#S937966 Mus musculus peroxiredoxin 3 (Prdx3), mRNA	gb=NM_007452	gi=6680689	ug=Mm.29821
	Mm#S1437442 Mus musculus peroxiredoxin 4 (Prdx4), mRNA	gb=NM_016764	gi=7948998	ug=Mm.19127
	Mm#S939699 Mus musculus thioredoxin 1 (Txn1), mRNA	gb=NM_011660	gi=6755910	ug=Mm.1275
	Mm#S1660289 Mus musculus thioredoxin 2 (Txn2), mRNA	gb=NM_019913	gi=9903608	ug=Mm.3533
	Mm#S2455825 Mus musculus glutaredoxin 1 (thioltransferase) (Glrx1), mRNA	gb=NM_053108	gi=16716404	ug=Mm.29728
cell adhesion molecule activity	Mm#S979051 Mus musculus integrin alpha 6 (Itga6), mRNA	gb=NM_008397	gi=7110658	ug=Mm.25232
	Mm#S4156 Murine mRNA for integrin beta subunit, mRNA	gb=Y00769	gi=52721	ug=Mm.4712
	Mm#S3085 Procollagen, type I, alpha 1, mRNA	gb=U08020	gi=470673	ug=Mm.22621
	Mm#S2080 Procollagen, type IV, alpha 1, mRNA	gb=J04694	gi=556296	ug=Mm.738
	Mm#S244281 Procollagen, type XV, mRNA	gb=AF011450	gi=2558824	ug=Mm.4352

Out of 8882 non-redundant clones, 5799 were classified into specific functional categories using a classification system designed by the Gene Ontology Consortium. Representative genes for each functional category are listed in this table. A complete list is available in Supplemental Information 3, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information3.html.

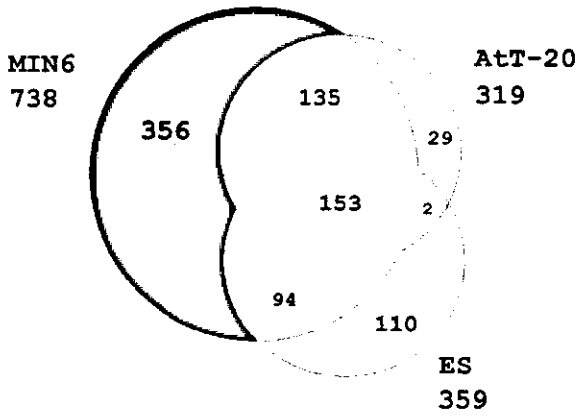

Figure 2. Comparison of expression profiles in MIN6, ES and AtT-20 cells. Genes with signal intensity of one-tenth or more of the housekeeping gene Gapd were compared in the three cell lines. Numbers of genes are shown in this diagram. A complete list of the expressed genes is available in Supplemental Information 5, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information5.html.

and other endocrine cells, respectively (Supplemental Information 4B and 4C, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information4.html). We used a threshold signal intensity of one-tenth that of the housekeeping gene Gapd, and compared the expressed genes in the three cell lines (Fig. 2). A list of the expressed genes is available (Supplemental Information 5, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information5.html). Among a total of 879 genes, 738 (84.0%) were expressed in MIN6 cells, while only 319 (36.3%) and 359 (40.8%) were expressed in AtT-20 and ES cells, respectively.

Comparison of expression profiles of three cell lines revealed both differentially and similarly expressed genes. A subset of 135 genes was expressed in both MIN6 and AtT-20 cells but not in ES cells. As both MIN6 and AtT-20 cells were derived from endocrine cells and preserve regulated exocytotic ability, the genes involved should be expressed in both endocrine cell types. Indeed, well-known endocrine-specific genes such as chromogranins,[23,24] prohormone convertase 1/3,[25] and carboxypeptidase E[26,27] were in this subset (Fig. 2 and Supplemental Information 5, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information5.html). In addition, 28 genes with unknown functions were contained in this subset, suggesting they may have other roles in regulated exocytosis.

By contrast, a subset of 356 genes was found to be expressed only in MIN6 cells. This subset included known pancreatic β-cell-specific genes, such as Ins1/2, Iapp, and Pdx1 as well as many uncharacterized genes (Supplemental Information 5, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information5.html) that may be involved in phenomena specific to pancreatic β-cells such as insulin synthesis. Thus, comparison of the mRNA expression patterns of various cell types should be useful in investigating the biological function of the clones newly identified in our cDNA library.

As the second experiment, we compared expression profiles between MIN6 cells and those infected with aden-

Table 3. A list of genes upregulated by overexpression of NeuroD1 in MIN6 cells.

No.	Signal intensity +LacZ[1)]	+NeuroD1[2)]	Ratio	Gene name[3)] / GenBank accession number
1	4.70	84.31	17.9324	gi: 26101046 / AK083076
2	14.59	251.26	17.2207	Neurod1 (NeuroD1)[4)] / NM_010894
3	3.60	24.37	6.7660	gi: 38328296 / BC062185
4	1.61	7.60	4.7122	Scnn1b / NM_011325
5	4.26	18.02	4.2319	Ipf1 (Pdx1) / NM_008814
6	16.42	62.79	3.8232	Cdkn1a (p21) / NM_007669
7	2.07	7.51	3.6218	Jund1 / NM_010592
8	7.48	24.16	3.2285	Ddx5 / BC009142
9	5.25	15.33	2.9198	Rangap1 / NM_011241
10	4.98	14.42	2.8952	gi: 12833685 / AK003180
11	4.47	12.80	2.8639	Atp2a2 / NM_009722
12	2.65	7.53	2.8440	Scly / NM_016717
13	7.50	19.71	2.6274	gi: 17391148 / BC018486
14	6.66	17.30	2.5976	Actn4 / NM_021895

Signal intensities are normalized by two independent experiments.
1) Signal intensities from MIN6 cells infected with Ad.CMVLacZ.
2) Signal intensities from MIN6 cells infected with Ad.CMVNeuroD1.
3) gi (gene index number in GenBank) or Gene name registered in NCBI Entrez Gene database.
4) Signal intensity of NeuroD1 from MIN6 cells infected with Ad.CMVNeuroD1 was influenced by the exogenous introduction of NeuroD1.

oviral vector to express the transcription factor NeuroD1 exogenously. Because NeuroD1 is critical in pancreatic islet β-cell development[6] and β-cell functions,[28] and a mutation in the human NeuroD1 gene is responsible for maturity-onset diabetes of the young (MODY6),[11,12] we overexpressed NeuroD1 in MIN6 cells to evaluate the usefulness of our microarray analyses in the search for novel target genes. So far, Pdx1,[29] p21,[30] and Sur1[31] have been reported to be transactivated by NeuroD1.

Infection of the adenoviral vector with a range of 5 or 10 MOI did not alter the expression profiles, but infection with 30 or 50 MOI caused significant alterations (data not shown). The results of infection with 50 MOI are shown in Table 3 and Supplemental Information 4D, http://www.dna-res.kazusa.or.jp/11/5/01/supplemental/information4.html. We confirmed that the genes known to be induced by NeuroD1, including Pdx1[29] and p21,[30] were upregulated, and also found that expression levels of another 11 genes increased significantly. None of these genes has previously been associated with NeuroD1, and all are candidates for transactivation by NeuroD1. This second experiment shows that our microarray analyses will be useful in the search for novel molecular targets of various β-cell transcription factors and in investigations of pancreatic islets and related diseases including diabetes.

Acknowledgements: We thank Y. Murahashi for assistance with computational analysis of nucleotide sequences. We also thank T. Watanabe and T. Hirano for technical assistance in constructing the mouse pancreatic islet library and in the microarray analysis. This work was supported by Grants-in-Aid for Specially Promoted Research and for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan; by a Grant from the Ministry of Economy, Trade and Industry, Japan, and the Futaba Electronics Memorial Foundation.

References

1. Kahn, C. R., Gordon, C., Weir, G. C. et al. 1994, *Joslin's Diabetes Mellitus, 13th Ed.*, Lippincott Williams & Wilkins, Philadelphia, PA.
2. Jonsson, J., Carlsson, L., Edlund, T., Edlund, H. 1994, Insulin-promoter-factor 1 is required for pancreas development in mice, *Nature*, **13**, 606–609.
3. Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K., and Edlund, H. 1998, Beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes, *Genes Dev.*, **12**, 1763–1768.
4. Lee, Y. H., Sauer, B., and Gonzalez, F. J. 1998, Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse, *Mol Cell Biol.*, **18**, 3059–3068.

5. Shih, D. Q., Screenan, S., Munoz, K. N. et al. 2001, Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism, *Diabetes*, **50**, 2472–2480.
6. Naya, F. J., Huang, H. P., Qiu, Y. et al. 1997, Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice, *Genes Dev.*, **11**, 2323–2334.
7. Stoffers, D. A., Ferrer, J., Clarke, W. L., and Habener J. F. 1997, Early-onset type-II diabetes mellitus (MODY4) linked to IPF1, *Nat Genet*, **17**, 138–139.
8. Yamagata, K., Oda, N., Kaisaki, P. J. et al. 1996, Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3), *Nature*, **384**, 455–458.
9. Yamagata, K., Furuta, H., Oda, N. et al. 1996, Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1), *Nature*, **384**, 458–460.
10. Horikawa, Y., Iwasaki, N., Hara, M. et al. 1997, Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY, *Nat Genet*, **17**, 384–385.
11. Malecki, M. T., Jhala, U. S., Antonellis, A. et al. 1999, Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus, *Nat Genet*, **23**, 323–328.
12. Bell G. I. and Polonsky, K. S. 2001, Diabetes mellitus and genetically programmed defects in beta-cell function, *Nature*, **414**, 788–791.
13. Ohara, O. and Temple, G. 2001, Directional cDNA library construction assisted by the in vitro recombination reaction, *Nucleic Acids Res.*, **29**, E22.
14. Ohara, O., Nagase, T., Mitsui, G. et al. 2002, Characterization of size-fractionated cDNA libraries generated by the in vitro recombination-assisted method, *DNA Res.*, **9**, 47–57.
15. Nakayama, M., Kikuno, R., and Ohara, O. 2002, Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs, *Genome Res.*, **12**, 1773–1784.
16. Miki, T., Nagashima, K., Tashiro, F. et al. 1998, Defective insulin secretion and enhanced insulin action in KATP channel-deficient mice, *Proc. Natl. Acad. Sci. USA*, **95**, 10402–10406.
17. Mitsui, G., Mitsui, K., Hirano, T., Ohara, O., Kato M., and Niwano, Y. 2003, Kinetic profiles of sequential gene expressions for chemokines in mice with contact hypersensitivity, *Immunol Lett.*, **86**, 191–197.
18. Kawaki, J., Nagashima, K., Tanaka, J. et al. 1999, Unresponsiveness to glibenclamide during chronic treatment induced by reduction of ATP-sensitive K^+ channel activity, *Diabetes*, **48**, 2001–2006.
19. Suzuki, M., Mizutani-Koseki, Y., Fujimura, Y. et al. 2002, Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice. *Development*, **129**, 4171–4183.
20. Buonassisi, V., Sato, G., and Cohen, AI. 1962, Hormone-producing cultures of adrenal and pituitary tumor origin, *Proc. Natl. Acad. Sci. USA*, **48**, 1184–1190.
21. Scearce, L. M., Brestelli, J. E., McWeeney, S. K. et al. 2002, Functional genomics of the endocrine pancreas: the pancreas clone set and PancChip, new resources for diabetes research, *Diabetes*, **51**, 1997–2004.
22. Kaestner, K. H., Lee, C. S., Scearce, L. M. et al. 2003, Transcriptional program of the endocrine pancreas in mice and humans, *Diabetes*, **52**, 1604–1610.
23. Ehrhart, M., Grube, D., Bader, M. F., Aunis, D., and Gratzl, M. 1986, Chromogranin A in the pancreatic islet: cellular and subcellular distribution, *J. Histochem Cytochem.*, **34**, 1673–1682.
24. Iacangelo, A., Affolter, H. U., Eiden, L. E., Herbert, E., and Grimes, M. 1986, Bovine chromogranin A sequence and distribution of its messenger RNA in endocrine tissues, *Nature*, **323**, 82–86.
25. Zhu, X., Zhou, A., Dey, A. et al. 2002, Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects, *Proc. Natl. Acad. Sci. USA*, **99**, 10293–10298.
26. Fricker, L. D., Adelman, J. P., Douglass, J., Thompson, R. C., von Strandmann, RP., and Hutton, J. 1989, Isolation and sequence analysis of cDNA for rat carboxypeptidase E [EC 3.4.17.10], a neuropeptide processing enzyme, *Mol. Endocrinol.*, **3**, 666–673.
27. Naggert, J. K., Fricker, L. D., Varlamov, O. et al. 1995, Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity, *Nat. Genet*, **10**, 135–142.
28. Huang, H. P, Chu, K, Nemoz-Gaillard, E, Elberg, D, and Tsai, M. J. 2002, Neogenesis of beta-cells in adult BETA2/NeuroD-deficient mice, *Mol Endocrinol.*, **16**, 541–551.
29. Kojima, H., Fujimiya, M., Matsumura, K. et al. 2003, NeuroD-betacellulin gene therapy induces islet neogenesis in the liver and reverses diabetes in mice, *Nat. Med.*, **9**, 596–603.
30. Mutoh, H., Naya, F. J., Tsai, M. J., and Leiter, A. B. 1998, The basic helix-loop-helix protein BETA2 interacts with p300 to coordinate differentiation of secretin-expressing enteroendocrine cells, *Genes Dev.*, **12**, 820–830.
31. Kim J. W., Seghers V., Cho J. H., et al. 2002, Transactivation of the mouse sulfonylurea receptor I gene by BETA2/NeuroD, *Mol. Endocrinol.*, **16**, 1097–1107.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 49, Issue of December 3, pp. 51376–51385, 2004
Printed in U.S.A.

Physical and Functional Interaction between Dorfin and Valosin-containing Protein That Are Colocalized in Ubiquitylated Inclusions in Neurodegenerative Disorders*

Received for publication, June 15, 2004, and in revised form, August 31, 2004
Published, JBC Papers in Press, September 29, 2004, DOI 10.1074/jbc.M406683200

Shinsuke Ishigaki‡§¶, Nozomi Hishikawa‡, Jun-ichi Niwa‡, Shun-ichiro Iemura‖, Tohru Natsume‖, Seiji Hori, Akira Kakizuka**‡‡, Keiji Tanaka§, and Gen Sobue‡§§**

*From the ‡Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya 466-8500, Japan, the §Department of Molecular Oncology, Tokyo Metropolitan Institute of Medical Science, Tokyo 113-8613, Japan, the ‖National Institute of Advanced Science and Technology, Biological Information Research Center, Tokyo 135-0064, Japan, the **Laboratory of Functional Biology, Kyoto University Graduate School of Biostudies, Kyoto 606-8502, Japan, and ‡‡CREST, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan*

Dorfin, a RING-IBR type ubiquitin ligase (E3), can ubiquitylate mutant superoxide dismutase 1, the causative gene of familial amyotrophic lateral sclerosis (ALS). Dorfin is located in ubiquitylated inclusions (UBIs) in various neurodegenerative disorders, such as ALS and Parkinson's disease (PD). Here we report that Valosin-containing protein (VCP) directly binds to Dorfin and that VCP ATPase activity profoundly contributes to the E3 activity of Dorfin. High through-put analysis using mass spectrometry identified VCP as a candidate of Dorfin-associated protein. Glycerol gradient centrifugation analysis showed that endogenous Dorfin consisted of a 400–600-kDa complex and was co-immunoprecipitated with endogenous VCP. *In vitro* experiments showed that Dorfin interacted directly with VCP through its C-terminal region. These two proteins were colocalized in aggresomes in HEK293 cells and UBIs in the affected neurons of ALS and PD. VCP^{K524A}, a dominant negative form of VCP, reduced the E3 activity of Dorfin against mutant superoxide dismutase 1, whereas it had no effect on the autoubiquitylation of Parkin. Our results indicate that VCPs functionally regulate Dorfin through direct interaction and that their functional interplay may be related to the process of UBI formation in neurodegenerative disorders, such as ALS or PD.

Amyotrophic lateral sclerosis (ALS)[1] is one of the most common neurodegenerative disorders, characterized by selective motor neuron degeneration in the spinal cord, brain stem, and cortex. Two genes, CuZn-superoxide dismutase (SOD1) and amyotrophic lateral sclerosis 2 have been identified as responsible genes for familial forms of ALS. Using mutant SOD1 transgenic mice, the pathogenesis of ALS has been partially uncovered. The proposed mechanisms of the motor neuron degeneration in ALS include oxidative toxicity, glutamate receptor abnormality, ubiquitin proteasome dysfunction, inflammatory and cytokine activation, dysfunction of neurotrophic factors, damage to mitochondria, cytoskeletal abnormalities, and activation of the apoptosis pathway (1, 2).

In a previous study (3), we identified several ALS-associated genes using molecular indexing. Dorfin was identified as one of the up-regulated genes in ALS, which contains a RING-IBR (in between ring finger) domain at its N terminus and mediated ubiquitin ligase (E3) activity (3, 4). Dorfin colocalized with Vimentin at the centrosome after treatment with a proteasome inhibitor in cultured cells (4). Dorfin physically bound and ubiquitylated various SOD1 mutants derived from familial ALS patients and enhanced their degradation, but it had no effect on the stability of wild-type SOD1 (5). Overexpression of Dorfin protected neural cells against the toxic effects of mutant SOD1 and reduced SOD1 inclusions (5).

Recent findings indicate that the ubiquitin-proteasome system is widely involved in the pathogenesis of Parkinson's disease (PD), Alzheimer's disease, polyglutamine disease, and Prion diseases as well as ALS (6). From this point of view, we previously analyzed the pathological features of Dorfin in various neurodegenerative diseases and found that Dorfin was predominantly localized not only in Lewy body (LB)-like inclusions in ALS but also in LBs in PD, dementia with Lewy bodies, and glial cell inclusions in multiple system atrophy (7). These characteristic intracellular inclusions composed of aggregated, ubiquitylated proteins surrounded by disorganized filaments are the histopathological hallmark of aging-related neurodegenerative diseases (8).

A structure called aggresome by Johnston *et al.* (9) is formed when the cell capacity to degrade misfolded proteins is exceeded. The aggresome has been defined as a pericentriolar, membrane-free, cytoplasmic inclusion containing misfolded ubiquitylated protein ensheathed in a cage of intermediate filaments, such as Vimentin (9). The formation of the aggresome mimics that of ubiquitylated inclusions (UBIs) in the affected neurons of various neurodegenerative diseases (10). Combined with the fact that Dorfin was localized in aggresomes in cultured cells and UBIs in ALS and other neurode-

* This work was supported by a grant for a Center of Excellence from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Research resident of the Japan Foundation for Aging and Health, Psychiatric and Neurological Diseases, and Mental Health.

§§ To whom correspondence should be addressed: Dept. of Neurology, Nagoya University Graduate School of Medicine, Nagoya 466-8500, Japan. Tel.: 81-52-744-2385; Fax: 81-52-744-2384; E-mail: sobueg@med.nagoya-u.ac.jp.

[1] The abbreviations used are: ALS, amyotrophic lateral sclerosis; E3, ubiquitin ligase; ERAD, endoplasmic reticulum-associated degradation; LB, Lewy body; MS, mass spectrometry; LC-MS/MS, liquid chromatography coupled to electrospray tandem mass spectrometry; PD, Parkinson's disease; SOD1, CuZn-superoxide dismutase; UBI, ubiquitylated inclusions; VCP, valosin-containing protein; FLAG-Parkin, pcDNA3.1/FLAG-Parkin; Ub, ubiquitin; MBP, maltose-binding protein; GST, glutathione *S*-transferase; PBS, phosphate-buffered saline; HA, hemagglutinin; WT, wild type.