Gene Expression Profiling in Human Fetal Liver and Identification of Tissue- and Developmental-Stage-Specific Genes through Compiled Expression Profiles and Efficient Cloning of Full-Length cDNAs

Yongtao Yu, Chenggang Zhang, Gangqiao Zhou, Songfeng Wu, Xianghu Qu, Handong Wei, Guichun Xing, Chunna Dong, Yun Zhai, Jinghong Wan, Shuguang Ouyang, Li Li, Shaowen Zhang, Kaixin Zhou, Yinan Zhang, Chutse Wu, and Fuchu He[1]

Department of Genomics and Proteomics, Beijing Institute of Radiation Medicine, Chinese National Human Genome Center at Beijing, Beijing 100850, China

Fetal liver intriguingly consists of hepatic parenchymal cells and hematopoietic stem/progenitor cells. Human fetal liver aged 22 wk of gestation (HFL22w) corresponds to the turning point between immigration and emigration of the hematopoietic system. To gain further molecular insight into its developmental and functional characteristics, HFL22w was studied by generating expressed sequence tags (ESTs) and by analyzing the compiled expression profiles of liver at different developmental stages. A total of 13,077 ESTs were sequenced from a 3′-directed cDNA library of HFL22w, and classified as follows: 5819 (44.5%) matched to known genes; 5460 (41.8%) exhibited no significant homology to known genes; and the remaining 1798 (13.7%) were genomic sequences of unknown function, mitochondrial genomic sequences, or repetitive sequences. Integration of ESTs of known human genes generated a profile including 1660 genes that could be divided into 15 gene categories according to their functions. Genes related to general housekeeping, ESTs associated with hematopoiesis, and liver-specific genes were highly expressed. Genes for signal transduction and those associated with diseases, abnormalities, or transcription regulation were also noticeably active. By comparing the expression profiles, we identified six gene groups that were associated with different developmental stages of human fetal liver, tumorigenesis, different physiological functions of Itoh cells against the other types of hepatic cells, and fetal hematopoiesis. The gene expression profile therefore reflected the unique functional characteristics of HFL22w remarkably. Meanwhile, 110 full-length cDNAs of novel genes were cloned and sequenced. These novel genes might contribute to our understanding of the unique functional characteristics of the human fetal liver at 22 wk.

[The sequence data described in this paper have been submitted to the GenBank data library under the accession nos. listed in Table 6 herein]

The liver is the largest gland in the human body. In addition to secreting bile, it functions in the metabolism of carbohydrates, fats, proteins, vitamins, and hormones. Hepatocytes undergo distinct phases of differentiation as they arise from the gut endoderm, coalesce to form the liver, and mature by birth.

Hematopoiesis occurs at three different primary sites during human embryonic and fetal development. It begins between day 15 and day 18 in the blood islands of the yolk sac. After 6 wk, hematopoietic stem cells (HSCs) migrate via the bloodstream to fetal liver (FL) and spleen, where erythropoiesis still predominates, but myeloid ontogenesis is also beginning. During the 20th wk of gestation, bone marrow hematopoiesis begins to occur, then becomes more and more myelopoietic, and finally represents the entire blood cell production. At the same time, hepatic and splenic hematopoietic activity decrease and disappear (Migliaccio et al. 1986; Tavassoli 1991; Huang and Auerbach 1993; Godin et al. 1995). The fetal liver at 22 wk of gestation (HFL22w) is a major site of fetal hematopoiesis in man, and is at the critical turning point between immigration and emigration of the hematopoietic system. Therefore, the unique characteristics of the fetal liver at this stage are worthy of investigation.

The diverse functions and complex regulation of HFL22w might be largely determined by well-regulated gene expression. Indeed, a number of important growth factors

[1]Corresponding author.
E-MAIL hefc@nic.bmi.ac.cn; FAX 86-10-68214653.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.175501.

11:1392–1403 ©2001 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/01 $5.00; www.genome.org

(Fausto 1991), transcription factors (Dabeva et al. 1995), and protein transportation regulators (Zhang et al. 2000) have been identified from HFL22w over the last two decades. Apart from classical factors, we recently cloned hepatopoietin (HPO) (Wang et al. 1999), a novel human hepatotrophic growth factor. It specifically stimulates proliferation of cultured primary hepatocytes in vitro, liver regeneration after liver partial hepatectomy in vivo, and autonomous growth of hepatoma cells by stimulation of the mitogen-activated protein kinase cascade and tyrosine phosphorylation of the epidermal growth factor receptor (Wang et al. 1999; Li et al. 2000). However, there are many unknown regulators and molecular signaling mechanisms, as well as the genetic control of fetal liver development to be explored. The mechanisms of migration, localization, and regulation of hematopoiesis at different stages of ontogeny are not well understood either.

The identification of genes of a given cell type, tissue, or corresponding to a pathological state that confer developmental or functional specificity will provide valuable molecular insight for the study of biological phenomena and cellular physiology. Like any specified tissue and cell population in the human body, the biological features of human fetal liver might be determined largely at the level of gene expression. Single-pass, partial sequencing of randomly selected cDNA clones from cDNA libraries to generate expressed sequence tags (ESTs) (Adams et al. 1991), combined with bioinformatics analysis, has proved useful for the discovery of novel genes (Adams et al. 1995), the characterization of gene function (Papadopoulos et al. 1994), the differential and quantitative analysis of expression patterns (Okubo et al. 1992), and for the evaluation of the gene expression profile in a given tissue (Adams et al. 1991, 1992; Okubo et al. 1992; Liew et al. 1994; Mao et al. 1998; Ryo et al. 1998; Sterky et al. 1998). It is obvious that the establishment of a detailed catalog of genes expressed in HFL22w, the discovery of novel genes from HFL22w, and identification of tissue- and developmental-stage-specific genes through compiled gene expression profiles will certainly facilitate our understanding of the mechanisms of coexistence of hepatic and hematopoietic systems in fetal liver and the regulation network of immigration/emigration of the hematopoietic system of fetal liver.

The present report is on the establishment of a gene expression profile of HFL22w based on the analysis of 13,077 ESTs as well as preliminary results of comparison of this expression profile with those of 10 different human cells or tissues associated with hepatic or hematopoietic systems, which are two major functional features of human fetal liver at the developmental stage of 22 wk of gestation. As a result, we found some tissue-specific and developmental-stage-specific gene groups that are likely to play important roles in some definite functional features.

RESULTS

cDNA Sequencing and General Data of ESTs from HFL22w

The HFL22w cDNA library had average insert sizes of 1.0–1.5 kb. By using automatic procedures for DNA sequencing, 14,400 clones were randomly picked up and sequenced partially from one end by using T7 or SP6 primer. Of them, 743 were considered trash, defined as sequences from bacterial DNA, sequences from primer polymers, sequences containing >1% of ambiguous bases (*N*), or sequences shorter than 100 bp; the other 13,077 sequences were considered good ones. The rate of successful sequences was therefore 90.8% and the average read-length for good sequences is 555 bp, which, to our knowledge, is among the best in the literature. Analysis of the 13,077 ESTs of satisfactory quality revealed three groups of sequences. Group I (5819 ESTs, 44.5%) matched to known genes in the GenBank nonredundant database and were considered labels of known functional genes, among which 5666 ESTs (43.3%) matched to human genes and the other 153 ESTs (1.2%) matched to previously described genes of other species. Group II (5460 ESTs, 41.8%) exhibited no significant homology to known genes, and 18.8% (1025 ESTs) of these overlapped EST sequences in the public database (dbEST). Group III (1798 ESTs, 13.7%) were genomic sequences of unknown function, mitochondrial DNA, or repetitive sequences.

Gene Expression Profile of Active Genes in HFL22w

A catalog of genes expressed in HFL22w was established by

Table 1. ESTs Distribution of HFL22w by Functional Categories

Serial no.	Gene categories	Gene (%)	EST (%)
I	Cell defense and homeostasis	29 (1.7)	193 (3.4)
II	Cell division and regulation	59 (3.6)	113 (2.0)
III	Cytokines and hormones	12 (0.7)	72 (1.3)
IV	Cytoskeleton	73 (4.4)	187 (3.3)
V	Development	40 (2.4)	122 (2.2)
VI	Genes associated with diseases or abnormalities	131 (7.9)	279 (4.9)
VII	Gene/protein expression	231 (13.9)	701 (12.4)
VIII	Hematopoiesis	185 (11.1)	1260 (22.2)
IX	Liver and lipoproteins	69 (4.2)	1047 (18.5)
X	Metabolism	296 (17.8)	599 (10.6)
XI	Protease and protease inhibitor	59 (3.6)	209 (3.7)
XII	Secretory proteins	11 (0.7)	147 (2.6)
XIII	Signal transduction	139 (8.4)	221 (3.9)
XIV	Transcription related gene	95 (5.7)	168 (3.0)
XV	Unclassified	231 (13.9)	348 (6.1)
Total		1660	5666

Human identical or similar genes of HFL22w were partitioned based upon biological roles and subcellular localization.

generating a large amount of ESTs, followed by bioinformatics analysis (data available through E-mail: hefc@nic.bmi.ac.cn). The uninformative sequences of Group III were put aside, and the remaining 11,279 ESTs of Group I and II were further analyzed and assembled into 1729 and 4768 clusters, respectively. After integration of overlapping sequences or sequences corresponding to different portions of the same gene, 5666 ESTs actually represented 1660 human genes and were summarized into 15 different functional categories in Table 1. HFL22w ESTs were partitioned based upon biological roles and subcellular localization to include cell defense and homeostasis, cell division and regulation, cytokines and hormones, cytoskeleton, development, genes associated with diseases or abnormalities, gene/protein expression, hematopoiesis, liver and lipoproteins, metabolism, proteases and protease inhibitors, secretory proteins, signal transduction, transcription-related genes, and unclassified.

An expression profile of active genes in HFL22w is shown in Table 2. In the list we can see several genes with certain frequency that could be expected based on the unique features and functions of HFL22w. First, in this developmental stage of human liver, cell proliferation and differentiation need a high productive level of protein synthesis as well as general metabolism, and a large amount of energy supply and protein synthesis is occurring. The genes expressed in HFL22w in the highest proportion were functionally related to the general housekeeping responsibilities of the cells, such as general metabolism, protein synthesis, and synthesis of nucleic acids and amino acids, which includes transcripts for various enzymes involved in the central reactions of metabolism, elongation factors, and ribosomal proteins, similar to EST databases previously generated from other tissues (Adams et al. 1995). Second, HFL22w is a major site of fetal hematopoiesis and immune development. ESTs associated with hematopoiesis formed the largest group of transcripts, for example, hemoglobins, globins, complement components, prothymosin-α, angiotensinogen, T-cell cyclophilin, and glycophorin A. Third, as expected, HFL22w highly expressed liver-specific genes such as serum albumin, fibrinogens, apolipoproteins, α-fetoprotein, haptoglobin, and high density lipoprotein-binding protein. In addition, genes for signal transduction, genes associated with diseases or abnormalities, and transcription-related genes were also noticeably active. Some cytokines and hormones such as insulinlike growth factor II (IGF-2), thymosin β-4, β-10, FGFR-4, lens epithelium-derived

Table 2. Expression Profile of Frequent Genes in HFL22w

I-Cell Defense and Homeostasis (193)
- Ferritin L chain (91)
- Heart mRNA for hsp90 (30)
- Hsp90 (16)

II-Cell Division and Regulation (113)
- S-protein (23)
- HT-1080 protein (6)
- Replication protein A 32-kDa subunit (4)

III-Cytokines and Hormones (72)
- Insulin-like growth factor II (IGF-2) (55)
- Thymosin beta-4 (5)
- Barrier-to-autointegration factor (2)

IV-Cytoskeleton (187)
- Protein HC (alpha 1-microglobulin) (24)
- Alpha 2-macroglobulin (18)
- Beta-actin (11)

V-Development (122)
- Retinol binding protein (RBP) (69)
- Mammary-derived growth inhibitor (3)
- Putative WHSC1 protein (3)

VI-Genes Associated with Disease or Abnormalities (279)
- H19 gene (70)
- Translationally controlled tumor protein (15)
- HFREP-1 mRNA for unknown protein (9)

VII-Gene/Protein Expression (701)
- Elongation factor EF-1-alpha (62)
- Ribosomal protein S16 (37)
- XP1PO ribosomal protein S3 (rpS3) (20)

VIII-Hematopoiesis (1260)
- Hemoglobin gamma-G (HBG2) (724)
- Alpha one globin (HBA1) (68)
- Complement component 3 (C3) (30)
- Prothymosin alpha (29)
- Hemoglobin, gamma A (HBG1) (23)

IX-Liver and Lipoproteins (1047)
- Serum albumin (694)
- Fibrinogen beta-chain (41)
- Fibrinogen gamma chain (38)
- Apolipoprotein B100 (30)
- Apolipoprotein AII (25)
- Albumin (ALB) (24)
- Apolipoprotein AI (apo AI) (20)

X-Metabolism (599)
- mRNA clone with similarity to L-glycerol-3-gene phosphate:NAD oxidoreductase and albumin sequences. (34)
- Isolate Asn6 cytochrome b (CYTB) (31)
- NADH dehydrogenase subunit 2 (31)

XI-Protease and Protease Inhibitor (209)
- Alpha 1-antitrypsin (62)
- Antithrombin III variant (13)
- Z type alpha 1-antitrypsin gene (13)

XII-Secretory Proteins (147)
- Alpha 2-HS-glycoprotein alpha and beta chain (67)
- Transferrin (46)
- 23 kD highly basic protein (11)

XIII-Signal Transduction (221)
- Coupling protein G (s) alpha-subunit (9)
- Calmodulin (CALM1) (8)
- Guanine nucleotide exchange factor p532 (7)

XIV-Transcriptional Related Genes (168)
- DNA-binding protein A (7)
- DNA-binding protein, TAXREB107 (7)
- H3.3 gene (7)

XV-Unclassified (348)
- Novel gene (12)
- Cl1 protein (10)
- KIAA0745 (10)

Numbers in parentheses indicated the frequency of ESTs matched to these genes. The three most frequent genes of each functional category or genes whose transcripts detected twenty times or more are presented.

growth factor, megakaryocyte-stimulating factor, osteoclast-stimulating factor, and transforming growth factor (TGF) were also encountered in the EST data.

Among 13,077 clones, 10.8% belong to two abundant transcripts, hemoglobin γ-G and serum albumin (HSA), which had 724 and 694 copies, respectively. Other frequent transcripts were ferritin light chain, H19 gene, retinol binding protein (RBP), α 1 globin, and so on. Besides serum albumin, some other liver-specific genes were detected also, including fibrinogen-β, -γ, and -α chains; apolipoprotein-B100, -AII, and -AI; albumin; α-fetoprotein (AFP); high density lipoprotein-binding protein; and heptoglobin α 2 and β subunits, which are known to be abundant in the liver. Fifty-eight species of ribosomal proteins (total 283 ESTs; 2.2%) were sequenced in 13,077 randomly selected clones. Because mammalian ribosomes are reported to be composed of ~70–80 distinct proteins (Wool 1986), most of the ribosomal proteins seemed to be represented, suggesting that gene/protein expression was very active in fetal liver of this developmental stage.

Table 3 shows some of the 153 ESTs of 69 different transcript species matched to nonhuman sequences. Several ESTs were found to be similar to the genes differentially regulated during development. Some of them may turn out to be involved in signal transduction during the differentiation and proliferation of the fetal liver. Further characterization would be necessary to find out the actual biological roles of these candidates. Together with the 5460 Novel ESTs (representing 4768 EST clusters, Group II), we identified 4837 EST clusters whose biological functions were not completely known that could be good candidates for full-length cDNA cloning of novel functional genes.

Table 3. ESTs Homologous to Nonhuman Sequences

Primary accession	Homologous gene definition	Species[a]	%ID	frequency
U28494	fibringen gamma A	M.Pu.	26.6	8
U42385	fibroblast growth factor inducible gene	M.M.	75.8	4
X00227	alpha2-globin	P.T.	43.4	3
X63209	CI-ASHI mRNA for ubiquinone oxidoreductas	B.T.	76.3	2
M23159	DHFR-coamplified protein	C.H.	61.3	2
AF082526	MEK binding partner 1 (Mp1)	M.M.	61.0	2
X63678	TRAM-protein	C.F.	62.3	2
AB013357	49 kDa zinc finger protein	M.M.	61.3	1
U35776	ADP-ribosylation factor-directed GTPase activiating protein	R.N.	52.0	1
AJ005073	Alix (ALG-2-interacting protein X)	M.M.	38.8	1
U78031	apoptosis inhibitor bcl-x (bcl-x) gene	M.M.	51.9	1
Y12577	Arl4 gene	M.M.	67.0	1
AB005549	atypical PKC specific binding protein	R.N.	81.8	1
L24753	BTLF3- B.T. lactoferrin	B.T.	35.0	1
U36340	CACCC-box binding protein BKLF mRNA	M.M.	93.8	1
U59166	casein kinase 1 alpha	O.C.	81.7	1
AB000517	CDP-diacylglycerol synthase	R.N.	83.9	1
M62419	clathrin-associated protein (AP47)	M.M.	82.4	1
X75931	cleavage and polyadenylation specificity factor	B.T.	37.2	1
Z54200	DNA-binding protein	M.M.	78.9	1
L27707	eukaryotic hemin-sensitive initiation factor 2	R.N.	65.9	1
X03110	fetal A-gamma-globin gene	Ch	38.1	1
M92295	gamma-1 and gamma-2 globin	G.G.	43.6	1
AF061582	heterogeneous nuclear ribonucleoprotein C (hnRNPC)	O.C.	31.2	1
AF135440.1	huntington yeast partner C (Hypc)	M.M.	37.0	1
AF061260	immunosuperfamily protein B12	M.M.	25.0	1
AF091047	KH domain RNA binding protein QKI-7B	M.M.	94.7	1
X75947	mCBP	M.M.	95.3	1
M30685	MHC class I protein mRNA (MHCPATRF1)	P.T.	70.3	1
L02897	nonerythroid beta-spectrin	Dog	49.9	1
M15825	nucleolin (C23)	C.H.	75.1	1
D84649	p27/Kip1	F.C.	30.3	1
X89969	polyA binding protein II	B.T.	86.9	1
U78090	potassium channel regulator 1 mRNA	R.N.	85.6	1
X89650	Rab7 protein	M.M.	33.2	1
U89254	retina specfic RGS protein (RET-RGS1)	B.T.	22.0	1
L02953	ribonucleoprotein (xrp1)	X.L.	46.3	1
D78188	SCID complementing gene 2	M.M.	42.9	1
AF084205	serine/threonine protein kinase	R.N.	27.4	1
U35245	vacuolar protein sorting homolog r-vps33b	R.N.	84.6	1
U40825	WW-domain binding protein 1 mRNA	M.M.	39.6	1

ESTs match to nonhuman sequences may represent the human homologs of these genes.
Quality of match is given as percent identity (%ID).
[a]Abbreviations: B.T., *Bos taurus;* C.F., *Canis familiaris;* C.H., Chinese hamster; Ch, Chimpanzee; F.C., *Felis catus;* G.G., *Gorilla gorilla;* M.M., *Mus musculus;* M.Pu., *Mustela putorius;* O.C., *Oryctolagus cuniculus;* P.T., *Pan troglodytes;* R.N., *Rattus norvegicus;* X.L., *Xenopus laevis.*

Identification of Tissue- and Developmental-Stage-Specific Genes by the Compilation of the Expression Profiles of HFL22w and the Other Functionally Associated Tissues or Cells

Although we were profiling the active genes in HFL22w based upon 13,077 ESTs, as yet the number of ESTs collected for each expression profile obtained from the published data was only approximately 1000. It was not possible to compare the genes that appeared at low abundance. However, with those genes whose transcripts appeared at high abundance and represent typical physiological and developmental status, relatively accurate comparisons could be made and the conclusion might even be more objective. Therefore, genes listed in the tables were extracted from each of these expression profiles, detected two or more times, and the abundance of their transcripts among total ESTs compiled. Through the comparison, several gene groups associated with definite physiological and/or molecular features were identified.

We collected five other liver-associated expression profiles including human fetal liver at 19 wk (HFL19w) or 40 wk (HFL40w) of gestation, human adult liver (HAL), Itoh cells, and HepG2 cells (http://bodymap.ims.u-tokyo.ac.jp/human_1.html) and compared them with the expression profile of HFL22w established here. We extracted 773 genes whose abundance was two or more in at least one of the six expression profiles and compiled their activities (EST frequency). Only the genes whose transcripts appeared 15 or more times in the compiled expression profile are shown in Table 4. These genes were categorized into three classes according to the number of libraries in which they were detected: ubiquitous—appeared in five or six origins (filled area in the Library column, lib); common—appeared in two–four origins (hatched area in the Library column), and unique—appeared

Table 4. Compiled Gene Expression Profile Associated with Liver

Primary accession	lib	L1	L2	L3	L4	IC	HC	Gene definition
M15386		–	724	–	–	–	–	hemoglobin gamma-G
V00494	■	54	694	45	35	–	17	Serum albumin (ALB)
M11147	■	4	91	2	2	3	5	ferritin L chain
M32053		–	70	–	–	–	–	H19 RNA gene
X00129		–	69	–	5	–	–	retinol binding protein (RBP)
AF105974		–	68	–	–	–	–	alpha one globin (HBA1)
M16961	▨	8	67	18	–	–	3	alpha 2-HS-glycoprotein alpha and beta chain
X01683	■	13	62	9	6	–	7	alpha 1-antitrypsin
J04617	▨	–	62	–	–	11	17	elongation factor EF-1-alpha
X07868	▨	–	55	5	–	–	–	insulin-like growth factor II
S95936	▨	3	46	6	10	–	–	transferrin
J00129	▨	21	41	15	8	–	–	fibrinogen beta-chain
X51473	▨	7	38	2	5	–	–	fibrinogen gamma chain
M60854	▨	–	37	2	–	–	4	ribosomal protein S16
U22961		–	34	–	–	–	–	mRNA clone with similarity to L-glycerol-3-phosphate:NAD oxidoreductase and albumin gene sequences
AF042513		–	31	–	–	–	–	isolate Asn6 cytochrome b (CYTB)
AF014894		–	31	–	–	–	–	NADH dehydrogenase subunit 2
M36676		–	30	–	–	–	–	apolipoprotein B100
J04763		–	30	–	–	–	–	complement component 3 (C3)
D87666		–	30	–	–	–	–	heart mRNA for hsp90
L20955	▨	–	29	–	–	–	2	prothymosin alpha
X00955	■	11	25	7	9	–	3	apolipoprotein AII
NM_000477.1		–	24	–	–	–	–	albumin (ALB)
X04225		–	24	–	–	–	–	protein HC (alpha 1-microglobulin)
NM_000559.1		–	23	–	–	–	–	hemoglobin, gamma A
X03168		–	23	–	–	–	–	S-protein
X02162		–	20	–	–	–	–	apolipoprotein AI (apo AI)
U14990		–	20	–	–	–	–	XP1PO ribosomal protein S3
M11313	▨	6	18	3	5	–	–	alpha 2-macroglobulin
V00497	▨	2	16	2	–	–	–	beta-globin
M64982	▨	6	16	15	2	–	–	fibrinogen alpha chain
AF028832		–	16	–	–	–	–	Hsp90
X16064	■	11	15	2	2	9	9	translationally controlled tumor protein
X14420	▨	–	10	–	–	16	–	pro-alpha 1 type 3 collagen
X00637	▨	23	4	15	28	–	–	haptoglobin alpha 1S (Hpa 1S)
X55656	▨	17	2	–	–	–	–	gamma-G globin
J03040	▨	–	2	–	–	19	–	SPARC/osteonectin
M13692	▨	–	–	4	15	–	–	alpha-1 acid glycoprotein
X13345		–	–	–	–	15	–	plasminogen activator inhibitor 1 (PAI-1)
J02775	▨	10	–	29	13	–	2	RFLP 3′ to the apolipoprotein B gene

Forty gene species matched to known human genes were listed in descending order of EST abundance of the genes in HFL22w and the order of gene definition. Only those genes whose transcripts appeared fifteen or more times are presented.
Lib, library; L1, human fetal liver aged 19 wk; L2, human fetal liver aged 22 wk; L3, human fetal liver aged 40 wk; L4, human adult liver; IC, Itoh cell; HC, HepG2 cell.

in only one origin (blank in the Library column). The functions of a gene could be assumed from the frequencies in random isolates from the different libraries shown in the compiled expression profiles. Among the 773 genes, nine (Gene Group I) appeared ubiquitously (Table 5). Some of them were likely to function in housekeeping, such as the three ribosomal proteins. The other six genes were actually tissue-specific, function-keeping genes of liver, including serum albumin, ferritin L chain, and apolipoprotein AII.

On the other hand, 636 genes appeared only in one library (Table 4, blanks in Library column). Because their relatively high expression was unique to one expression profile among the listed six, they were the candidate genes whose products exerted unique functions in Itoh cells, HepG2 cells, or the liver in the different stages of development, respectively.

Eleven genes were expressed only in HFL19w and HFL22w but not in HFL40w or HAL (Table 5, Gene Group II). They were α-fetoprotein (AFP), 23-kD highly basic protein, thymosin-4, insulinoma *rig-analog* mRNA encoding DNA-binding protein, and seven ribosomal proteins. Genes expressed only in HAL and HFL40w but not in HFL19w or HFL22w are also listed (Table 5, Gene Group III). They, together with the genes of Gene Group II (Table 5), are developmental-stage-specific genes, which are suitable candidates for molecular probes to characterize the developmental stage of fetal liver. Further analysis of them would give impetus to the research of the molecular mechanism of liver development.

We also identified two other gene groups through systematic analysis of the mRNA population differences between the normal cells and the tumor cells in the liver. Gene Group IV consists of the genes expressed only in the three fetal livers and the adult liver but not in the hepatoblastoma HepG2 cells (Table 5). These genes might be candidate tumor suppressor genes or genes that were inhibited during tumorigenesis. On the contrary, Gene Group V consisted of genes expressed only in the HepG2 cells but not in the normal liver in various developmental stages (data not shown). These genes might be associated with tumorigenesis of the liver. Six genes in Gene Group II (Table 5) such as α-fetoprotein (AFP); ribosomal proteins L9, L19, S3a, and L6; and insulinoma *rig-analog* mRNA encoding DNA-binding protein were expressed in HepG2 cells and human fetal liver in the early stage of development (age

Table 5. Classification of Gene Groups Associated with Liver Development

Primary acc.	L1	L2	L3	L4	IC	HC	Gene definition	Gene cat.
Gene Group I: Genes ubiquitously expressed in liver								
V0094	54	694	45	35	–	17	serum albumin (ALB)	IX
M11147	4	91	2	2	3	5	ferritin L chain	I
X01683	13	62	9	6	–	7	alpha 1-antitrypsin	XI
X00955	11	25	7	9	–	3	apolipoprotein AII	IX
X16064	11	15	2	2	9	9	translationally controlled tumor protein	VI
X02761	3	6	7	3	10	–	fibronectin	XIII
L22154	2	5	4	–	9	5	ribosomal protein L37a	VII
X89401	3	3	–	3	3	6	ribosomal protein L21	VII
U14968	4	3	2	2	9	6	ribosomal protein L27a	VII
Gene Group II: Genes expressed only in HFL19w and HFL22w, but not in HFL40w nor HAL								
D14531	5	13	–	–	8	5	homolog of rat ribsomal protein L9	VII
V01514	3	12	–	–	–	2	alpha-fetoprotein (AFP)	IX
X56932	2	11	–	–	9	–	23 kD highly basic protein	XII
U14966	2	9	–	–	3	–	ribosomal protein L5	VII
D14530	2	6	–	–	3	6	homolog of yeast ribosomal protein S28	VII
S56985	2	6	–	–	5	3	ribosomal protein L19	VII
X69150	2	5	–	–	2	–	fibosomal protein S18	VII
M77234	2	5	–	–	–	3	ribosomal protein S3a	VII
M17733	3	5	–	–	9	–	thymosin beta-4	III
J02984	2	2	–	–	–	–	insulinoma rig-analog mRNA encoding DNA-binding protein	VI
X69391	3	2	–	–	2	2	ribosomal protein L6	VII
Gene Group III: Genes expressed only in HAL and HFL40w, but not in HFL19w nor HFL22w								
M13692	–	–	4	15	–	–	alpha-1 acid glycoprotein	XII
X05151	–	–	3	7	–	–	apoC-II preproapolipoprotein C-II	IX
M20496	–	–	3	2	4	–	cathepsin L	XI
D00097	–	–	4	2	–	–	serum amyloid P component (SAP)	X
Gene Group IV: Genes expressed only in HFL19w, 22w, 40w and HAL, but not in HepG2 cell								
S95936	3	46	6	10	–	–	transferrin	XII
J00129	21	41	15	8	–	–	fibrinogen beta chain	IX
X51473	7	38	2	5	–	–	fibrinogen gamma chain	IX
M11313	6	18	3	5	–	–	alpha 2-macroglobulin	IV
M64982	6	16	15	2	–	–	fibrinogen alpha chain	IX
X02747	3	10	2	5	–	–	aldolase B	X
X02761	3	6	7	3	10	–	fibronectin	XIII
M10050	9	6	6	8	–	–	liver fatty acid binding protein (FABP)	IX
X00637	23	4	15	28	–	–	haptoglobin alpha 1S (Hpa 1S)	IX

L1, human fetal liver aged 19 wk of gestation; L2, human fetal liver aged 22 wk of gestation; L3, human fetal liver aged 40 wk of gestation; L4, adult liver; IC, Itoh cell; HC: HepG2 cell.

19 and 22 wk of gestation) but not in HFL40w or HAL. Because tumor cells often express embryonic genes in abnormal ways, these six genes might represent oncogenic status in hepatoma cells.

Although Itoh cells are located in the liver, their gene expression profile was obviously different from those of hepatocytes at various developmental stages and of the hepatoma cell line HepG2. Out of 120 genes that had two or more EST copies in Itoh cells, 60 were not expressed in any of the five other liver-associated expression profiles. Genes commonly expressed with high levels in liver, such as serum albumin (ALB), fibrinogen, transferrin, apolipoprotein AI, and haptoglobin, were not detected in Itoh cells. The different expression profile of Itoh cells contributed to its different physiological function from other types of liver cells.

The compiled gene expression profile associated with hematopoiesis (data not shown) consisted of five gene expression profiles including the CD34+ hematopoietic progenitor/stem cell (Mao et al. 1998), CD4 T cell, CD8 T cell, granulocyte, and myeloblastic leukemia cell line HL60 cell (http://bodymap.ims.u-tokyo.ac.jp/human_1.html). They had 134, 38, 45, 20, and 46 genes that also expressed in HFL22w, respectively. It was obvious that the CD34+ hematopoietic progenitor/stem cell shared the most active genes with HFL22w. Among the 595 genes whose frequency was two or more in the expression profile of HFL22w, 134 (22.5%) genes were also expressed in CD34+ hematopoietic stem/progenitor cells. Some of them were hematopoietic system-specific, for example, hemoglobin γ-G (HMG), β-globin, and T-cell cyclophilin. But the similarity between the expression profile of HFL22w and granulocytes was much less. This result matched the fact that there were few differentiated granulocytes in HFL22w.

Full-Length cDNA Cloning from HFL22w

Based on the bioinformatics analysis, 110 EST clusters have been chosen initially for full-length cDNA cloning. The clone inserts were sequenced with end-sequencing, primer extension, and sequencing after partial deletion/subcloning. After assembling ESTs into contigs, we found that 74 (67.3%) of the 110 cDNA clones already contained a complete open reading frame (ORF). In the other 36 cDNA clones, an obvious but incomplete reading frame was present. In silico cloning with dbEST extension allowed us to obtain 22 (20.0%) putative entire ORFs, which were then confirmed by sequencing of material cDNA clones obtained by appropriately designed RT-PCR. For the remaining 14 (12.7%) cDNA clones that could not be extended properly with an electronic approach, rapid amplification of cDNA ends (RACE) was applied to get the 5′ or 3′ ends from appropriate tissue origins. In total, 110 cDNAs with putatively entire ORFs were obtained. Table 6 shows all 110 new full-length cDNAs from HFL22w. Among these 110 full-length cDNAs, 71 contained multiple exons and 87 had a consensus polyadenylation signal near the 3′ end; the 14 polyA tails might correspond to an A-rich region of the genome when they were searched against GenBank's working draft of the human genome. It is worth pointing out that, although a polyadenylation signal was found in the majority (73/110) of cDNAs as evidence of containing the complete 3′ UTR, the integrity of the 5′ UTR needs further experimental confirmation as in reports like that of the RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium (Kawai et al. 2001). Among these novel genes, the majority, 76 (69.1%), encode 80–500 amino acid residues deduced from their encoding frames. According to their homology with known genes and domains, some genes might be associated with signal transduction, such as the human homolog of mouse c-Jun leucine zipper interactive protein (cDNA JZA-20), the *Kluyveromyces lactis* transcription initiation factor IIIB 70-kD subunit, or *Bos taurus* guanine nucleotide-binding protein. And some genes might be new members of certain gene families, for example, the gene for the human homolog of *Schizosaccharomyces pombe* Arf GTPase-activating protein, now termed human ADP-ribosylation factor GTPase-activating protein (ARFGAP3), belonging to the ARF GAP family (Zhang et al. 2000; Liu et al. 2001). In addition, some genes are very conserved in the species' evolution because their encoded proteins exhibit similar primary structure with those derived from such organisms as *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Plasmodium chabaudi*, *Tetrahymena thermophila*, *Caenorhabditis elegans*, *Drosophila melanogaster* (Table 6), and other mammals (Qu et al. 2001). These novel genes might be involved in critical biological processes according to their homology to known genes with established significant functions like signal transduction, metabolism, protein expression, and hematopoiesis.

In further investigations, the chromosomal localization of 77 novel genes was determined, 70 of which were located by using database information of UniGene, dbSTS, dbHTGS, and Human Chromosome Databases; the location of the other 7 was determined by radiation hybrid (RH) mapping. The remaining 33 novel genes could not be mapped by either of the above methods.

DISCUSSION

The major objective of the human genome project is the identification of the complete set of human genes. Single-pass, partial sequencing of cDNA clones in different organs, tissues, or cells of the human body is complementary to the genomic DNA sequencing. The analysis of ESTs generated from cDNA libraries has been shown to provide an extensive and quantitative measure of the transcriptional activity of expressed genes (Adams et al. 1991; Okubo et al. 1992). Here we have undertaken the EST sequencing of the cDNA library of HFL22w as the first step of a long-term effort to explore the genes expressed in this specific developmental stage of human fetal liver. A preliminary profile of gene expression in this cell population was set up based on the analysis of 13,077 ESTs.

Current estimates place the total number of genes in the human genome at about 30,000 (Lander et al. 2001; Venter et al. 2001). The portion of the genome expressed in any given cell type or tissue is not precisely known. The mRNAs from most genes are at low levels and from a smaller number of genes at intermediate levels of expression. Only a few genes are expressed at high levels (Sargent 1987). The highly abundant species are often tissue-specific, and the majority of the rare messages are shared among all tissues examined, implying a housekeeping function (Bishop et al. 1974). As expected, gene categories IX (liver and lipoproteins) and VIII (hematopoiesis) consisted of tissue-specific and stage-specific genes of HFL22w. These two gene categories have 22 highly expressed genes, about one-third of the total abundant species. Meanwhile, two gene categories—X (metabolism) and VII (gene/protein expression)—which included most of the housekeep-

Table 6. List of the Full-Length cDNA from HFL22w and Their Homologous Genes

Primary accession	Homologous gene definition[a]	cDNA (bp)	ORF (aa)	Chromosome localization	Species[b]
AF078841	c-Jun leucine zipper interactive (cDNA JZA-20)	1233	237	1	M
AF078842	CG11323 gene product	1684	292	3p25.1-25.2	D
AF078843	CG9253 gene product	1647	401	12	D
AF090898	Novel (HQ0149)	1675	67		
AF090900	Novel (HQ0189)	2390	57	11	H
AF090908	extensin-like protein	1737	177	6p24.1-25.3	A.T.
AF090911	unnamed protein product	1775	409		M
AF090915	uncharacterized bone marrow protein BM-037 H	2268	441	15q13.3-21.1	H
AF090917	OPA-containing rotein	1318	102		H
AF090919	mus308 gene product	2504	77		D
AF090921	Novel (HQ0365)	2790	118		
AF090929	conserved hypothetical protein	1292	130	16	T.M.
AF090935	hypothetical protein F57C2.5	2122	424	20p11.21-11.23	C
AF090939	Novel (HQ0641)	2162	50		
AF090945	Novel (HQ0670)	1220	92	22q11	
AF090947	CG13232 gene product	1502	217	15	D
AF111847	ArfGaap GTPase activating protein	2768	516	22q13.2-13.3	S
AF111851	Novel (HQ0611)	1824	90	16	
AF113009	KIAA1413 protein	1423	140	13q12-13	H
AF113012	Novel (HQ0767)	1606	63	9	
AF113687	Novel (HQ1158)	1619	82	14	
AF113691	CG14407 gene product	851	49	14	D
AF113697	kinesin-II homologue	1905	102		T.T.
AF116607	hypothetical protein	1604	96		H
AF116608	CG4603 gene product	1490	87		D
AF116609	CG4180 gene product	1837	264	16p13.3	D
AF116610	(X66286) tensin	1588	235		G
AF116611	Novel (HQ0943)	1563	59	11p15.5	
AF116617	erythrocyte membrane antigen	1747	309	9	P
AF116618	putative protein kinase	1730	418	1	A.T.
AF116620	system A transporter isoform 2 (SAT2) mRNA	2270	506		R
AF116637	Novel (HQ1489)	1321	53		
AF116638	hypothetical protein (L1H 3′ region)	1201	105		H
AF116642	Novel (HQ1618)	1240	117	X	
AF116643	Novel (HQ1635)	1301	62	14	
AF116646	unnamed protein product	2286	109	X	H
AF116652	X-linked PEST-containing transporter	2410	201	6	H
AF116655	Novel (HQ1082)	1876	100		
AF116657	Novel (HQ1310)	1646	164		
AF116662	Novel (HQ1446)	1998	75	3q27	
AF116672	Novel (HQ1905)	653	102		
AF116677	Novel (HQ1966)	1200	52	11q23.2	
AF116678	Novel (HQ1995)	975	128	9	
AF116682	putative protein	876	141		A.T.
AF116688	serum albumin (ALB)	1247	97	5	H
AF116692	CG8972 gene product	1262	356	3	D
AF116694	predicted coding region AF0392	1083	183		A.F.
AF116701	CG6516 gene product	1234	408	9	D
AF116703	CG1354 gene product	1303	396	22q13.1-13.2	D
AF116704	CG17180 gene product	1839	357	17p13.3-17qter	D
AF116707	mRNA for KIAA1147 protein	1137	189	7q32	H
AF116708	putative NADH oxidoreductase complex I subunit	1053	251	8	D
AF118062	Novel (HQ1386)	2352	75	5	
AF118068	Novel (HQ1596)	1341	75		
AF118077	Novel (HQ1808)	1580	68	Xp22	
AF118080	tetratricopeptide repeat protein	1263	91	11q13.2-13.4	H
AF118082	tRNA selenocysteine associated protein	1634	287		R
AF118084	Novel (HQ1914)	1737	67	X	
AF118087	CG16947 gene product	1543	253	4	D
AF118088	KIAA1240 protein	1934	579	8	H
AF119843	protein serine/threonine phosphatase 4 regulatory subunit 1	2608	342	20	H
AF119857	ribosome-binding protein p34	2341	307	17	R
AF119864	mitochondrial carrier family protein	1402	351	17	H
AF119865	Novel (HQ2176)	1673	90	5	
AF119869	CG9241 gene product	1658	364		D
AF119870	Novel (HQ2266)	2211	122		
AF119872	Novel (HQ2272)	1597	88	7q31	
AF119878	hypothetical HTLV-1 related endogenous sequence HRES1 25K	1581	75	2p14-2p12	H

Table 6. (*Continued*)

Primary accession	Homologous gene definition[a]	cDNA (bp)	ORF (aa)	Chromosome localization	Species[b]
AF119880	Novel (HQ2372)	1270	70	14	
AF119881	acetyltransferase Tubedown-1	1408	51		M
AF119882	Novel (HQ2492)	1526	94	17	
AF119884	unnamed protein product	1646	394	17	H
AF119907	Novel (HQ2949)	1992	137		
AF119908	Novel (HQ2955)	1750	77	18	
AF130049	CG7611 gene product	2553	139	1	D
AF130058	transcription initiation factor IIIB 70 KD subunit	1955	419	8p11.2	K
AF130060	erythrocyte membrane antigen	2048	481	9	P
AF130061	polybromo 1 protein	3185	306	3p24.3-3p13	G
AF130066	CG17665 gene product	1280	582	1	D
AF130072	CG5087 gene product	5732	1068		D
AF130074	Novel (HQ2523)	1741	117	9	
AF130076	RNA binding protein	1131	213	16	H
AF130079	Novel (HQ2852)	1769	169	9	
AF130081	KIAA0680 protein	2006	339	1	H
AF130083	Novel (HQ1737)	2299	62		
AF130091	FH protein interacting protein FIP2	1942	148	5	A.T.
AF130096	CG7288 gene product	1874	404	2	D
AF130104	Novel (HQ0756)	1054	63	14q32.1	
AF130106	guanine nucleotide binding protein (G protein), gamma 2 subunit	2108	71	14p13-14q32.33	B
AF130107	Novel (HQ1433)	2243	91		
AF130112	Novel (HQ1953)	2153	127	14	
AF130114	Novel (HQ2459)	1248	90	14q24.3	
AF132198	probable membrane protein	2755	627		D
AF132206	Novel (HQ2397)	1878	81		
AF138861	Novel (HQ0848}	2614	61	14	
AF138863	hypothetical protein	1524	264	7	H
AF305815	CG11190 gene product	2169	94	20q12-13.12	D
AF305816	Novel (HQ0633)	1946	56	4	
AF305817	Novel (HQ0715)	1799	75	20	
AF305818	Novel (HQ0764)	2225	133	1	
AF305819	Novel (HQ0777)	3774	114		
AF305820	Novel (HQ0875)	1979	64	6q14	
AF305821	Novel (HQ0902)	2626	67	6q23.1-24.3	
AF305822	Novel (HQ0996)	1241	79	7p15.3-p21	
AF305823	CG5850 gene product	1660	138	4q28.3-32.3	D
AF305824	hypothetical protein R53.5	1172	152		C
AF305825	Novel (HQ2869)	1326	87	11q12	
AF305826	putative acid phosphatase	1628	193	2	C
AF305827	hypothetical protein	2017	88		H
AF305828	synapse-associated protein	1813	186		D

[a]"Novel" means that this novel gene has no significant match to previously-deposited genes. The number following HQ in parentheses is the clone_ID of this gene.
[b]Abbreviations: A.T., *Arabidopsis thaliana*; A.F., *Archaeoglobus fulgidus*; B, *Bos taurus*; C, *Caenorhabditis elegans*; D, *Drosophila melanogaster*; G, *Gallus gallus*; H, *Homo sapiens*; K, *Kluyveromyces lactis*; M, *Mus musculus*; P, *Plasmodium chabaudi*; R, *Rattus norvegicus*; T.T., *Tetrahymena thermophila*; T.M., *Thermotoga maritima*.

ing genes, had 30.9% (445/1442) of the genes expressed at low levels, whose frequency is equal to or less than 3.

Our initial goal was to gain a broad understanding of both the diversity and the abundance of gene expression in HFL22w. HFL22w has its tissue-specific and stage-specific functions. In the liver of a human fetus, besides the general metabolism of carbohydrates, fats and proteins, hematopoiesis, which originated in the yolk sac, occurs in the liver from the 6th wk to the 7th month of gestation. After the immigration of the hematopoietic system into the fetal liver at 2 months of gestation, human fetal liver gradually becomes a major site of embryonic hematopoiesis, and, intriguingly, coexistence of hepatic and hematopoietic systems appears. Moreover, at 22 wk of gestation, human fetal liver displays the balance of immigration and emigration of the hematopoietic system. Therefore, HFL22w is an excellent model for unraveling the mechanisms of interaction between hepatic and hematopoietic systems and of immigration and emigration of the hematopoietic system during mammalian development, and is a suitable resource for identification of novel significant genes.

Although gene activities were not simply reflected by the abundance of various mRNAs, gene expression profiling leads to the best approximation about them. Because there was a satisfactory representation of ESTs generated from HFL22w, the gene expression profile could be analyzed in terms of both patterns and levels. The profile dramatically reflected the hepatic and hematopoietic activities of HFL22w as described above. The quantitative ratios should help us understand its major functional feature. For instance, the mRNA of hemoglobin γ-G was the most abundant mRNA in HFL22w, which had 724 EST copies. Considering that it plays a pivotal role in

hematopoiesis, its high abundance in expression profiling of HFL22w strongly indicated that HFL22w was a major site of embryonic hematopoiesis and that the expression profiling of HFL22w reported here could objectively represent the molecular features of human fetal liver. Hemoglobin is composed of four kinds of polypeptide chains, each of which is the product of a specific gene. Choi et al. (1995) reported the appearance of adult-type hemoglobin (hemoglobin β) and concluded that the transition of hemoglobin type from fetal to adult form has already begun in the 22-wk-old fetal liver before the bone marrow takes over the hematopoietic function. However, we found the appearance of embryonic-type hemoglobin (hemoglobin ζ) but no hemoglobin β in HFL22w. This showed that the transition of hemoglobin type from fetal to adult form had not yet begun and the transition of hemoglobin type from embryonic to fetal form had not completely finished at this stage. In addition, serum albumin had 694 EST copies in our profiling. It has been known as a main component for maintaining the colloid-osmotic pressure of plasma, as well as for binding bilirubin or lipids for eventual excretion. It could therefore be concluded that albumin synthesis, the typical liver-specific function, has begun in HFL22w. These results showed that the typical fetal liver functions of either hepatic biochemical metabolism or hematopoiesis were maintained through high rates of transcription of specific genes. Meanwhile, since the number of sequenced clones was large enough, it is possible to identify those genes with low level expression, or those with unknown functions. Actually, hepatopoietin (HPO) (Wang et al. 1999; Li et al. 2000) expression was detected in HFL22w, indicating that it may also function in fetal liver development. Through the comparison of the liver-associated expression profiles, we found 11 genes only expressed in the fetal liver during the early stage of liver development, which might be tissue-specific and stage-specific. Of them, α-fetoprotein (AFP) was highly expressed as expected. It was a serum glycoprotein normally present in high concentration in fetal and maternal serum but in low concentration in normal adult liver (Kew 1990). As the most typical liver oncodevelopmental protein, reappearance of AFP in high concentrations in adulthood is a strong pointer to the diagnosis of hepatocellular carcinoma, and in childhood to either hepatoblastoma or hepatocellular carcinoma. 23-kD highly basic protein is a protein whose precise physiologic function is unknown. As a kind of thymic hormone, thymosin β-4 is necessary for differentiation of stem cell precursors into mature cells (Kamani and Douglas 1991). The expression of thymosin β-4 in early fetal liver confirmed that during the 22 wk of gestation, human fetal liver was actually a major site of embryonic immune development. Insulinoma *rig-analog* mRNA encodes a DNA-binding protein, and the deduced 145-amino acid sequence remains invariant in hamster, human, and rat insulinomas, suggesting that *rig* has evolved under extraordinarily strong selective constraints (Inoue et al. 1987). *rig* was found to be expressed in rat regenerating liver and in rat primarily cultured hepatocytes. The level of *rig* mRNA was increased at the proliferative phase of liver regeneration. In synchronously cultured hepatocytes, the *rig* mRNA level was elevated at the G_1 phase of the cell cycle and the rig protein accumulated in the nuclei during the S phase (Inoue et al. 1988). These results indicate that *rig*, and the insulinoma *rig-analog* mRNA expressed in the early stage of development of human fetal liver, could be involved in a more general way in growth or cell proliferation.

The timing course of the successive developmental processes is one of the most fundamental aspects of ontogenesis. The liver development during various stages was apparently under the control of sequential gene expression as the dominant, though perhaps not exclusive, mechanism. Therefore, single-pass sequencing of randomly selected cDNAs, which is a rapid and efficient method for discovering new transcripts and for expression profiling the active genes, with consequent comparison of the profiles for determining patterns of gene expression during the different stages of liver development, did help us understand more about the functional features of HFL22w and identify gene groups consisting of candidate genes playing important roles during human liver development.

Actually, through the comparison of the expression profiles, we found that along with the development of the liver (from HFL19w to HAL), the expression level of translationally controlled tumor protein (TCTP) and its rank position of expression frequency among all the genes expressed in the tissues obviously dropped. In comparison, the expression level of TCTP in HepG2 cells was conversely very high and close to that of the fetal liver at early developmental stages (Table 7). Therefore, TCTP may be a dedifferentiation marker of liver or hepatocytes.

Generally speaking, most of the highly expressed genes have already been identified. So far, a large number of human genes have been labeled by dbESTs, and the proportion could be even higher in the databases of some genomic industries. However, the poor representation of some important genes in dbEST indicates that completion of the list of human genes, especially those with low-level expression or temporally and/or spatially restricted expression, needs continuous effort. Therefore, the Group II ESTs (5460), accounting for 41.8% of all ESTs obtained, are worth paying particular attention to in the future discovery of novel genes. Based on the novel ESTs and the homologous ESTs with nonhuman matches identified in HFL22w and taking advantage of the UniGene information in public databases and the available rapid amplification of cDNA ends PCR technology, we cloned 110 full-length cDNAs of novel genes. The tools of bioinformatics not only help to clone novel genes through dbEST assembly, but also provide important clues to the function of novel genes through comparison of homology of known genes with established functions and those genes from model organisms. Among the 110 novel genes, we have found that at least 4 may participate in signal transduction and that 8 genes were similar to the *D. melanogaster* genes predicted based on the genome sequence of *D. melanogaster* (Adams et al. 2000). However, to systematically characterize these genes involved in the molecular mechanism of fetal liver development, em-

Table 7. Expression Pattern of Translationally Controlled Tumor Protein

			Expression pattern		
Tissues/cells	ESTs	Genes	EST	ratio (%)	rank
HFL19w	570	57	11	1.93	6
HFL22w	13077	1660	15	0.11	33
HFL40w	529	48	2	0.38	44
HAL	620	64	2	0.32	61
Itoh cell	1120	120	9	0.80	14
HepG2 cell	741	75	9	1.21	4

bryonic hematopoiesis, and tumorigenesis, several approaches, such as microarray and yeast two-hybrid system technologies, should be used in grouping analysis of gene expression kinetics and protein interaction in human fetal liver.

METHODS

DNA Sequencing

Bacteria growth and plasmid extractions of the HFL22w cDNA library (CLONTECH) were performed by a QIAprep 96 Turbo Miniprep Kit (QIAGEN). Sequencing reactions were performed on a GeneAmp PCR System 9700 thermal reactor (Perkin-Elmer) by using a BigDye Terminator Cycle Sequencing Kit (Perkin-Elmer) with T7 or SP6 primers. After removing the unincorporated dye terminators from sequencing reactions with DyeEx Spin Kits (QIAGEN), the reaction products were electrophoresed on an ABI 377-XL DNA sequencer (Perkin-Elmer–Applied Biosystems), and raw sequence data were automatically recorded.

Data Management and Bioinformatics Analysis

Sequences were edited manually by using PHRED and Sequencher (version 3.0) to remove vector sequence and identify trash sequences, defined as sequences from bacterial DNA, sequences from primer polymers, sequences containing >1% of ambiguous bases (*N*), or sequences shorter than 100 bp. All sequence data were preserved on record tape. An in-house database for EST sequences generated from a cDNA library of HFL22w was established. The individual ESTs were searched against the GenBank nonredundant database (Release 105.0) for homology comparison by using BLASTN on the BLAST network server at the National Center for Biotechnology Information (NCBI). ESTs with an Alignment Score of the Basic Local Alignment Search (BLAST) >200 were considered to identify known genes or to have partial homology to known genes; the others were considered novel ones. Clustering of the ESTs generated in this work was performed by using PHRAP with default parameters.

Full-Length cDNA Cloning

The new sequences, considered as part of novel genes, confirmed by similarity searching against GenBank, were selected for full-length cDNA cloning. The program ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was applied to analyze the open reading frames. For those clones containing partial reading frames, in silico cloning and RACE were performed. In silico cloning was carried out using dbEST information, starting from the sequences obtained from the HFL22w cDNA library and then confirming these by sequencing of material cDNA clones obtained by appropriately designed RT-PCR. Sequence ambiguity existing in these contigs was clarified by further sequencing. A Smart RACE cDNA Amplification Kit (Clontech) was used to facilitate full-length cDNA cloning.

Genomic Mapping of Full-Length cDNA Clones

The chromosomal assignment of novel genes was mapped by two strategies: searching sequence databases such as Unigene, dbSTS, Human Chromosome Databases, dbHTGS at the National Center for Biotechnology Information; or radiation hybrid (RH). The Genebridge 4 RH panel (Research Genetics) was used in RH mapping.

ACKNOWLEDGMENTS

This work was partially supported by the Chinese National Key Project of Basic Research, the Chinese National High-tech Program, the Chinese National Distinguished Young Scholar Awards, the Chinese National Natural Science Foundation Key Project, and the Beijing City Municipal Key Project.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

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. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252:** 1651–1656.

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C., and Venter, C. 1992. Sequence identification of 2,375 human brain genes. *Nature* **355:** 632–634.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulder, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., et al. 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377** (6547 Suppl.): 3–174.

Adams, M.D., Celniker, S.E., Holt, R.O., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

Bishop, J.O., Morton, J.G., Rosbash, M., and Richardson, M. 1974. Three abundance classes in HeLa cell messenger RNA. *Nature* **250:** 199–204.

Choi, S.S., Yun, J.W., Choi, E.K., Cho, Y.G., Sung, Y.C., and Shin, H.-S. 1995. Construction of a gene expression profile of a human fetal liver by single-pass cDNA sequencing. *Mamm. Genome* **6:** 653–657.

Dabeva, M.D., Hurston, E., and Shafritz, D.A. 1995. Transcription factor and liver-specific mRNA expression in facultative epithelial progenitor cells of liver and pancreas. *Am. J. Pathol.* **147:** 1633–1648.

Fausto, N. 1991. Growth factors in liver development, regeneration and carcinogenesis. *Prog. Growth Factor Res.* **3:** 219–234.

Godin, I., Dieterlen-Lievre, F., and Cumano, A. 1995. Emergence of multipotent hematopoietic cells in the yolk sac and paraaortic splanchnopleura in the mouse embryos, beginning at 8.5 days postcoitus. *Proc. Natl. Acad. Sci. USA* **92:** 773–777.

Huang, H. and Auerbach, R. 1993. Identification and characterization of hematopoietic stem cells from the yolk sac of the early mouse embryo. *Proc. Natl. Acad. Sci.* **90:** 10110–10114.

Inoue, C., Shiga, K., Takasawa, S., Kitagawa, M., Yamamoto, H., and Okamoto, H. 1987. Evolutionary conservation of the insulinoma gene *rig* and its possible function. *Proc. Natl. Acad. Sci.* **84:** 6659–6662.

Inoue, C., Igarashi, K., Kitagawa, M., Terazono, K., Takasawa, S., Obata, K., Iwata, K., Yamamoto, H., and Okamoto, H. 1988. Expression of the insulinoma gene *rig* during liver regeneration and in primary cultured hepatocytes. *Biochem. Biophys. Res. Commun.* **150:** 1302–1308.

Kamani, N.R. and Douglas, S.D. 1991. Structure and development of the immune system. In *Basic and clinical immunology,* 7th ed. (eds. D.P. Stites and A.I. Terr), 9–33. Appleton and Lange, Norwalk, CT.

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. (The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium). Functional annotation of a full-length mouse cDNA collection. *Nature* **409:** 685–690.

Kew, M.C. 1990. Tumors of the liver. In *Hepatology: A textbook of liver disease,* 2nd ed. (eds. D. Zakim and T.D. Boyer), 1206–1240. Saunders, Philadelphia, PA.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. (International Human Genome Sequencing Consortium). Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Li, Y., Li, M., Xing, G., Hu, Z., Wang, Q., Dong, C., Wei, H., Fan, G., Chen, J., Yang, X., et al. 2000. Stimulation of the mitogen-activated protein kinase cascade and tyrosine phosphorylation of the epidermal growth factor receptor by hepatopoietin. *J. Biol. Chem.* **275:** 37443–37447.

Liew, C.C., Hwang, D.M., Fung, Y.W., Laurenssen, C., Cukerman, E., Tsui, S., and Lee, C.Y. 1994. A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc. Natl. Acad. Sci.* **91:** 10645–10649.

Liu, X., Zhang, C., Xing, G., Chen, Q., and He, F. 2001. Functional characterization of novel human ARFGAP3. *FEBS Lett.* **490:** 79–83.

Mao, M., Fu, G., Wu, J., Zhang, Q., Zhou, J., Kan, L., Huang, Q., He, K., Gu, B., Han, Z., et al. 1998. Identification of genes expressed in human CD34+ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc. Natl. Acad. Sci.* **95:** 8175–8180.

Migliaccio, G., Migliaccio, A.R., Petti, S., Mavilio, F., Russo, G., Lazzaro, D., Testa, U., Marinucci, M., and Peschle, C. 1986. Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac–liver transition. *J. Clin. Invest.* **78:** 51–60.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., and Matsubara, K. 1992. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* **2:** 173–179.

Papadopoulos, N., Nicolaides, N.C., Wei, Y.F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., et al. 1994. Mutation of a *mutL* homolog in hereditary colon cancer. *Science* **263:** 1625–1629.

Qu, X., Zhang, C., Zhai Y., Xing, G., Wei, H., Yu, Y., Wu, S., and He, F. 2001. Characterization and tissue expression of a novel human gene *npdc1*. *Gene* **264:** 37–44.

Ryo, A., Kondoh, N., Wakatsuki, T., Hada, A., Yamamoto, N., and Yamamoto, M. 1998. A method for analyzing the qualitative and quantitative aspects of gene expression: A transcriptional profile revealed for Hela cells. *Nucleic Acids Res.* **26:** 2586–2592.

Sargent, T.D. 1987. Isolation of differentially expressed genes. *Methods Enzymol.* **152:** 423–432.

Sterky, F., Regon, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, A., Villarroel, R., et al. 1998. Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Proc. Natl. Acad. Sci.* **95:** 13330–13335.

Tavassoli, M. 1991. Embryonic and fetal hematopoiesis: An overview. *Blood Cells* **17:** 269–281.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291:** 1304–1351.

Wang, G., Yang, X., Zhang, Y., Wang, Q., Chen, H., Wei, H., Xing, G., Xie, L., Hu, Z., Zhang, C., et al. 1999. Identification and characterization of receptor for mammalian hepatopoietin that is homologous to yeast ERV1. *J. Biol. Chem.* **274:** 11469–11472.

Wool, I.G. 1986. *Studies of the structure of eukaryotic (mammalian) ribosomes*. Springer Verlag, New York.

Zhang, C., Yu, Y., Zhang, S., Liu, M., Xing, G., Wei, H., Bi, J., Liu, X., Zhou, G., Dong, C., et al. 2000. Characterization, chromosomal assignment, and tissue expression of a novel human gene belonging to the *ARF GAP* family. *Genomics* **63:** 400–408.

Received December 15, 2000; accepted in revised form May 14, 2001.