

# Gene expression profiling in human insulinoma tissue: genes involved in the insulin secretion pathway and cloning of novel full-length cDNAs

X-C Wang, S-Y Xu<sup>1</sup>, X-Y Wu<sup>2</sup>, H-D Song<sup>1</sup>, Y-F Mao<sup>1</sup>, H-Y Fan<sup>2</sup>, F Yu<sup>1</sup>, B Mou<sup>1</sup>, Y-Y Gu<sup>1</sup>, L-Q Xu<sup>1</sup>, X-O Zhou<sup>1</sup>, Z Chen<sup>2</sup>, J-L Chen<sup>1</sup> and R-M Hu

State Key Laboratory of Genetics Engineering, Institute of Endocrinology and Diabetology, Department of Endocrinology, Huashan Hospital, Fu Dan University, 12 Middle Urumqi Road, Shanghai 200040, People's Republic of China

Shanghai Institute of <sup>1</sup>Endocrinology and <sup>2</sup>Hematology, Ruijin Hospital, Shanghai Second Medical University, 197 Ruijin Er Road, Shanghai 200025, People's Republic of China

(Requests for offprints should be addressed to R-M Hu; Email: renminghu@fudan.edu.cn)

## Abstract

Insulinoma is a clinically common cause of organic hypoglycemia. The prominent characteristic of insulinoma is endogenous hyperinsulinism. Until now, the molecular biology of human insulinoma has been little understood. In this study, gene expression profiling of human insulinoma was established by expressed sequence tag (EST) sequencing and cDNA array. A total of 2063 clones were obtained, of these, 1589 clones were derived from EST sequencing, 975 clones were derived from cDNA array and 501 clones were shared by the two methods. G protein  $\alpha$ -stimulating activity polypeptide (Gs $\alpha$ ) and carboxypeptidase E (CPE) were the most highly expressed genes in human insulinoma, as derived by EST sequencing and cDNA array respectively. The genes involved in the protein/insulin secretion pathway were strongly expressed in human insulinoma tissue. Meanwhile, eight full-length cDNAs of novel genes were cloned and sequenced. The results demonstrated the molecular biology of human insulinoma tissue at the level of transcript abundance and validated the efficacy of EST sequencing combined with cDNA array in the construction of gene expression profiling. In conclusion, the predominance of the genes participating in the secretory pathway suggested that regulation of secretion might be a major mechanism by which insulin release is abnormally increased in patients with insulinomas. It was also concluded that overexpression of the Gs $\alpha$  gene played an important role in the pathogenesis of insulinoma.

*Endocrine-Related Cancer* (2004) 11 295–303

## Introduction

The maintenance of plasma glucose concentration is critical to the survival of most organisms; this is due to plasma glucose being the predominant fuel utilized by organic cells under most conditions, especially cells of the nervous system. The  $\beta$ -cells of the pancreatic islets of Langerhans play a central role in glucose homeostasis by secreting the polypeptide hormone insulin. Under physiological conditions,  $\beta$ -cells sense plasma glucose concentration and maintain tight control over insulin biosynthesis and secretion. Functional disorders of  $\beta$ -cells could result in hyperglycemia or hypoglycemia.

Diabetes is characterized by hyperglycemia resulting from dysfunction of  $\beta$ -cells, and/or insensitivity to insulin, of peripheral tissues. Hypoglycemia can be the result of drugs, critical illness including hepatic or renal failure, hormonal deficiencies, endogenous hyperinsulinism (insulinoma), etc. (Larsen *et al.* 2002).

An insulinoma is the most common cause of hypoglycemia resulting from endogenous hyperinsulinism. The most consistent insulin secretory abnormalities associated with insulinoma are oversecretion of insulin and failure of the normal decrease in insulin secretion which occurs as plasma glucose levels decline in the postabsorptive state. Insulinomas are rare, the estimated

incidence being one case per 250 000 patient-years in the United States, but it is a curable cause of potentially lethal hypoglycemia. Insulinomas almost always come to clinical attention because of hypoglycemia (Larsen *et al.* 2002). Most insulinomas are derived from monoclonal proliferation of the  $\beta$ -cells of the pancreatic islets, which means insulinomas are an optimal model for studying  $\beta$ -cell biology. Insulinoma cells offer several advantages over cultured  $\beta$ -cell lines and isolated islet material: (1) studies of  $\beta$ -cell lines are conducted *in vitro*; (2) isolated islet material is a complex tissue composed of varied cell types. Insulinoma cells offer not only a pure  $\beta$ -cell population but also a study *in vivo*.

To illuminate the molecular biology of insulinoma/ $\beta$ -cells, gene expression profiling of human insulinoma tissue has been established using expressed sequence tag (EST) sequencing and cDNA array. As a result, the genes involved in the secretory pathway of protein/insulin are strongly expressed in insulinoma tissue and several new full-length cDNAs have been cloned.

## Materials and methods

### RNA extraction

Insulinoma and normal pancreatic tissues were obtained from patients with insulinomas and adult males who had died of accidents respectively. Total RNA was extracted from frozen tissues using TRIzol reagent (Invitrogen). mRNA was separated and purified using oligo(dT) (Qiagen).

### cDNA library construction

cDNA synthesis was performed with a CapFinder PCR cDNA library construction kit (Clontech).

### cDNA sequencing

Bacteria growth and plasmid extraction were performed in a 96-well format (Qiagen). Sequencing reactions were performed on a 9600 thermal reactor (Perkin-Elmer, Shelton, CT, USA) using a dye primer cycle sequencing kit (Perkin-Elmer), and partial cDNA sequencing of each clone was taken from the 5' end. The reaction products were analyzed using an ABI 3700 DNA sequencer (Perkin-Elmer).

### Bioinformatics analysis and data management

Sequencing data were transmitted to a Sun station (Sun Microsystems, Santa Clara, CA, USA). Analysis of sequences, including quality assessment and quality base trimming of ESTs, was performed with a GCG software

package (Genetic Computer Version 9.0). The good-quality sequences were referred to those longer than 100 bp and containing less than 3% ambiguous bases. After being compared with public databases such as GenBank, dbEST and Unigene by using Blast and Fasta software, analyzable sequences were categorized into: known genes that shared at least 95% homology with certain genes over at least 100 bp of DNA sequences; known ESTs that couldn't meet the criteria of known genes but shared at least 95% homology over at least 100 bp of ESTs in dbEST; novel ESTs that shared no homology or little homology with known genes and known ESTs. ESTs corresponding to known genes were divided into several categories according to the Gene Ontology Consortium functions (Ashburner *et al.* 2000) (<http://www.geneontology.org>). Clustering of the ESTs generated in the work was performed using CAT 3.2 from Pangea (Oakland, CA, USA) with default parameters.

### cDNA microarray

A total of 14 000 clones were spotted onto nylon membrane-based cDNA arrays by using Spotter (Bio-Robotics, Cambridge, UK); these clones were derived from human liver and hemopoietic stem cells and hypothalamus–pituitary–adrenal (HPA) libraries or purchased from Research Genetics (Huntsville, AL, USA). Approximately 1–2  $\mu$ g mRNA of human insulinoma tissue were labeled in a reverse transcription (RT) in the presence of 200  $\mu$ Ci [ $\alpha$ -<sup>33</sup>P]deoxyadenosine 5'-triphosphate (DuPont NEN, Boston, MA, USA) using Moloney murine leukemia virus reverse transcriptase according to the manufacturer's instructions (Promega). Prehybridization was carried out in 20 ml prehybridization solution (6 $\times$ SSC, 0.5% SDS, 5 $\times$ Denhardt's solution, 100  $\mu$ g/ml denatured salmon sperm DNA) at 68°C for 4 h. Overnight hybridization with the <sup>33</sup>P-labeled cDNA in 10 ml hybridization solution (6 $\times$ SSC, 0.5% SDS, 100  $\mu$ g/ml salmon sperm DNA) was followed by stringent washing (0.1 $\times$ SSC, 0.5% SDS, at 65°C for 1 h). Membranes were exposed to a phosphor screen overnight and scanned using a FLA-3000A plate/fluorescent image analyzer (Fuji Photo Film, Tokyo, Japan). The data analyses of hybridized membranes were performed using Array Gauge software (Fuji Photo Film).

### Semi-quantitative RT-PCR

Total RNA (1  $\mu$ g) was incubated with DNase I and reverse-transcribed with oligo(dT) using the superscript II RT-PCR system (Life Technologies). RT product (1  $\mu$ l) was amplified with primer pairs specific for the genes to be studied. Semi-quantitative multiplex PCR assays were designed to compare the RT-PCR products of the genes

under study with  $\beta$ -actin transcript. Conditions and primer sequences are available on request. Each RT-PCR product was loaded on 1.5% agarose gel containing 0.5  $\mu$ g/ml ethidium bromide. Gel images were made and saved using the Gel Documentation System (UVP, Upland, CA, USA).

### Cloning of full-length cDNAs

ESTs from dbEST and other databases were selected out; these ESTs were highly homologous with known ESTs from our cDNA library. Assemblies were carried out among known ESTs and selected ESTs using Autoassembler software (PE Applied Biosystems, Foster City, CA, USA). Overlapping EST sequences were assembled into contigs. The contigs were checked to see whether they contained open-reading frames using DNA Strider 1.2 software (PE Applied Biosystems). The 5' or 3' sequences of the contigs were selected to be blasted for homologous ESTs in dbEST and other databases, then assemblies were repeated until complete open-reading frames were finished. Nucleic acid sequences and translated amino acid sequences of full-length cDNAs were blasted in GenBank and protein databases or other databases. If there was no or little homology with known genes and proteins, a novel gene was cloned. Sequences of novel genes were amplified by RT-PCR and confirmed by sequencing analyses.

## Results

### General information about gene expression in human insulinoma tissue

A total of 2650 clones were picked out randomly from a cDNA library of insulinoma. Partial cDNA sequencing of each clone was taken from the 5' end using appropriate primers. As a result, 2271 good-quality clones were produced, with a success rate of sequencing reaching 85.7%. After mitochondrial, ribosomal and repeated sequences were screened out, a total of 1589 valid ESTs were obtained. Of these, after integration using Cat software, 671 known genes (42.2% of 1589) accounted for 513 clusters, 582 known ESTs (36.6%) for 508 clusters and 336 novel ESTs (21.2%) for 289 clusters.

cDNA probes transcribed from mRNA of human insulinoma tissue were hybridized with a cDNA array containing 14 000 genes or clones of ESTs. The results of hybridization were processed with Array Gauge software (Fuji Photo Film). After subtraction of background chosen from an area where no cDNA was spotted, genes with intensities  $> 10$  and a ratio of the intensities between the double-offset spots  $< 1.5$  were considered positive signals (i.e. expressed genes) to ensure that they were distinguished from background with statistical

significance  $> 99.9\%$ . As a result, a total of 975 clones were found expressed in human insulinoma tissue: 616 known genes accounted for 63.2%, 275 known ESTs accounted for 29.2% and 74 novel ESTs for 7.6%. To confirm the accuracy and validity of the results of cDNA array, several expressed clones, including the highest and lowest expressed, were selected for RT-PCRs. The outcomes of RT-PCRs proved the cDNA array to be accurate and valid.

The highly expressed genes in human insulinoma are listed in Tables 1 and 2 and some of these genes were chosen to undergo semi-quantitative RT-PCR, as shown in Fig. 1. Using ESTs sequencing, G protein  $\alpha$ -stimulating activity polypeptide  $Gs\alpha$  (14 copies) and insulin (13 copies) were the two most highly expressed genes in insulinoma tissue. In the cDNA array, carboxypeptidase E (CPE) was the most highly expressed. These data demonstrate that the resource of insulinoma tissue was full of reliability, and the established profiling reflected the actual gene expression in human insulinoma tissue.

Furthermore,  $Gs\alpha$  expression was determined in the tumoral tissues of four different insulinomas. It was also examined in normal pancreatic tissues from four adult males who had died of accidents. The results shown in Fig. 3 demonstrate that the expression of  $Gs\alpha$  was significantly elevated in human insulinomas.

### The assignment of gene ontology to known genes in the profiling

After different ESTs derived from the same gene were integrated, a total of 872 clusters were produced using EST sequencing and cDNA array. A directed acyclic graphical (DAG) classification system defined by the Gene Ontology (GO) Consortium (Ashburner *et al.* 2000) was adopted to categorize the cDNA clusters corresponding to known genes expressed in human insulinoma tissue. The usefulness of the GO ontologies was confirmed repeatedly in the annotations of gene expression data, especially after these had been clustered by similarities in patterns of gene expression (Spellman *et al.* 1998). The assignment of GO functions to the proteins represented by the cDNA clusters of insulinoma was performed computationally using an algorithm associating the translated protein domains with GO functions (Schug *et al.* 2002). The distributions of the top-level GO function assignments for the cDNA clusters corresponding to known genes of insulinoma are shown in Table 3.

The data of GO function assignments indicated that the genes involved in nucleic acid binding and in signal transduction were strongly expressed in insulinoma tissue; this was largely in agreement with previous studies pertaining to gene expression profilings in other human tissues (Liew *et al.* 1994, Mao *et al.* 1998, Hu *et al.* 2000,

**Table 1** Most highly expressed genes in human insulinoma using EST sequencing

UniGene	Gene name	Number	%
Hs.273385	Guanine nucleotide binding protein (G protein), $\alpha$ stimulating activity polypeptide 1 (GNAS1)	14	0.88
Hs.89832	Insulin	13	0.82
Hs.323806	Gastric-related protein VRG107	12	0.76
Hs.324720	cDNA FLJ14112 fis, clone MAMMA1001683	11	0.69
Hs.164280	Adenine nucleotide translocator, member 6	10	0.63
Hs.367825	Ferritin-associated protein	8	0.5
Hs.1032	Regenerating islet-derived 1 $\alpha$ (REG1A)	8	0.5
Hs.116428	Secretagogin	6	0.38
Hs.74564	Signal sequence receptor, $\beta$ (SSR2)	5	0.31
Hs.177656	Calmodulin 1 (CALM1)	5	0.31

Kaestner *et al.* 2003). The study of Hu *et al.* (2000) also showed that genes participating in signaling/cell communication were expressed abundantly in the HPA axis. It is worth noting, however, that the proportion of genes involved in signal transduction in the insulinoma tissue was higher than those involved in the HPA axis.

### Genes involved in the protein/insulin secretion pathway

Most eukaryotic cells are equipped with a process exporting biosynthesized proteins to the cell exterior which is referred as secretion. Previous studies indicated that the secretory pathways were similar in different animal cells or yeast cells, although different proteins were secreted by those cells (Palade 1975, Deshaies *et al.* 1988). As a cluster of endocrine cells, the  $\beta$ -cells of the pancreatic islets take on robust activities of secretion; this was also confirmed by our study. The genes involved in the secretory pathway of proteins were highly expressed in insulinoma tissue. Although it was not asserted that all those genes participated in the insulin secretory pathway, most of them were probably associated with insulin secretion due to the predominance of insulin in the secretory pathways of proteins in insulinoma tissue. The proposed insulin secretion-related genes found in our

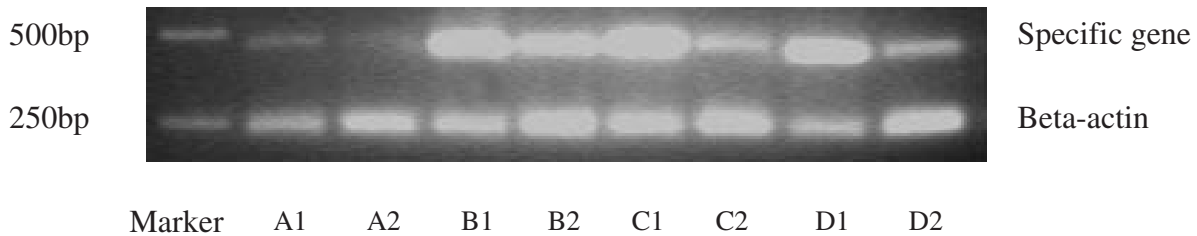
study are shown in Table 4. Furthermore, some of those were picked out to undergo RT-PCRs, as shown in Fig. 2.

### Cloning of novel full-length cDNAs in human insulinoma tissue

In total, eight novel full-length cDNAs were cloned in human insulinoma tissue. Intensive studies on the novel genes were conducted using bioinformatical and experimental methods. In order to confirm the sequences of the novel genes, we designed appropriate primers capable of stretching over the open reading frame (ORF) regions, then the segments of expected sizes were amplified in insulinoma tissue by RT-PCR. After the segments were purified from PCR products, they were subcloned into plasmid pGEM-T and sequenced. In addition, functional predictions of the proteins deduced from the novel genes were carried out using appropriate software. Some evolutionary conserved structures that had implications for the proteins' functions were detected; Pfam 6.6 software automatically recognized to which of the existing protein families these new proteins belonged (<http://pfam.wustl.edu/>). By using CBS prediction servers (<http://www.cbs.dtu.dk/servers/>), signal peptides and transmembrane structures were found to exist in some of these new proteins. As a result, three proteins with

**Table 2** Most highly expressed genes in human insulinoma using cDNA array

UniGene	Gene name	Signal/Noise Ratio
Hs.75360	Carboxypeptidase E (CPE)	25.38
Hs.110029	Protein disulfide isomerase A3 (PDIA3)	22.94
Hs.273385	Guanine nucleotide-binding protein (G protein), $\alpha$ stimulating activity polypeptide 1 (GNAS1)	21.97
Hs.8074	Brain-specific angiogenesis inhibitor 3 (BAI3)	21.96
Hs.225979	Small G proteins, especially RAP-2A (LOC57826)	21.61
Hs.154387	KIAA0103 gene product (KIAA0103)	20.6
Hs.166091	Ligase IV, DNA, ATP-dependent (LIG4)	19.36
Hs.44205	Cortistatin (CORT)	18.64
Hs.2230	Epidermal growth factor ( $\beta$ -urogastrone) (EGF)	17.82
Hs.75183	Cytochrome P450IIE1 (ethanol-inducible)	17.74



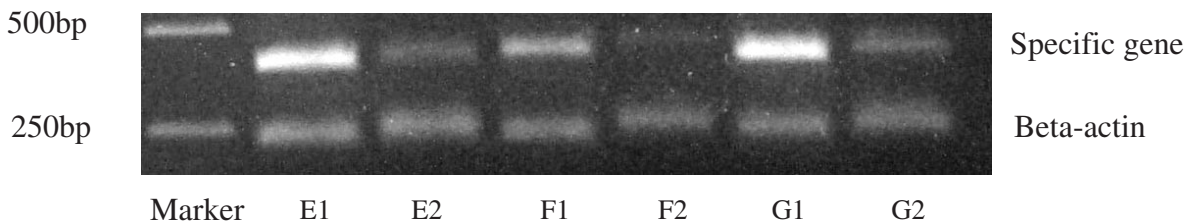
**Figure 1** The result of semi-quantitative RT-PCR of some highly expressed genes in human insulinoma. 1, insulinoma tissue; 2, normal pancreatic tissue. A, B, C and D represent insulin amyloid polypeptide (IAPP), GNAS1, CPE and insulin respectively.

signal peptides and two with obvious transmembrane structures were elucidated. Detailed information about the novel genes was released in GenBank (accession numbers: AY194292, AY194293 and AY194294; the others are submitted).

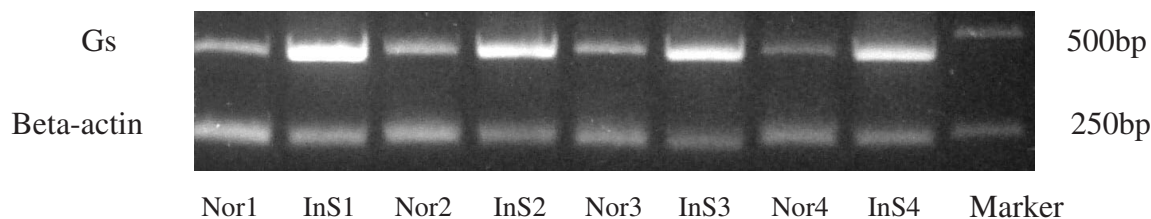
## Discussion

In a previous study, our group characterized the gene expression profile of the HPA axis using EST analysis and molecular cloning of full-length cDNA of novel genes identified; this formed a basis for a more profound understanding of the regulation of the neuroendocrine system under physiological conditions as well as for the further analysis of possible disease association (Hu *et al.* 2000). From this basis, we constructed a cDNA library and established gene expression profiling in human insulinoma tissue. Furthermore, cDNA array was used as an additional method for EST sequencing to elucidate the gene expression of insulinomas; the cDNA array provided a simple and natural vehicle for exploring the genome in a way that is both systemic and comprehensive. In total, 1589 valid ESTs corresponding to 1310 clusters and 975 clones were produced out of human insulinoma tissue using EST sequencing and cDNA array respectively. Using the two methods, the genes involved in the protein/insulin secretory pathway were highly expressed in insulinoma tissue. Meanwhile, the data from GO functions indicated that genes related to signal transduction were strongly expressed in human insulinoma tissue compared with other studies in this field.

Heterotrimeric G proteins function as molecular switches between varieties of seven spanning receptors and intracellular effectors in the signal transduction pathway. It is well established that the activating mutations of the gene encoding the  $\alpha$ -subunit of Gs( $Gs\alpha$ ) are involved in the pathogenesis of several tumors such as growth hormone-secreting pituitary adenomas, toxic thyroid adenomas, differentiated thyroid adenocarcinomas and McCune–Albright syndrome, etc. (Landis *et al.* 1989, Lyons *et al.* 1990, O’Sullivan *et al.* 1991, Suarez *et al.* 1991, Schwindinger *et al.* 1992, Ballare *et al.* 1998). These activating mutations usually occur at two hot spots, codon 201 or 227, resulting in constitutive activation of adenylyl cyclase by impairing the intrinsic guanosine triphosphatase activity of the subunit. In the cDNA library of insulinoma we constructed, the  $Gs\alpha$  gene was the most highly expressed and it was also found strongly expressed in the cDNA array. We also examined the expression of  $Gs\alpha$  in three other insulinomas. The results indicated that the gene for  $Gs\alpha$  was much more highly expressed in insulinomas than in normal pancreatic tissues. Similarly, Zeiger & Norton (1993) found that the gene for  $Gs\alpha$  showed a 30-fold higher expression in insulinoma than in normal pancreas. There was also a three-fold increase in expression of  $Gs\alpha$  by insulinomas over that of normal human islet cells. Several researchers also indicated that the levels of wild-type  $Gs\alpha$  protein in GH-secreting tumors and thyroid tumors were much higher than in non-functioning pituitary tumors and normal thyroid tissue respectively (Siperstein *et al.* 1991, Bertherat *et al.* 1995). Until now, no mutations in the gene for  $Gs\alpha$  have



**Figure 2** The result of semi-quantitative RT-PCR of some genes involved in the secretory pathway of insulin in human insulinoma. 1, insulinoma tissue; 2, normal pancreatic tissue. E, F and G represent SSR2, GOSR1 and secretagogin respectively.



**Figure 3** The result of semi-quantitative RT-PCR of the expression of the gene for *Gsα* in human insulinomas and normal pancreatic tissues. Nor, normal pancreatic tissues; InS, human insulinoma tissues.

been identified in human insulinomas. Vessey *et al.* (1994) reported the absence of mutations in the *Gsα* gene in nine insulinomas. There was also no evidence of a mutation in the gene for *Gsα* in Zeiger's study (Zeiger & Norton 1993). In our study, the mutated regions of the *Gsα* gene implicated in other endocrine tumors were examined and no mutations were found. Both activating mutations and overexpression of the *Gsα* gene could constitutively stimulate adenylyl cyclase and result in elevated accumulation of second messenger cAMP, which could consequently activate downstream effectors. In this regard, Bertherat *et al.* (1995) demonstrated that overexpression of wild-type *Gsα* constitutively stimulates the phosphorylation of cAMP response-element binding protein (CREB) and cAMP-regulated (CRE)-dependent transcription in somatotrophs, which promotes the transformation of pituitary somatotrophs. Regarding overexpression of the *Gsα* gene in human insulinoma tissue compared with normal pancreatic tissue, it was

strongly implied that the tumorigenesis of insulinomas was related to the overexpression of the *Gsα* gene.

The prominent characteristic of insulinoma is hyperinsulinism. In other words, the release of insulin into blood is abnormally elevated as a result of excessive secretion and/or biosynthesis of insulin in the  $\beta$ -cells of pancreatic islets. Surprisingly, the transcripts encoding genes necessary for insulin biosynthesis, such as transcriptional factor PDX-1 (one copy in the cDNA library), were not found at high levels of expression in insulinoma tissue. Significantly, a large cluster of genes functioning in the secretory pathway were detected with abundant expression. The predominance of the genes participating in the secretory pathway suggested that regulation of secretion might be a major mechanism by which insulin release was abnormally increased in patients with insulinomas. Of these genes, some had been reported previously by other researchers to be related to insulin secretion. For example, CPE which specifically expressed in neuroendocrine tissues was the most highly expressed gene in insulinoma tissue, using the cDNA array. CPE plays an important role in the secretory pathway and the maturation of insulin. It serves as a regulated secretory pathway-sorting receptor to co-aggregate with insulin and correctly target insulin into regulated secretory granules (Rindler 1998). Moreover, CPE removes C-terminal basic amino acid residues following endopeptidase cleavage of proinsulin and prompts proinsulin conversion to mature and active insulin (Fricker 1988). Secretagogin recently cloned from a human pancreatic  $\beta$ -cell cDNA library by Wagner *et al.* (2000) was also found highly expressed in human insulinoma tissue (eight copies in the cDNA library). Intracellular  $\text{Ca}^{2+}$ -binding proteins such as calbindin and calretinin, which are highly homologous to secretagogin in structure, have been determined as playing central roles in the insulin release triggered by  $\text{Ca}^{2+}$  flux (Redecker & Cetin 1997, Sooy *et al.* 1999). It has been shown that insulin secretion of secretagogin-transfected RIN cells was much higher than that of controls (Wagner *et al.* 2000). We speculate that high expression of secretagogin is associated with insulin oversecretion in human insulinoma tissue.

**Table 3** GO function assignments of the gene clusters in human insulinoma

GO function	Number of clusters	Percentage of total
Nucleic acid binding	167	19.2
Signal transducer	158	18.1
Ligand binding or carrier	114	13.1
Transporter	93	10.7
Structural protein	83	9.5
Enzyme	78	8.9
Cell adhesion molecule	39	4.5
Chaperone	37	4.2
Cell-cycle regulator	30	3.4
Motor	25	2.9
Microtubule binding	23	2.6
Apoptosis regulator	10	1.2
Enzyme activator	5	0.6
Defense/immunity protein	4	0.5
Enzyme inhibitor	2	0.2
Cytoskeletal regulator	2	0.2
Protein tagging	1	0.1
Antioxidant	1	0.1
Total	872	100

**Table 4** Genes involved in the secretory pathway of protein/insulin in human insulinoma

Functional categories	Gene
Protein translocation	Signal-recognition particle (SRP) receptor $\alpha$
	SRP receptor $\beta$
	Translocon-associated protein $\beta$ (TRAP $\beta$ )
	Translocon-associated protein $\delta$ (TRAP $\delta$ )
Protein folding quality control	Protein disulfide isomerase A3
	Peptidylprolyl isomerase D (cyclophilin D)
	Prohormone convertase 3
	Calnexin
	Endoplasmic reticulum chaperone SIL1, homolog of yeast Stress-associated endoplasmic reticulum protein 1
Protein glycosylation	Mannosidase, $\alpha$ , class 2A, member 2 (MAN2A2)
	O-linked <i>N</i> -acetylglucosamine (GlcNAc) transferase (UDP- <i>N</i> -acetylglucosamine: polypeptide- <i>N</i> acetylglucosaminyltransferase) (OGT)
	Lectin, galactoside-binding, soluble, 2 (galectin2) (LGALS2)
	Glucuronyltransferase I
Vesicle-mediated transportation	Golgi SNAP receptor complex member 1 (GOSR1)
	Adaptor-related protein complex 4, $\epsilon$ -1 subunit (AP4E1)
	SEC24( <i>S.cerevisiae</i> )-related gene family, member D (SEC24D)
	Pleckstrin homology, Sec7 and coiled/coil domains 1 (cytohesin 1)(PSCD1), transcript variant 1
	Pleckstrin homology, Sec7 and coiled/coil domains 3 SEC14L
	Catenin (cadherin-associated protein), $\alpha$ -1 (102 kD) (CTNNA1)
	Annexin V (ANX5)
	Annexin A6 (ANXA6), transcript variant 1
	Oatmer protein complex, subunit $\alpha$ (COPA)
	Vesicle transport-related protein Transport-secretion protein 2.2
Secretory vesicles budding/exocytosis	Carboxypeptidase E (CPE)
	Chromogranin B (secretogranin 1) (CHGB)
	Secretory carrier membrane protein 1
	$\beta$ SNAP (soluble NSF attachment protein)
	Secretagogin (SECRET) Suppressor of K <sup>+</sup> transport defect 1

Apart from the genes previously reported to participate in insulin secretion, several genes encoding the proteins related to the secretory pathway were first found expressed in the  $\beta$ -cells of pancreatic islets from our insulinoma cDNA library. Protein trafficking along the secretory pathway is performed by transport vesicles, which bud from a donor membrane and fuse with a target acceptor membrane (Allan *et al.* 2002). The formation of transport vesicles requires co-participation of several proteins among which adaptor protein complexes (APs) serve as assembly particles to select the cargo for inclusion in transport vesicles. Three types of APs: AP1, AP2 and AP3, have been identified and found to mediate specific transport steps (Robinson & Bonifacino 2001). The fourth adaptor protein complex (AP4) had been cloned by Hirst *et al.* (1999). AP4 and the other three APs are heterotetrameric complexes consisting of  $\epsilon$ ,  $\beta$ 4,  $\mu$ 4 and  $\sigma$ 4

subunits. Until now, there have only been eight hits for the  $\epsilon$  subunit in the human EST database; these are derived from human testis, spinal cord, breast, lymph and uterus tissues. Many more hits for its counterparts of the other three APs (184 for  $\gamma$  of AP1, 204 for  $\alpha$  of AP2, and 485 for  $\delta$  of AP3), however, have been generated from various human tissues. It is suggested that the expression level of AP4 may be as many as two orders of magnitude less abundant than other APs. It is also suggested that the AP4 pathway is not very extensively used and AP4 may be essential for the sorting of a very specific type of cargo (Hirst *et al.* 1999). It was noted that the  $\epsilon$  subunit of AP4 was found with relatively high expression (three copies) in the insulinoma cDNA library. From the results of RT-PCR analysis (data not shown), it could be seen that the expression level of the  $\epsilon$  subunit in insulinoma tissue was much higher than that of normal pancreatic tissue.

Although it is not asserted that AP4 specifically participates in the sorting of insulin, there is a strong likelihood that high expression of AP4 is tightly associated with oversecretion of insulin in insulinoma tissue.

In conclusion, the global molecular biology of human insulinoma was well elucidated at transcriptional levels in this study. The strategy of EST sequencing for establishing gene expression profiles has been widely accepted. The data from our study also validated the use of cDNA microarray to describe gene expression patterns. The elevated G protein-mediated signal transduction pathway, especially the high levels of  $G\alpha$  gene, strongly suggested a significant role in the pathogenesis of insulinoma. The large cluster of genes encoding secretory pathway components suggested a mechanism by which  $\beta$ -cells regulate insulin secretion. The genes specifically involved in insulin secretion might be discovered by intensive study of those genes, such as AP4. Meanwhile, further studies on the novel genes cloned from the insulinoma tissue, especially those with signal peptides, are ongoing at our laboratory.

## Acknowledgements

The study was funded by grants to R-MH from the Key Project of National Natural Science Foundation of China (30230380), the National Natural Science Foundation of China (39900072), the Chinese High Tech Program (2002BA711A05 and 2001AA221201), the National Key Basic Research and Development Program (2002CB713703) and the Shanghai Commission for Science and Technology (01JC14026). The work was also funded by a grant to X-CW from the National Natural Science Foundation of China (30300162). We thank Drs Zhao-Yun Zhang and Ji Hu for their help in preparing the manuscript.

## References

- Allan VJ, Thompson HM & McNiven MA 2002 Motoring around the Golgi. *Nature Cell Biology* **4** E236–242.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM & Sherlock G 2000 Gene ontology: tool for the unification of biology. *Nature Genetics* **25** 25–29.
- Ballare E, Mantovani S, Lania A, Di Blasio AM, Vallar L & Spada A 1998 Activating mutations of the Gs alpha gene are associated with low levels of Gs alpha protein in growth hormone-secreting tumors. *Journal of Clinical Endocrinology and Metabolism* **83** 4386–4390.
- Bertherat J, Chanson P & Montminy M 1995 The cyclic adenosine 3',5'- monophosphate-responsive factor CREB is constitutively activated in human somatotroph adenomas. *Molecular Endocrinology* **9** 777–783.
- Deshaies RJ, Koch BD & Schekman R 1988 The role of stress proteins in membrane biogenesis. *Trends in Biochemical Science* **13** 384–388.
- Fricker LD 1988 Activation and membrane binding of carboxypeptidase E. *Journal of Cell Biochemistry* **38** 279–289.
- Hirst J, Bright NA, Rous B & Robinson MS 1999 Characterization of a fourth adaptor-related protein complex. *Molecular Biology of the Cell* **10** 2787–2802.
- Hu RM, Han ZG, Song HD, Peng YD, Huang QH, Ren SX, Gu YJ, Huang CH, Li YB, Jiang CL, Fu G, Zhang QH, Gu BW, Dai M, Mao YF, Gao GF, Rong R, Ye M, Zhou J, Xu SH, Gu J, Shi JX, Jin WR, Zhang CK, Wu TM, Huang GY, Chen Z, Chen MD & Chen JL 2000 Gene expression profiling in the human hypothalamus–pituitary–adrenal axis and full-length cDNA cloning. *PNAS* **97** 9543–9548.
- Kaestner KH, Lee CS, Scearce LM, Brestelli JE, Arsenlis A, Le PP, Lantz KA, Crabtree J, Pizarro A, Mazzarelli J, Pinney D, Fischer S, Manduchi E, Stoeckert CJ Jr, Gradwohl G, Clifton SW, Brown JR, Inoue H, Cras-Meneur C & Permutt MA 2003 Transcriptional program of the endocrine pancreas in mice and humans. *Diabetes* **52** 1604–1610.
- Landis CA, Masters SB, Spada A, Pace AM, Bourne HR & Vallar L 1989 GTPase inhibiting mutations activate the alpha chain of Gs and stimulate adenylyl cyclase in human pituitary tumours. *Nature* **340** 692–696.
- Larsen PR, Kronenberg HM, Melmed S & Polonsky KS 2002 *Williams Textbook of Endocrinology*, 10th edition. Pennsylvania: Saunders.
- Liew CC, Hwang DM, Fung YW, Laurensen C, Cukerman E, Tsui S & Lee CY 1994 A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *PNAS* **91** 10645–10649.
- Lyons J, Landis CA, Harsh G, Vallar L, Grunewald K, Feichtinger H, Duh QY, Clark OH, Kawasaki E & Bourne HR 1990 Two G protein oncogenes in human endocrine tumors. *Science* **249** 655–659.
- Mao M, Fu G, Wu JS, Zhang QH, Zhou J, Kan LX, Huang QH, He KL, Gu BW, Han ZG, Shen Y, Gu J, Yu YP, Xu SH, Wang YX, Chen SJ & Chen Z 1998 Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *PNAS* **95** 8175–8180.
- O'Sullivan C, Barton CM, Staddon SL, Brown CL & Lemoine NR 1991 Activating point mutations of the gsp oncogene in human thyroid adenomas. *Molecular Carcinogens* **4** 345–349.
- Palade G 1975 Intracellular aspects of the process of protein synthesis. *Science* **189** 347–358.
- Redecker P & Cetin Y 1997 Rodent pancreatic islet cells contain the calcium-binding proteins calcineurin and calretinin. *Histochemistry and Cell Biology* **108** 133–139.
- Rindler MJ 1998 Carboxypeptidase E, a peripheral membrane protein implicated in the targeting of hormones to secretory granules, co-aggregates with granule content proteins at acidic pH. *Journal of Biological Chemistry* **273** 31180–31185.
- Robinson MS & Bonifacio JS 2001 Adaptor-related proteins. *Current Opinions in Cell Biology* **13** 444–453.



- Schug J, Diskins S, Mazzarelli J, Brunk BP & Stoeckert CJ Jr 2002 Predicting gene ontology functions from ProDom and CDD protein domains. *Genome Research* **12** 648–655.
- Schwindinger WF, Francomano CA & Levine MA 1992 Identification of a mutation in the gene encoding the alpha subunit of the stimulatory G protein of adenylyl cyclase in McCune–Albright syndrome. *PNAS* **89** 5152–5156.
- Siperstein AE, Miller RA, Landis C, Bourne H & Clark OH 1991 Increased stimulatory G protein in neoplastic human thyroid tissues. *Surgery* **110** 949–955.
- Sooy K, Schermerborn T, Noda M, Surana M, Rhoten WB, Meyer M, Fleischer N, Sharp GW & Christakos S 1999 Calbindin-D(28k) controls  $[Ca^{2+}]_i$  and insulin release. Evidence obtained from calbindin-d(28k) knockout mice and beta cell lines. *Journal of Biological Chemistry* **274** 34343–34349.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D & Futcher B 1998 Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* **9** 3273–3297.
- Suarez HG, Duvillard JA, Caillu B, Schlumberger M, Parmentier C & Monier R 1991 Gsp mutations in human thyroid tumours. *Oncogene* **6** 677–679.
- Vessey SJR, Jones PM, Wallis SC, Schofield J & Bloom SR 1994 Absence of mutations in the Gs $\alpha$  Gi2 $\alpha$  genes in sporadic parathyroid adenomas and insulinomas. *Clinical Science* **87** 493–497.
- Wagner L, Oliyarnyk O, Gartner W, Nowotny P, Groeger M, Kaserer K, Waldhausl W & Pasternack MS 2000 Cloning and expression of Secretagoin, a novel neuroendocrine and pancreatic islet of Langerhans-specific Ca<sup>2+</sup>-binding protein. *Journal of Biological Chemistry* **275** 24740–24751.
- Zeiger MA & Norton JA 1993 Gs alpha-identification of a gene highly expressed by insulinoma and other endocrine tumors. *Surgery* **114** 458–463.