Gene expression profiling in the human hypothalamus-pituitary-adrenal axis and full-length cDNA cloning

R.-M. Hu*†, Z.-G. Han†‡§, H.-D. Song*†, Y.-D. Peng*†‡, Q.-H. Huang*†‡, S.-X. Ren‡, Y.-J. Gu*‡, C.-H. Huang‡, Y.-B. Li‡, C.-L. Jiang‡, G. Fu‡, Q.-H. Zhang*, B.-W. Gu*, M. Dai*, Y.-F. Mao*, G.-F. Gao*, R. Rong*, M. Ye*, J. Zhou‡, S.-H. Xu‡, J. Gu‡, J.-X. Shi‡, W.-R. Jin‡, C.-K. Zhang‡, T.-M. Wu‡, G.-Y. Huang‡¶, Z. Chen*‡¶, M.-D. Chen*, and J.-L. Chen*§

*Rui-Jin Hospital, Shanghai Institute of Endocrinology, Shanghai Second Medical University, 197 Rui-Jin Road II, Shanghai 200025, China; ‡Chinese National Human Genome Center, 351 Guo Shou-Jing Road, Shanghai, 201203, China; and ¶Morgan-Tan International Center for Life Science, Institute of Genetics, Fudan University, 220 Han Dan Road, Shanghai 200433, China

Communicated by Jiazhen Tan, Fudan University, Shanghai, China, June 13, 2000 (received for review February 2, 2000)

The primary neuroendocrine interface, hypothalamus and pituitary, together with adrenals, constitute the major axis responsible for the maintenance of homeostasis and the response to the perturbations in the environment. The gene expression profiling in the human hypothalamus-pituitary-adrenal axis was catalogued by generating a large amount of expressed sequence tags (ESTs), followed by bioinformatics analysis (http://www.chgc.sh.cn/database). Totally, 25,973 sequences of good quality were obtained from 31,130 clones (83.4%) from cDNA libraries of the hypothalamus, pituitary, and adrenal glands. After eliminating 5,347 sequences corresponding to repetitive elements and mtDNA, 20,626 ESTs could be assembled into 9,175 clusters (3,979, 3,074, and 4,116 clusters in hypothalamus, pituitary, and adrenal glands, respectively) when overlapping ESTs were integrated. Of these clusters, 2,777 (30.3%) corresponded to known genes, 4,165 (44.8%) to dbESTs, and 2,233 (24.3%) to novel ESTs. The gene expression profiles reflected well the functional characteristics of the three levels in the hypothalamus-pituitary-adrenal axis, because most of the 20 genes with highest expression showed statistical difference in terms of tissue distribution, including a group of tissue-specific functional markers. Meanwhile, some findings were made with regard to the physiology of the axis, and 200 full-length cDNAs of novel genes were cloned and sequenced. All of these data may contribute to the understanding of the neuroendocrine regulation of human life.

The neuroendocrine system plays a primordial role in the regulation of major physiological processes such as development, growth, metabolism, reproduction, and adaptation to environment. The primary neuroendocrine interface, hypothalamus (HT) and pituitary gland (NP), together with adrenal glands (AD), constitute the major axis responsible for the maintenance of homeostasis and the response to perturbations in the environment (1). Hypothalamus-pituitary-adrenal (HPA) response to stress also serves to mobilize the defensive mechanisms of the body including those modulating the intensity of the immune response and some inflammatory components, such as changes in vascular tone and vascular permeability, etc.

The diverse functions and complex regulation of the HPA axis are largely determined by well-regulated gene expression in tissues at different levels of the axis. The use of molecular biology techniques over the last two decades has allowed molecular cloning of a number of genes encoding hormones or secretory proteins. Apart from classical endocrine hormones, new active peptides related to neuroendocrine function, such as CART (cocaine and amphetamine-regulated transcript peptide) (2), PACAP (pituitary adenylate cyclase-activating polypeptide) (3), leptin, orexin (4), neuropeptide Y (5), endothelin (6), and many others, recently were identified. However, there are still a lot of unknown functions as well as regulatory mechanisms to be explored in the HPA axis. The establishment of a detailed catalog of genes expressed in the HPA axis, and the discovery of new genes from HT, NP, and AD, will certainly help the functional characterization of this utmost important system for human life and contribute to functional genomics as a whole.

In the present work, a relatively large-scale generation of expressed sequence tags (ESTs) was conducted in combination with analysis using bioinformatics tools. A catalog of genes expressed in the human HPA axis was established, and full-length cDNA cloning was carried out based on the EST database.

Materials and Methods

RNA Extraction. Normal human hypothalami, pituitaries, and adrenals were removed within 4 h postmortem from two adult males aged 20 and 35 years who died in traffic accidents. Total RNA was extracted from frozen tissues, and the selection of poly(A) RNA was performed by using oligo(dT) (Qiagen, Chatsworth, CA).

cDNA Library Construction. cDNA synthesis was performed either with a CapFinder PCR cDNA library construction kit (CLONTECH) or by using conventional procedures (7).

DNA Sequencing. Bacteria growth and plasmid extractions were performed in a 96-well format (Qiagen). Sequencing reactions were performed on a 9600 Thermal reactor (Perkin–Elmer) by using a Dye Primer Cycle Sequencing Kit (Perkin–Elmer), and partial cDNA sequencing of each clone was taken from the 5′

Abbreviations: HPA, hypothalamus-pituitary-adrenal; EST, expressed sequence tag; HT, hypothalamus; NP, pituitary gland; AD, adrenal gland; CART, cocaine and amphetamine-regulated transcript peptide; RT, reverse transcriptase; GH, growth hormone; POMC, proopiomelanocortin; TSH, thyroid-stimulating hormone; RH, radiation hybrid; CRH, corticotropin-releasing hormone.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF070666, AF077046, AF077049, AF077052, AF077054, AF078845, AF078846, AF078849, AF078851, AF078853–AF078855, AF078865–AF078867, AF092128, AF092131, AF092137, AF092139, AF093680, AF100740–AF100744, AF100751–AF100757, AF100759, AF100760, AF100762, AF106681–AF106685, AF110644, AF110645, AF110647, AF110774–AF110778, AF112200–AF112204, AF112206–AF112222, AF113122–AF113129, AF113534–AF113540, AF117229–AF117237, AF119662–AF119666, AF125392–AF125394, AF125530–AF125535, AF126020, AF126021, AF126023, AF126024, AF136970–AF136972, AF136974–AF136978, AF150732–AF150735, AF155569, AF155648–AF166658, AF155660–AF155662, AF157316–AF157327, AF160212–AF160215, AF164790–AF164799, AF183409, AF183413, AF183414, AF183416, AF183420, AF183422–AF183428, AF184213, AF211481, AF212253, AF220182–AF220193, AF221595, AF223466–AF223470, and AF226732).

†R.-M.H., Z.-G.H., H.-D.S., Y.-D.P., and Q.-H.H. contributed equally to this work.

§To whom reprint requests should be addressed. E-mail: zchen@ms.stn.sh.cn or hanzg@chgc.sh.cn.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.160270997. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.160270997

end. The reaction products were analyzed by using an ABI 377 DNA Sequencer (Perkin–Elmer).

Bioinformatics Analysis and Data Management. Quality assessment and quality base trimming of ESTs were performed by using PE SEQUENCING ANALYSIS (version 3.0) and PHRED. "Good" sequences referred to those sequences containing less than 3% ambiguous bases that were longer than 100 bp. ESTs were considered as part of known genes if they shared at least 95% homology over at least 100 bp of DNA sequences with a BLAST search. ESTs corresponding to known genes were divided into eight categories according to the gene functions as proposed in the literature (7, 8). Clustering of the ESTs generated in the work was performed by using CAT3.2 from Pangea (Oakland, CA) with default parameters.

Full-Length cDNA Cloning. The new sequences confirmed by homology comparison with GenBank and other databases were selected for full-length cDNA cloning. A Marathon Ready cDNA kit was used to facilitate full-length cDNA cloning. *In silico* cloning was carried out by using dbEST information started from the sequences obtained from HT, NP, and AD libraries. Overlapping EST sequences were assembled to establish contigs. Sequence ambiguity existing in these contigs was clarified by further sequencing.

Quantitative Reverse Transcriptase (RT)-PCR. The quantitative RT-PCR for growth hormone (GH), proopiomelanocortin (POMC), and thyroid-stimulating hormone (TSH) genes was performed. Total RNA with DNase I treatment was used to synthesize first-stand cDNA with RT (GIBCO/BRL) and oligo(dT) 15 Primer (Promega). The products of the RT reactions was used to seed real-time PCR by using an ABI Prism 7700 Sequence Detector by comparing with glyceraldehyde-3-phosphate dehydrogenase (internal control) and individual standard curve with three time repeats. Probes were labeled with quencher and fluorescent dye 6-FAM by 5′ and 3′ ends, respectively.

Mapping of Full-Length cDNA Clones. We mapped novel genes to chromosomes by two strategies: searching UNIGENE or radiation hybrid (RH). In RH mapping, the GENEBRIDGE G4 panel (Research Genetics, Huntsville, AL) was used according to previously described protocol (9).

Results

Quality Evaluation of cDNA Libraries and an Overview of ESTs from HT, NP, and AD. A total of 31,130 clones were selected at random from cDNA libraries of HT, NP, and AD. Partial cDNA sequence of each clone was taken from the 5′ end by using appropriate primers, generating 25,973 ESTs of good quality, with an overall success rate reaching 83.4% (http://www.chgc.sh.cn/database) (Fig. 1*A*). Of these, 24,044 ESTs were generated from the conventional cDNA libraries whereas only 1,929 ESTs were derived from the CapFinder libraries. When ESTs corresponding to the same known genes were grouped into clusters, the percentages of clusters containing more than five clones in the HT, NP, and AD cDNA libraries were 9.1% (114 clusters), 8.0% (85 clusters), and 10.3% (158 clusters), respectively. Among those 273 clusters with high EST copy number, 67.4% were genes previously reported to be ubiquitously expressed in more than 20 tissues (8). Hence, ESTs generated from the three tissues in the HPA axis seemed to have a good representation of gene expression. Based on these data, we believe that EST copy number for a given gene could be considered as a reflection of the gene expression level.

After bioinformatics analysis, 5,347 ESTs homologous to Alu, L1, and other repetitive elements and mtDNA were put aside. The remaining 20,626 ESTs were further analyzed and assembled into clusters for identical or overlapping clones. The numbers of clusters from HT, NP, and AD were 3,979, 3,075, and 4,116, respectively. If the EST clusters from the three tissues were combined and integrated to remove those appearing in more than one tissue, then a total of 9,175 clusters were identified. These clusters then were divided into three groups. The first group included sequences identical to known genes in GenBank, the second group contained clusters matching with the sequences in dbEST, and the third group comprised clusters not present in the public database and thus could be temporarily considered as novel ESTs. To address whether the novel ESTs represented previously unidentified mRNA species, we also sequenced the 3′ ends of 1,156 clones whose 5′ ends were considered novel. As a result, 14.6% (169 clones) were identical to known genes in GenBank, 37.7% (390 clones) matched with the sequences in dbEST, and 49.3% (570 clones) were truly novel ESTs.

To evaluate the proportion of cDNAs containing full-length ORF, we selected at random 353 ESTs corresponding to known genes from HT, NP, and AD cDNA libraries and made an bioinformatics analysis. Among these sequences the proportion of cDNA containing the first ATG in the 5′ terminal sequence reached 36%. Therefore, our libraries were relatively enriched in full-length cDNA.

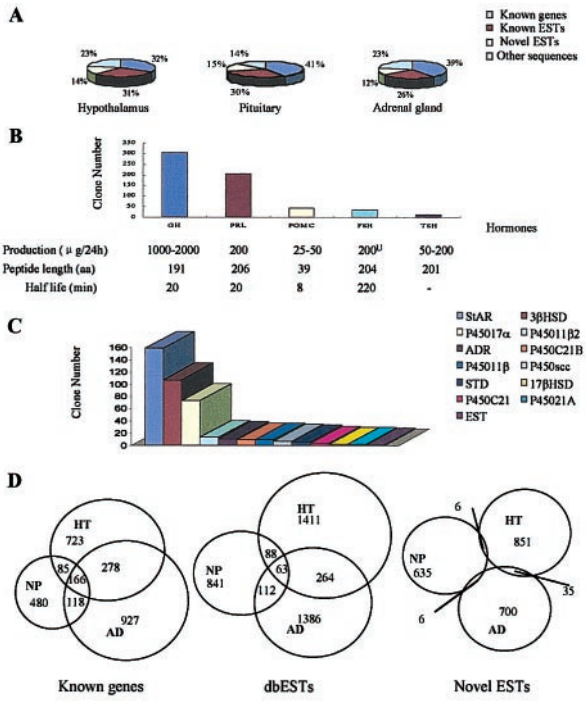

Fig. 1. Gene expression profile in whole HT, NP, and AD. (*A*) Schematic chart of ESTs from HT (8,876 ESTs), NP (7,221 ESTs), and AD (9,875 ESTs) libraries. (*B*) The order of gene expression levels of five classical NP hormones. Corticotropin was grouped into POMC. Real-time PCR performed for GH, POMC, and TSH generated gene expression levels (32×10^5, 5.9×10^5, and 2.4×10^5 copies/100 ng RNA, respectively, P value all < 0.001) proportional to those estimated by EST copy number. (*C*) The order of steroids biosynthesis-related enzymes and proteins in AD. StAR, steroidogenic acute regulatory protein; 3βHSD, type II 3β hydroxysteroid dehydrogenase/5-δ-4-δ isomerase; P450 17α, cytochrome P450, subfamily XVII (steroid 17α-hydroxylase); P45011b2, cytochrome P450 XIB2 (aldosterone synthase); ADR, adrenodoxin; P45°C21B, 21-hydroxylase B; P45011β, cytochrome P450 11β; P450scc, cholesterol side-chain cleavage enzyme P450scc; STD, dehydroepiandrosterone sulfotransferase; 17βHSD: 17-β-hydroxysteroid dehydrogenase; P45°C21, steroid 21-hydroxylase [P450(C21)]; P45021A, 21-hydroxylase A; EST, estrogen sulfotransferase. (*D*) Overlapping of EST clusters corresponding to known genes, known ESTs, and novel ESTs expressed in HT, NP, and AD.

Table 1. Copies of the first 20 known genes mostly highly expressed in HT, NP, and AD

HT		NP		AD	
Gene name	Copies (%)	Gene name	Copies (%)	Gene name	Copies (%)
Proteolipid protein	70 (0.789)*†	Growth hormone	302 (4.182)*‡	Steroidogenic acute regulatory protein	159 (1.610)†‡
Myelin basic protein	66 (0.744)*†	Prolactin	206 (2.853)*‡	Type II 3β hydroxysteroid dehydrogenase	107 (1.084)†‡
Glial fibrillary acidic protein	22 (0.248)*†	POMC	43 (0.595)*‡	Steroid 17α-hydroxylase	73 (0.739)†‡
Neuronal membrane glycoprotein M6b	19 (0.214)*†	FSH α subunit (or chorionic gonadotropin)	36 (0.499)*‡	Aldose reductase	46 (0.466)†‡
PGP 9.5 (neuroendocrine marker protein)	14 (0.158)*	Glyoxalase II	19 (0.263)*‡	Mitochondrial matrix protein P1	35 (0.354)†‡
Amyloid precursor protein	10 (0.113)*†	Tumor susceptibility protein (TSG101)	16 (0.222)*‡	Prostatic binding protein	29 (0.294)†‡
Carboxypeptidase E	10 (0.113)	Carboxypeptidase E	14 (0.194)	KIAA0018	26 (0.263)†‡
Selenoprotein P	10 (0.113)*	Thyroid hormone receptor coactivating protein	14 (0.194)*‡	SH3 binding protein	17 (0.172)†‡
Glutamate transporter	9 (0.101)*†	Brain-expressed HHCPA78 homolog	14 (0.194)*‡	Apo1_Human	14 (0.142)‡
Homologue of mSNAP25	9 (0.101)*†	Electron transfer flavoprotein-ubiquinone oxidoreductase	11 (0.152)*‡	IEF 7442	14 (0.142)†‡
Osteopontin	9 (0.101)*†	Lysosomal pepstatin insensitive protease (CLN2)	10 (0.138)*‡	13-kDa differentiation-associated protein	13 (0.132)†‡
Semaphorin F homolog	9 (0.101)*†	Hs-cul-3	8 (0.111)*‡	Apolipoprotein E	13 (0.132)‡
Tyrosine 3-monooxygenase, zeta polypeptide	9 (0.101)*	Pro-galanin	8 (0.111)*‡	Cytochrome P450 XIB2 (aldosterone synthase)	13 (0.132)†‡
Carbonic anhydrase II	8 (0.090)*†	Reticulocalbin	8 (0.111)*‡	Adrenodoxin	11 (0.111)†‡
Clusterin	8 (0.090)*	RNA-binding protein regulatory subunit	8 (0.111)*	Glutathione peroxidase	11 (0.111)†‡
Neuroendocrine-specific protein A (NSP)	8 (0.090)*†	Secretogranin I (chromogranin B)	8 (0.111)	KIAA0026	11 (0.111)‡
Prostatic binding protein	8 (0.090)*†	hCART	7 (0.097)*‡	KIAA0108	10 (0.101)
RACHI (RACHI)(urea transporter)	8 (0.090)*†	HIV-1 TAR RNA binding protein (TARBP-b)	7 (0.097)*‡	Tazarotene-induced gene 2 (TIG2)	10 (0.101)†‡
Splicing factor SRp40-2 (SRp40)	8 (0.090)†	Insulin-induced protein I (INSIGI)	7 (0.097)*‡	Chromogranin B (secretogranin I)	9 (0.091)‡
Acid ceramidase	7 (0.079)	Ionizing radiation resistance conferring protein	7 (0.097)*‡	CLA-I	9 (0.091)‡

Genes listed do not include those widely expressed (8). % indicates the percentage of a given EST in all ESTs obtained from a tissue. The statistical analysis was performed as described (11) (http://igs-server.cnrs-mrs.fr). *, Significant differences of gene expression ($P < 0.05$) in HT vs. NP. †, Significant differences of gene expression ($P < 0.05$) in HT vs. AD. ‡, Significant differences of gene expression ($P < 0.05$) in NP vs. AD.

Gene Expression Profile in the HPA Axis. In the present work, a catalog of genes expressed in the human HPA axis was established by generating a large amount of ESTs, followed by bioinformatics analysis (available on the web at http://www.chgc.sh.cn/database). A total of 20,626 ESTs were assembled into 9,175 clusters (3,979, 3,074, and 4,116 clusters in HT, NP, and AD, respectively) when overlapping ESTs were integrated. A total of 2,777 clusters (30.3%) corresponded to known genes, 4,165 (44.8%) to dbESTs, whereas 2,233 (24.3%) were novel ESTs (see Fig. 1 for details). A total of 1,252, 1,065, and 1,532 EST clusters from HT, NP, and AD, respectively matched known genes, which were grouped into eight categories (endocrine-related, cell division, signaling/cell communication, cell structure/mobility, cell/organism defense, gene/protein expression, metabolism, and unclassified). The genes expressed in three tissues by the highest proportion were those participating in gene/protein expression (16.4%, 23.5%, and 20.4% from HT, NP, and AD, respectively) and metabolism (16.5%, 14.0%, and 17.0% from HT, NP, and AD, respectively). Proportions of the genes related to endocrinology were close in the three tissues (11.8%, 12.9%, and 10.7% from HT, NP, and AD, respectively). However, compared with some other tissues such as $CD34^+$ hematopoietic precursors (10.6%) (7) and cardiovascular system (12.1%) (10), the percentages of genes involved in signaling/cell communication were higher in the HPA axis (18.7%, 18.2%, and 15.3% from HT, NP, and AD, respectively) (P values all <0.01). The gene expression profiles reflected well the functional characteristics of the three levels in the HPA axis, because most of the 20 genes with highest expression showed statistical difference in terms of tissue distribution, including a group of tissue-specific functional markers (Table 1). Except for those genes listed in Table 1, some neuroendocrine-specific genes such as neuroendocrine-specific protein A, pro-melanin-concentrating hormone, prolactin receptor-associated protein, CART, OB-R gene-related protein, and leptin receptor short form were hit in cDNA libraries from HT. In addition, some cytokines and hormones such as vascular endothelial growth factor, c-sis/platelet-derived growth factor (PDGF) 2, PDGF A, insulin-like growth factor 1, adrenomedullin, angiotensinogen, basic fibroblast growth factor, cholecystokinin, endothelin 3, lens epithelium-derived growth factor, transforming growth factor α, and ALK-3 (members of the transforming growth factor-β cytokine family) were encountered in ESTs data from HT. In ESTs from NP, some hormones and cytokines including luteinizing hormone, TSH, pro-galanin, CART, GH (20 kDa), HE1 (a major secretory protein of the human epididymis), growth factor FIGF, angiopoietin-2, amyloid protein (AD-AP), tumor necrosis factor, epithelins 1 and 2, endothelin 3, kidney epidermal growth factor, preprocortistatin (Cort), and preproenkephalin also were hit. Unexpectedly, steroidogenic enzyme/protein gene transcription such as type II β-hydroxysteroid dehydrogenease/5-δ-4-δ isomerase, 17β-hydroxysteroid dehydrogenase, steroidogenic acute regulatory protein, and prepro form of corticotropin-releasing factor were expressed in both HT and AD. It may be interesting to note that transcription factors with tissues specificity were found, such as specific zinc finger protein 2 (A1–5) in HT, Pit1

Table 2. List of the full-length cDNA of HPA axis and their homology with some model organisms in the present work

Accesion number	Gene name	ORF (bp)	ORF (aa)	Chromosome localization	E S C D M	Accesion number	Gene name	ORF (bp)	ORF (aa)	Chromosome localization	E S C D M
AF112200	ubiquinone oxidoreductase complex	411	137	D3S1309-D3S3694		AF125392	insulin-induced growth-response protein	714	238	D2S121-D2S110	
AF112201	hNP25	846	282			AF157322	Human nucleoporin p54 protein	1515	505	D4S392-D4S2947	
AF112202	Human prenylated rab acceptor 1	555	185	D19S425-D19S418		AF125394	geranylgeranyl pyrophosphate synthase	900	300	D1S459-D1S304	
AF112204	hSFD isoform	1449	483	D8S517-D8S509		AF110774	AD-001	372	124	D11S913-D11S916	
AF112215	Human SKD1 protein	1314	438	D16S3031-D16S3139		AF110775	AD-002	684	228	D11S1311-D11S917	
AF112216	Human UMP-CMP kinase	588	196	D1S2843-D1S417		AF110776	AD-003	657	219	9q34.2-34.3	
AF112218	Human 3-7 gene product	2226	742			AF110777	AD-004	516	172	D5S647-D5S637	
AF112219	Human esterase D	846	282	D13S328-D13S168		AF110778	AD-005	948	316	16p11.1	
AF112220	Human cyclic ARPP-21	267	89	D3S1609-D3S1260		AF112203	Human px19	624	208	D5S408-qTEL	
AF113122	Human AP-2rep transcription factor	807	269	D13S1260-D13S152		AF112206	Human rab-14	645	215	D8S507-D8S510	
AF113125	Human E-1 enzyme (masA)	783	261	D4S2947-D4S400		AF112207	Human eIF-2B d subunit	1569	523	D2S165-D2S352	
AF113126	Human PEMT	708	236	D17S922-D17S798		AF112208	13 KD differentiation-associated protein	435	145		
AF113129	vacuolar ATPase (isoform VA68)	1851	617	D3S1610-D3S1278		AF112209	Human Ena-VASP like protein	1248	416	14q32.2	
AF113534	Human HP1-74 protein	1662	554			AF112210	Human NST-1	1527	509		
AF113535	Human maternal transcript Maid	696	232	15p14		AF112211	Human p47	1113	371	D20S117-D20S113	
AF113536	Human MO25	1023	341	D2S2158-D2S125		AF112212	Human peroxisomal-like protein	486	162		
AF113537	Human P19	513	171	D5S498-D5S408		AF112213	Human putative Rab5-interacting protein	387	129	D20S106-D20S107	
AF113538	Huamn RIP110	1047	349	D5S498-D5S408		AF112214	Human 50s ribosomal protein l13	516	172	8q23.3-24.1	
AF113539	HT001	960	320	D3S3606-D3S3554		AF112217	human UCCR complex protein	219	73	D22S1144-D22S280	
AF113540	HT002	669	223	8q24.3		AF112221	Human rap2 interacting protein 8	1149	383	D4S392-D4S2947	
AF117235	Human glycine rich protein	417	139			AF112222	Human nuclear protein SDK3	2151	717	D14S70-D14S281	
AF117236	Human matrin 3	2544	848	D5S500-D5S436		AF113123	Human carbonyl reductase	732	244	D17S784-qTEL	
AF126021	Human BCR associated protein 37	897	299	12p13		AF113124	Human FEZ2	960	320	D2S367-D2S2230	
AF136977	Human SM-x5	285	95	D6S1558-D6S1616		AF113127	Human S1R protein	714	238	D12S83-D12S350	
AF155660	Human mitochondrial solute carrier	561	187	D8S258-D8S560		AF113128	Human steroid 5-α-reductase isoform	633	211	D5S678-D5S675	
AF157323	Human protein homolog with p45skp2	2022	674	D4S412-D4S1601		AF117229	AD-009	852	284	12q24	
AF157325	hSREBP-3	1371	457	D12S333-D12S325		AF117230	AD-006	1008	336	top of Chr.3	
AF164790	oxidoreductase UCPA	735	245			AF117231	AD-007	918	306	20q13	
AF183423	Human reticulocabin precursor isolog	984	328	D19S425-D19S418		AF117232	AD-008	486	162	19p13.3	
AF184213	HT006	609	203	D10S540-D10S597		AF117233	Human znf-xp homologue	984	328	D7S2450-D7S550	
AF183422	human mRNA for RNA helicase	1434	478			AF117234	Human flotillin-isoform	759	253	D6S1558-D6S1616	
AF220193	HT007	1014	338	pTEL-D6S1640		AF117237	Human prefoldin subunit 2	462	154	D8S257-D8S508	
AF183414	human eHSIF 2a kinase	1890	630			AF119662	Human E46	1425	475	D22S272-D22S274	
AF183409	human transmembrane protein PT27	972	324	D4S1619-D4S392		AF119663	Human G-protein d12 subunit	216	72	D1S203-D1S2865	
AF183420	peptide methionine sulfoxide reductase	693	231			AF119664	Human HCNGP	924	308	D17S1352-D17S785	
AF183427	HT005	1443	481	D3S3582-D3S1588		AF119665	Human inorganic pyrophosphatase	867	289	D10S210-D10S537	
AF183426	HT004	1137	379	D13S285-qTEL		AF119666	Human IRTKS	1218	406		
AF183413	dolichyl-phosphate β-glucosyltransferase	972	324	D13S267-D13S1253		AF125530	Human JM4 protein homologue	564	188	3p11.2 -12	
AF183425	HT003	351	117			AF125531	mitochondrial carrier family protein	1014	338		
AF221595	HT014	630	210	D1S2843-D1S417		AF125532	Human Muk2	1242	414	D1S2843-D1S417	
AF223466	HT015	1044	348	D8S560-D8S1820		AF125533	NADH-cytochrome b5 reductase isoform	915	305	D15S146-D15S117	
AF220193	HT-007	657	219			AF125534	Human NFE2-related factor 1	1197	399	7p15	
AF220182	HT-008	687	229			AF125535	Human pp21 homolog	312	104	DXS990-DXS1059	
AF220183	HT-009	564	188	pTEL-D10S558		AF126020	Human bcr-associated protein 29	726	242	7q21.3	
AF220184	HT-010	1191	397	D2S2257-D2S115		AF126023	Human stromal cell protein	663	221	D22S272-D22S274	
AF220185	HT-011	1080	360	D10S531-D10S209		AF126024	Human stromal cell protein isoform	537	179	D22S272-D22S274	
AF220186	HT-012	420	140	D6S1665-D6S1660		AF136970	Human sarcosine oxidase (SOX)	1296	432	D12S99-D12S358	
AF220187	HT-013	930	310			AF136971	AD-011	420	140	D14S76-D14S270	
AF220189	HT-Bex2	375	125			AF136972	Human protein phosphatase 2C b	1161	387	D2S119-D2S337	
AF220192	HT-HCDase	858	286			AF136974	Human ras-related protein (rab18)	618	206	D10S197-D10S588	
AF220190	HT-ARP11	1251	417	D14S274-D14S1038		AF136975	Human RAMP4	198	66	3q24.3	
AF220191	HT-SMNP1	783	261			AF136976	Human sec13 homolog protein	1092	364	D18S71-D18S478	
AF220188	HT-TMP	972	324	D4S1619-D4S392		AF136978	AD-010	942	314	10q22	
AF212253	HCOBP	1128	376	D11S1307-D11S921,		AF150732	Human protein tyrosine phosphatase	2397	799	D1S418-D1S514	
AF077049	Human lambda-crystallin	891	297			AF150733	AD-014	165	55	D5S504-D5S677	
AF070666	Human KRAB homolog	282	94	D19S425-D19S418		AF150734	Human PC326 protein	1599	533	D1S196-D1S210	
AF077046	Human GEF-2	351	117	D16S512-D16S515		AF150735	AD-013	570	190		
AF077052	Human sui1iso1 homolog	339	113	D3S1260-D3S3582		AF160212	Human VAP-33 homolog	729	243	D20S183-D20S173	
AF077054	Human unr protein	2310	770	D1S418-D1S514		AF155651	Human P21 activated kinase-3	1632	544	Xq22.3-23	
AF078845	Human hypothetical 16.7 kd protein	453	151			AF155569	Human MCBP E isolog	564	188	D5S628-D5S474	
AF078846	Human brain-specific protein	522	174	D16S3031-D16S3139		AF155652	Human PCMF	1143	381		
AF078849	Human dynein light chain-A	1569	523	D3S1609-D3S1260		AF157316	Human kinesin heavy chain homolog	504	168	D5S500-D5S436	
AF078851	Human secretogranin III (SgIII)	1404	468	D15S146-D15S117		AF155653	Human 50s ribosomal protein l36	309	103		
AF078853	NPD001	411	137	D16S407-D16S414		AF155654	Human ribosomal protein Sx	297	99	D4S412-D4S1601	
AF078854	NPD002	516	172	D3S1267-D3S1269		AF155656	Human zinc finger protein ZNFshge	1233	411		
AF078855	NPD003	909	303	D8S1820-D8S505		AF157326	Human TIP 120 protein	3690	1230	D12S83-D12S350	
AF226732	NPD007	1302	434			AF155648	Human putative zinc finger protein	3636	1212		
AF223467	NPD008	618	206	16p13		AF155661	Human pyruvate dehydrogenase	1722	574	D8S270-D8S257	
AF078865	Human RNA-binding protein	828	276			AF155649	Human 15kd protein	420	140	D6S1665-D6S1660	
AF078866	Human SURF-4 protein	807	269	D9S159-qTEL		AF155662	Human 16.7kd protein	432	144	D12S366-D12S340	
AF078867	Human SURF-4 protein isolog	477	159	D9S159-qTEL		AF155655	AD-016	522	174		
AF092128	putative transmembrane protein E3-16	798	266	D13S168-D13S153		AF155657	Human 17.9KD protein	480	160	D6S1558-D6S1616	
AF092131	Human 51 kDa NADH dehydrogenase	1392	464	D11S913-D11S916		AF155650	Human goliath protein	828	276		
AF092137	Human FK506-binding protein	666	222			AF157317	AD-015	612	204	D15S157-qTEL	
AF092139	Human vesicle transport-related protein	1920	640	D14S275-D14S262		AF155658	Human 55kd protein	1515	505	D22S1144-D22S280	
AF093680	Human TfIIB	579	193	D1S306-D1S491		AF157327	AD-012	1749	583	D9S165-D9S1874	
AF100740	Human ARF-like protein 5	537	179	D20S183-D20S173		AF157318	AD-017	1113	371	D3S3582-D3S1588	
AF100741	Human vacuolar H-ATPase subunit D	741	247	D14S63-D14S1069		AF157319	AD-018	729	243		
AF100742	Human M-phase phosphoprotein homolog	1254	418	D5S504-D5S677		AF157320	AD-019	948	316		
AF100743	Human NADH-Ubiquinone reductase	789	263	D11S1357-D11S136		AF164796	Human MLRQ subunit of NADH	261	87	D12S325-D12S1691	
AF100744	Human hypothetical 19.5 KD protien	1101	367	D18S459-D18S482		AF183416	Human HGR74 homolog	375	125	DXS990-DXS1059	
AF100751	Human FK506-binding protein isoform	777	259			AF164795	Human SRP JANUS-A	375	125	D9S159-qTEL	
AF100752	transitional endoplasmic reticulum ATPase	2418	806	D9S1874-D9S273		AF157324	Human Rer1 protein	642	214		
AF100753	ancient ubiquitous 46 kDa protein AUP1	1230	410	D2S145-D2S286		AF157321	Human 30 Kd protein	783	261	D3S3591-D3S1283	
AF100754	ancient ubiquitous protein AUP1 isoform	1428	476	D2S145-D2S286		AF183424	Human SCO1 & SCO2 protein	903	301	D17S804-D17S799	
AF100755	Human homeobox Gene CLE-7	732	244	D2S145-D2S286		AF160214	Human SIR 2 homolog	1119	373	D19S425-D19S418	
AF100756	Human coat protein gamma-cop	2622	874	D3S3606-D3S3554		AF160215	ubiquitin-conjugating enzyme E2	591	197	D1S2622-D1S306	
AF100757	Human COP9 complex subunit 4	1215	405	D4S400-D4S1534		AF164799	Human 28 kd protien	756	252		
AF100759	transmembrane 4 protein homolog	612	204			AF160213	Human emp70 precusor isolog	1758	586		
AF100760	Human MNUDC protein	993	331	D2S156-D2S376		AF164791	Human 38.3 kd protein	1014	338	18q12.1	
AF100762	Human TRIP15	1329	443	D15S146-D15S117		AF164792	Human 47 kd protein	1242	414		
AF106681	Human ras-related GTP-binding protein	600	200	D2S171-D2S165		AF164797	Human ribosomal protein L17	525	175	D11S1318-D11S909	
AF106682	Human spindlin (Spin)	741	247	D9S283-D9S1842		AF183428	Human 28.4 kd protein	753	251	D11S1357-D11S913	
AF106683	Human WSB-1	1263	421			AF164793	AD-020	222	74		
AF106684	Human WSB-1 isoform	657	219			AF164794	Human Dil33-homolog	1359	453	D6S1712-D6S407	
AF106685	Human myelin gene expression factor	1641	547	D1S2843-D1S417		AF164798	hypothetical transmembrane protein	246	82	12q24.2	
AF110644	Human GH-variant	513	171	D17S794-D17S795		AF211481	human 43.8kd protein	1182	394	D4S1579-D4S1604	
AF110645	Human p33ING1 isolog	747	249	D12S99-D12S358		AF223468	AD021	738	246		
AF110647	Human TRAP-complex g subunit	555	185	D3S1275-D3S1605		AF223469	AD022	1086	362		
AF125393	Human rab-27a isolog	639	213			AF223470	KIAA0971 isoform	2130	710		

*The different color boxes represent disntinct homology with model organisms according to their identity rate by over 30 amino acid residues. Red box indicate 80-100% identity rate; yellow box: 60-89%; green box: 40-59%; blue box: 25-39%; and black box: <25%. E:Escherichia coli; S:Saccharomyces cerevisiae; C:Caenorhabditis elegans; D:Drosophila melanogaster; M:Mus musculus. Vacancy in chromosome localization indicates that the chromosome mapping of that gene failed by electronic PCR or RH method.

Table 3. Variant splicing pattern of some genes

Gene name	Splicing pattern	cDNA, bp	Amino acid	Characteristics of splicing
Human stromal cell protein	Classical	1316	221	Atypical, aa-gc
	Isoform	1183	179	
Human steroid 5-α-reductase	Classical	2222	260	Atypical, gt-cg spanning exons 1 and 2
	Isoform	1178	211	
Human FK506-binding protein	Classical	1231	222	Atypical, gt-cg
(FKBP23)	Isoform	1067	259	
Human ancient ubiquitous protein	Classical	1466	410	Typical gt-ag
	Isoform	1664	476	
Human 17.9-kDa protein	Classical	None	113	Typical, gt-ag, increase 1 new exon
	Isoform	1121	160	
Human growth hormone	Classical	821	217	Atypical, tt-ag, spanning 2 exons
	Isoform	642	171	
Human flotillin	Classical	1698	427	Atypical, gg-ga
	Isoform	1493	253	
Human WSB-1	Classical	1925	421	??-ag
	Isoform	1657	219	
Human SURF-4 protein	Classical	1557	269	Atypical, at-ac
	Isoform	1218	159	
Human small GTP-binding protein	Classical	2496	221	Atypical, ga-ag
(Rab27) Isoform	Isoform	1156	213	
hTRIP15-iso	Classical	1963	443	Typical, gt-ag
	Isoform	1984	450	

and zinc finger protein 74 in NP, and ZNF185 in AD. Further study of these gene functions will be beneficial to the disclosure of the specific function of each tissue. Notably, the expression of GTP-binding protein α (stimulatory) was high in these three tissues, in agreement with the concept that the HPA axis is centrally positioned in neuroendocrine signaling. The most highly expressed classical NP hormone genes, as estimated by the number of ESTs, were GH (4.182% of total ESTs from NP), followed by prolactin (2.853%), POMC (0.595%), follicle-stimulating hormone α (0.499%), and TSH (0.06%). This order was in parallel to the output of the hormones from the NP during a period of 24 h (Fig. 1*B*). To confirm the order of these hormones in NP, real-time PCR performed for GH, POMC, and TSH normalized by comparison with respective glyceraldehyde-3-phosphate dehydrogenase curve-generated gene expression levels (32×10^5, 5.9×10^5, and 2.4×10^5 copies/100 ng RNA, respectively, P value all < 0.001), which are proportional to those estimated by EST copy number. In the EST profile of AD, we found the expression of steroidogenic acute regulatory protein, a vehicle of cholesterol related to steroidogenesis, and all 12 key enzymes involved in the biosynthesis of steroid hormones (Fig. 1*C*). Surprisingly, when clusters from HT, NP, and AD were compared with each other, only relatively minor portions of clusters were expressed in all three tissues (6.0%, 1.5%, and none for known genes, dbESTs, and novel ESTs, respectively), whereas the majority of clusters, especially in novel ESTs, were found only in one of the three given tissues. These data suggest a significantly functional divergence at distinct levels of the HPA axis and an enrichment of tissue-specific genes in novel ESTs (Fig. 1*D*). All of these data may contribute to the understanding of the neuroendocrine regulation of human life.

Full-Length cDNA Cloning from the HPA Axis. Table 2 shows all 200 new full-length cDNAs, with 97 cloned from AD, 49 from NP, and the remaining 54 from HT. Among these novel genes, the majority, 153 (76.5%) and 139 (69.5%), contain 500–2,000 bp in cDNA length and encode 100–400 amino acid residues deduced from their encoding frames. Some genes might be new members of certain gene families, such as zinc finger family, leucine zipper family, Ras-related protein family, and vesicle-associated membrane protein family, according to their homology with known genes and domains. In addition, some genes are very conserved in the progress of evolution because the proteins derived from these genes exhibit similar primary structure with organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus*, and even some plants or viruses (Table 2). These novel genes might be involved in critical biological process according to their homology to known genes with established functions, such as secretory proteins, signal transduction molecules, enzyme and genes involved in secretion of vesicle, development, differentiation, and transcriptional gene expression.

The chromosomal locations of 154 nove1 genes were determined, of which 127 were located by using UniGene database information and 27 by RH mapping. The remaining 46 new genes could not be chromosome mapped by either of the above methods.

The Discovery of Novel Splicing Patterns of Known Genes. The EST data acquired from the HPA axis also provide useful information on alternative splicing of known genes, as shown in Table 3 for 11 genes. Of note, among 302 ESTs for the GH gene in NP, a splicing form lacking 138 bp was discovered, which might result in a new isoform of the hormone missing 46 aa (amino acids 92–137). Comparison of the mRNA sequence with the corresponding genomic region suggests a possible splicing within both exons 4 and 5 of the gene, deleting 11 aa from exon 4 and 35 aa from exon 5. This splicing, however, seems not to fit the classical gt-ag rule but uses atypical splicing donor tt-ag. To confirm the existence of this splicing pattern in NP, primers capable of distinguishing classical GH gene from its new isoform were synthesized, and two bands of expected sizes were amplified in NP by RT-PCR. Both bands then were subcloned into plasmid pGEM-T and sequenced, and their sequences were shown to correspond exactly to our original finding (not shown). New splicing forms of FKBP and WSB genes also were confirmed by the above methods.

Discussion

With the growing interest in research into human genome and the advent of high-throughput DNA sequencing technologies, Adams *et al*. (12) carried out the measurement of cDNAs on a large scale, and then thousands of human ESTs were produced

from different organs, tissues, or cell types. However, not all tissue/cell types are equally represented for gene expression profiles in dbEST. Concerning the endocrine system, very few ESTs could be found from the HT and NP whereas there were only about 900 ESTs from the normal adrenal tissues until the end of 1999. The present study thus aimed at characterization of the gene expression profile of the HPA axis by using EST analysis and molecular cloning of full-length cDNA of novel genes identified, to lay 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. Totally, we have produced 25,973 ESTs from the HPA axis and have basically set up the expression spectrum for known genes in each tissue. Known genes, known ESTs, and novel ESTs make up 79.5% of all ESTs, whereas repetitive elements and mtDNA sequences accounted for 21.5%. This distribution is quite close to those previously reported in other tissues (10). The fact that a relatively low proportion (13.3%) of novel ESTs is from the HPA axis suggests that the majority of human genes may have already been labeled in dbEST (13).

Because there was a satisfactory representation of ESTs generated from HT, NP, and AD, the gene expression of the HPA axis could be analyzed in terms of both patterns and levels. The overall expression profile from the three tissues corresponds very well to the known functions, with neuroendocrine markers/hypothalamic hormones/receptors, classical NP hormones, and enzymes involved in the biosynthesis of steroid hormones as the most highly expressed genes in HT, NP, and AD, respectively. In addition, the three tissues showed relatively distinct gene expression profiles, as there was a relatively low proportion (23.3%) of genes with expression occurring in two or three tissues. In the meantime, however, among genes with overlapped expression, especially those expressed at the three levels of the HPA axis, some important endocrine-related genes were found, which may extend our knowledge about endocrine functions and/or regulations. For example, CART, previously considered as a key hypothalamic hormone controlling appetite and energy metabolism, showed expression in the tissues of HT, NP, and AD. The highest expression of CART (seven EST copies), however, was found in NP instead of HT, suggesting that CART could also be a NP hormone. Another interesting finding was the expression of both corticotropin-releasing hormone (CRH) and urocortin in AD. It was reported that the receptor for CRH as well as the downstream POMC genes could be expressed by AD whereas urocortin recently was found to be able to stimulate secretion of corticotropin through CRH receptor (14), with an even stronger effect than CRH (15). It is thus possible that there be local paracrine/autocrine pathways, with a corresponding regulatory network of feedback for CRH and urocortin in AD. Another example is the hypothalamic expression of adrenomedullin, a secreted protein isolated from human pheochromocytoma (16), but also expressed in the zona glomerulosa of the cortex of AD, which increases the secretion of aldosterone through paracrine mechanism (17).

Recently, full-length cDNA cloning and sequencing have become major tasks in the Human Genome Project's next 5-year plan (18, 19). Mass production of ESTs combined with bioinformatics analysis is a milestone in the discovery of novel transcription units. Based on the known or novel ESTs identified in the HPA axis, and taking advantage of the full-length cDNA-enriched libraries constructed in our work, the UniGene information in public databases, and the available 5′ rapid amplification of cDNA ends PCR technology, we cloned 200 full-length cDNA of novel genes. The tools of bioinformatics not only help to clone novel genes through dbEST assembly, but also provide important clues to the functions of novel genes through comparison of homology to known genes with established function and those genes from model organisms. For example, among 200 novel genes, we have found at least three that may participate in the signal transduction process of the receptors with tyrosine kinase activity. Human UNR gene was cloned from NP and its homology with the known gene in rat is as high as 93%, the latter being located in tandem to *N*-ras gene with a distance of only 150 bp (20). Recently, the UNR gene product was found to inhibit the expression of *N*-ras (21), which is an important molecule in the signaling of receptor tyrosine kinase. Another novel gene, human insulin receptor tyrosine kinase substrate cloned from AD, shows 40.9% homology with the insulin receptor tyrosine kinase 53-kDa substrate gene from Crietinase gen. SP. The known 53-kDa substrate contains phosphorylation site of receptor tyrosine kinase and participates in the process of signal transduction of insulin receptor (22). The third novel gene cloned from the NP, designated as human insulin-induced growth response protein 2, shows 83% homology to human insulin-induced growth response protein 1 at the protein level, whereas the latter plays a role in growth and differentiation of tissues involved in metabolic control (23).

Moreover, the estimated total number of human genes may need modification, considering the discrepancy between EST data and genomic DNA data generated from chromosomes 21 and 22 (24–26). There might be overestimation of the number of genes according to EST clusters because many genes may yield different splicing transcripts. Actually, in our study, some known genes and new genes indeed present alternative splicing patterns.

This work was supported in part by the National Key Basic Research (973) program of China, Chinese High Tech Program (863), National Natural Science Foundation of China, Shanghai Commission for Science and Technology, and SmithKline Beecham pharmaceuticals.

1. Reichlin, S. (1993) *N. Engl. J. Med.* **329,** 1246–1253.
2. Kristensen, P., Judge, M. E., Thim, L., Ribel, U., Christjansen, K. N., Wulff, B. S., Clausen, J. T., Jensen, P. B., Madsen, O. D., Vrang, N., *et al*. (1998) *Nature (London)* **393,** 72–76.
3. Arimura, A. (1998) *Jpn. J. Physiol.* **48,** 301–331.
4. Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R. M., Tanaka, H., Williams, S. C., Richardson, J. A., Kozlowski, G. P., Wilson, S., *et al*. (1998) *Cell* **92,** 573–585.
5. Small, C. J., Todd, J. F., Ghatei, M., Smith, D. M. & Bloom, S. R. (1998) *Regul. Pept.* **75–76,** 301–307.
6. Kanyicska, B., Lerant, A. & Freeman, M. E. (1998) *Endocrinology* **139,** 5164–5173.
7. Mao, M., Fu, G., Wu, J. S., Zhang, Q. H., Zhou, J., Kan, L. X., Huang, Q. H., He, K. L., Gu, B. W., Han, Z. G., *et al*. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8175–8180.
8. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., *et al*. (1995) *Nature (London)* **377,** 3–174.
9. Cox, D. R., Burmeister, M., Price, E. R., Kim, S. & Myers, R. M. (1990) *Science* **250,** 245–250.
10. Liew, C. C., Hwang, D. M., Fung, Y. W., Laurenssen, C., Cukerman, E., Tsui, S. & Lee C. Y. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10645–10649.
11. Audic, S. & Claverie, J. M. (1997) *Genome Res.* **7,** 986–995.
12. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., *et al*. (1991) *Science* **252,** 1651–1656.
13. Schuler, G. D., Boguski, M. S., Hudson, T. J., Hui, L., Ma, J., Castle, A. B., Wu, X., Silva, J., Nusbaum, H. C., Birren, B. B., *et al*. (1996) *Science* **274,** 540–546.
14. Vaughan, J., Donaldson, C., Bittencourt, J., Perrin, M. H., Lewis, K., Sutton, S., Chan, R., Turnbull, A. V., Lovejoy, D., Rivier, C., *et al*. (1995) *Nature (London)* **378,** 287–292.
15. Turnbull, A. V., Vaughan, J., Rivier, J. E., Vale, W. W. & Rivier, C. (1999) *Endocrinology* **140,** 71–78.
16. Kitamura, K., Kangawa, K., Kawamoto, M., Ichiki, Y., Nakamura, S., Matsuo, H. & Eto, T. (1993) *Biochem. Biophys. Res. Commun.* **192,** 553–560.
17. Kapas, S., Martinez, A., Cuttitta, F. & Hinson, J. P. (1998) *J. Endocrinol.* **156,** 477–484.
18. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. & Walters, L. (1998) *Science* **282,** 682–689.
19. Strausberg, R. L., Feingold, E. A., Klausner, R. D. & Collins, F. S. (1999) *Science* **286,** 455–457.
20. Boussadia, O., Jacquemin-Sablon, H. & Dautry, F. (1993) *Biochim. Biophys. Acta* **1172,** 64–72.
21. Boussadia, O., Amiot, F., Cases, S., Triqueneaux, G., Jacquemin-Sablon, H. & Dautry, F. (1997) *FEBS Lett.* **420,** 20–24.
22. Yeh, T. C., Ogawa, W., Danielsen, A. G. & Roth, R. A. (1996) *J. Biol. Chem.* **271,** 2921–2928.
23. Peng, Y., Schwarz, E. J., Lazar, M. A., Genin, A., Spinner, N. B. & Taub, R. (1997) *Genomics* **43,** 278–284.
24. Ewing, B. & Green, P. (2000) *Nat. Genet.* **25,** 232–234.
25. Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S. L. & Quackenbush, J. (2000) *Nat. Genet.* **25,** 239–240.
26. Dunham, I., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., Ainscough, R., Almeida, J. P., Babbage, A., *et al*. (2000) *Nature (London)* **402,** 489–495.