Available online at www.sciencedirect.com
SCIENCE DIRECT
Developmental Biology 294 (2006) 482–496
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

Genomes & Developmental Control

HNF factors form a network to regulate liver-enriched genes in zebrafish

Wei Cheng [1], Lin Guo [1], Zhenhai Zhang [1], Hui Meng Soo, Chaoming Wen, Wei Wu, Jinrong Peng *

Functional Genomics Lab, Institute of Molecular and Cell Biology, 61 Biopolis Drive, Protesos, 138673, Singapore

Received for publication 19 December 2005; revised 23 February 2006; accepted 15 March 2006
Available online 24 April 2006

Abstract

Defects in some of liver-enriched genes in mammals will cause liver- and/or blood-related diseases. However, due to the fact that embryogenesis happens intrauterinally in the mammals, the function of these liver-enriched genes during liver organogenesis is poorly studied. We report here the identification of 129 genuine liver-enriched genes in adult zebrafish and show that, through in situ hybridization, 69 of these genes are also enriched in the embryonic liver. External embryogenesis coupled with the well-established morpholino-mediated gene knock-down technique in zebrafish offers us a unique opportunity to study if this group of genes plays any role during liver organogenesis in the future. As an example, preliminary study using morpholino-mediated gene knock-down method revealed that a novel liver-enriched gene *leg1* is crucial for the liver expansion growth. We also report the analysis of promoter regions of 51 liver-enriched genes by searching putative binding sites for Hnf1, Hnf3, Hnf4 and Hnf6, four key transcription factors enriched in the liver. We found that promoter regions of majority of liver-enriched genes contain putative binding sites for more than one HNF factors, suggesting that most of liver-enriched genes are likely co-regulated by different combination of HNF factors. This observation supports the hypothesis that these four liver-enriched transcription factors form a network in controlling the expression of liver-specific or -enriched genes in the liver.

Keywords: Microarray; Liver-enriched genes; HNF factors; Liver development; Promoter analysis; Functional genomics

Introduction

Adult liver is one of the largest organs in the body and consists of two lobes. Approximately 60% of cells in the adult rat liver are hepatocytes, while the remaining cells consist largely of cholangiocytes (bile duct cells), Kuppfer cells, stellate cells and a variety of endothelial cells including those lining the sinusoids (sinusoidal endothelial cells) (Blouin et al., 1977). Liver is an essential organ that plays a number of vital functions in the body. The functions of the liver include processing nutrients from ingested food, forming and maintaining metabolite and serum protein concentrations in the blood and serving as a site for detoxification and for hematopoiesis during gestation. In response to the metabolic demand of the body, the liver expresses vast varieties of genes to encode enzymes and blood components to comply with these diverse functions (Kawamoto et al., 1996). Among the vast number of genes expressed in a mature liver, expression of many genes, including genes for transcription factors and metabolic enzymes, has been found to be specific to or enriched in the liver (Cereghini, 1996; Kawamoto et al., 1996).

The liver is originated from the definitive endoderm. Based on extensive studies (cell lineage tracing, anatomic, molecular and genetic studies) in chick and mouse, the process of liver organogenesis is arbitrarily divided into the following stages: (1) the ventral foregut endodermal cells proximal to cardiogenic mesoderm to gain competence to become hepatoblasts; (2) specification of hepatoblast; (3) formation of liver bud; (4) fast proliferation of hepatoblasts and (5) differentiation of hepatoblast cells into hepatocytes and billiary lineages (Duncan, 2003; Zaret, 2002). The accomplishment of liver organogenesis is under the control of a genetic network that relies on the precise coordinate actions of many genes that are essential for normal development. Gene expression profiling, for example, has revealed that gene expression patterns in the liver change accordingly during the different stage of liver development (Jochheim et al., 2003; Petkov et al., 2004). Although genes enriched in a mature liver have been well-established, genes

* Corresponding author. Fax: +65 67791117.
E-mail address: pengjr@imcb.a-star.edu.sg (J. Peng).
[1] These authors contributed equally to this work.

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.03.018

enriched in the developing liver have been less systematically studied. Only a handful of such genes were obtained in *Xenopus* through screening the expression cDNA libraries (Chen et al., 2003; Zorn and Mason, 2001). In addition, it is an intriguing question if mature liver-enriched genes are also enriched in the embryonic liver since this group of genes might play a crucial role for maintaining the liver structure or the status of liver stem cells that are important for liver regeneration after hepatectomy (Su et al., 2002).

Due to some of its unique advantages for genetic studies, zebrafish (*Denrio rerio*) has been adopted as a genetic model system to study the vertebrate development in recent years (Driever et al., 1996; Haffter et al., 1996). Zebrafish is particularly amenable for the study of liver organogenesis. External development and optical clarity of embryo facilitate histological studies of liver development during early embryogenesis. In mammals, since the embryonic liver is an early hematopoietic organ, mutations affecting liver development will cause anemia that will lead to early lethality (Reimold et al., 2000), thus making in vivo studies of the progress of mouse liver morphogenesis difficult. Hematopoiesis in zebrafish takes place in the intermediate cell mass (ICM) and subsequently in the kidney, not the liver (Thisse and Zon, 2002), thus liver defects do not lead to anemia. Meanwhile, zebrafish embryos receive enough oxygen through diffusion to allow embryonic development to proceed relatively normally for several days even without blood circulation (Stainier, 2001), eliminating some of the problems encountered with mammalian model organisms. There have been a few reports describing the early liver organogenesis in zebrafish (Field et al., 2003; Ober et al., 2003; Wallace and Pack, 2003). By using a gut GFP line, Field et al. reveals two phases of liver morphogenesis: budding and growth. The budding period, which can be further subdivided into three stages, starts when hepatocytes first aggregate, shortly after 24 h post-fertilization (hpf), and ends with the formation of a hepatic duct at 50 hpf. The growth phase immediately follows and is responsible for a dramatic alteration of liver size and shape (Field et al., 2003). The adult zebrafish liver, as in mouse and human, consists of two lobes and is surrounded by blood vessels. Recently, through screening for hepatomegaly in 297 zebrafish lines bearing mutations in genes that are essential for embryogenesis, seven genes have been identified to encode factors important for regulating liver expansion growth, including the class C vacuolar sorting protein Vps18, transcription factor Sox9a, novel factor foie gras (Fgr), etc. (Sadler et al., 2005). On the other hand, loss-of-function of *def*, a novel pan-endoderm gene, and of *nil per os* (*npo*) that encodes an RNA-binding protein, caused hypoplastic liver, pancreas and gut (Chen et al., 2005; Mayer and Fishman, 2003). Other studies have shown that *gata5* (Reiter et al., 1999), *hhex* (Wallace et al., 2001), *hnf1* (Sun and Hopkins, 2001), *hnf6* (Matthews et al., 2004) and *prox1* (Liu et al., 2003), as known in mammals and other vertebrates, play crucial roles in controlling liver organogenesis and cell differentiation in zebrafish.

Our primary interest is to identify genes that are important for liver initiation and development in zebrafish. For this purpose, we have been taking both genomic and genetic approaches. In this report, we present our work on identification of 129 genuine adult liver-enriched genes in zebrafish via the genomic approach. Further analysis based on database mining revealed that majority of these liver-enriched genes in zebrafish share high homology with their mammalian counterparts, suggesting that the profile of liver-enriched genes in an adult zebrafish liver is similar to that identified in mammals and other vertebrates. Based on analysis of the zebrafish genome sequence available in both NCBI and ESEMBLE database, promoter regions for 51 genes were obtained and used for identification of putative binding sites for Hnf1, Hnf3, Hnf4 and Hnf6, four well-known liver-enriched transcription factors (Cereghini, 1996; Schrem et al., 2002). We found that, only in very rare cases a liver-enriched gene is solely regulated by a single of these four transcription factors. Majority of liver-enriched genes are most likely co-regulated by different combination of these transcription factors. We also examined the expression of all 129 adult liver-enriched genes during embryonic stages from 3–5 days post-fertilization (dpf) and found that more than half of these genes are also enriched in the embryonic liver. As an initial effort to study the function of these liver-enriched genes during the early stage of liver development, a novel gene, *liver-enriched gene 1* (*leg1*), was chosen from the list and was subjected to preliminary functional study via morpholino-mediated gene knock-down method. We found that *leg1* morphant conferred a smaller liver phenotype, demonstrating that *leg1* is essential for the expansion growth of the embryonic liver in zebrafish. Extensive efforts will be made to study if and how this group of genes is involved in controlling liver development in zebrafish in the future.

Materials and methods

Affymetrix zebrafish Genechip hybridization

Total RNA from the adult liver and the liver-free remaining body from three independent batches of zebrafish (wild type AB strain) was extracted using TRIzol (Gibco-BRL, USA), treated with DNaseI and purified through Qiagen RNeasy kit (Qiagen, Germany). cDNA synthesis, RNA probe labeling, GeneChip hybridization, washing and staining were performed following the manufacturer's instructions (Affymetrix, Santa Clara, California, USA). GeneChip arrays were scanned on an Affymetrix probe array scanner. Data were analyzed using a statistics software MAS5.0 from Affymetrix.

cDNA microarray hybridization

Total RNA from the adult liver and the liver-free remaining body (wild type AB strain) was extracted using TRIzol (Gibco-BRL, USA) followed by mRNA purification (PolyA Tract, *Promega*). The mRNA samples were labeled, reciprocally, with fluorescent dye Cy3 and Cy5. Generation of zebrafish cDNA microarray, microarray hybridization and microrray data analysis were performed as described (Lo et al., 2003; Wen et al., 2005).

RNA analysis

Ten micrograms of total RNA was separated on a formaldehyde gel and then transferred to nylon membrane (Hybond N+, Amersham). Probes for candidate genes were DIG-labeled (Roche Molecular Biochemicals) through PCR amplification using vector primer pairs T3/T7, respectively. For candidate genes that have matches in our own EST collection, the corresponding EST clones were retrieved and used for probe labeling. For 22 candidate genes

without match in our database, primers were designed based on available sequence (Supplementary Table S4) and used in RT-PCR to clone the corresponding gene fragment. The identity of the cloned gene was confirmed by DNA sequencing. Gel-blot hybridization was performed as described previously (Lee et al., 2002).

Search for Hnf1, Hnf3, Hnf4 and Hnf6 putative binding sites

Corresponding full-length cDNA sequences of genuinely liver-enriched genes were obtained by blasting NCBI database with EST sequences and were then used to blast against zebrafish genomic DNA sequence deposited either in NCBI or ESENMBL. The genomic sequence 5 kb upstream of the first base of each full-length cDNA was retrieved and used for searching binding site (Supplementary Table S2). Consensus sequence GTTAQTNNTY (Q: A/C; Y: T/C; N: any base) (Odom et al., 2004; Tronche et al., 1997) is used to search for Hnf1 binding sites, consensus sequence AWTRTTURYTY (W: A/T; R: A/G; U: G/T) (Cereghini, 1996) for Hnf3, consensus sequence GNNYAAQGUNC (Odom et al., 2004) for Hnf4 and consensus sequence ATTGAYTW (Odom et al., 2004; Samadani and Costa, 1996) for Hnf6. An in-house developed software 'Java Promoter Analysis Program' was used to perform the search. To use this program, 5′-regulatory genomic sequences of all 51 genes are deposited in a mini-database in FASTA format. For each binding consensus sequence, the program is designed to allow one or more base pairs in the consensus sequence to be replaced with anonymous characters as defined, for example, Y for C or T, W for A or T, N for A, C, G or T. The program is run to search through the min-database containing 5′-regulatory sequences for all possible combinations, both forward and reverse, of consensus sequence. Any match of a fragment of genomic sequence to the consensus sequence will be returned with further information (e.g. the position and sequence of the target binding site in the genomic sequence; the exact nature of the consensus sequence defined in this match). The starting position and ending position returned are in numeric format separated by symbol '|'. The minus symbol indicates the distance of the match found from the corresponding first base of each full-length cDNA sequence.

Whole-mount in situ hybridization

Embryos (wild type AB strain) at 3 dpf, 4 dpf and 5 dpf were collected and pre-treated and fixed as described (Chen et al., 2005). Purified plasmids was linearized by selected restriction enzymes and used as templates for in vitro transcription using T3 or T7 RNA polymerase (Roche Molecular Biochemicals) to generate DIG-labeled (Roche Molecular Biochemicals) sense and anti-sense probes. In situ hybridization was performed as described (Lo et al., 2003). Photos were taken under a Leica M216 optics.

leg1 morpholino injection and whole-mount in situ hybridization with three probes

Two morpholinos, one specifically targets the start codon ATG of the *leg1* gene (MO^{leg1}: CCATCTCAGACATCTAGCAGGACTG) and one is the mismatch control (MO^{mis}: CCATGTCACACATGTAGCACGAGTG), were designed and obtained from Gene Tools (Philomath, USA). 1 nL (0.5 mM) was injected into the one-cell-stage embryos each time. Embryos 4 days post-injection were collected and used for whole-mount in situ hybridization with triple markers. For three probe (triple-marker) whole-mount in situ hybridization, a *secreted immunoglobulin 4* probe (liver marker, clone ID: 127-D07-2) and an *insulin* probe (endocrine pancreas marker) were DIG-labeled while a *trypsin* probe (exocrine pancreas marker) was labeled with Fast Red (Roche) (Chen et al., 2005).

Results

Use of zebrafish genechip to identify candidate liver-enriched genes

To identify liver-enriched genes in zebrafish, total RNA samples from the adult female liver and the remaining body (liver-free body) were labeled with biotin and hybridized, respectively, to the zebrafish GeneChip (Affymetrix) carrying ~14,900 unigenes. Data from three independent experiments were analyzed using a statistics software MAS5.0 from Affymetrix. The logarithm base 2 of the signal ratio of liver vs. remaining body ≥3 in all three experiments was used as the cut-off value to identify candidate genes enriched in the liver. A total of 332 genes were identified as candidate liver-enriched genes. Because each of the probes on the zebrafish GeneChip is designed based on a Unigene cluster, the unigene accession number corresponding to each of these candidate liver-enriched genes was used to blast the NCBI Unigene database to examine its identity. We found that some of these 332 genes have been re-clustered and have been given new unigene accession numbers in the updated database (Supplementary Table S1). In fact, in many cases, genes previously under distinct unigene accession numbers are now clustered under the same unigene accession number (data not shown). After removing redundant genes based on updated unigene accession numbers, a total of 295 genes were identified as candidate liver-enriched genes (Supplementary Table S1). Blast search revealed that 227 genes can be assigned a biochemical function while 68 are represented by transcribed sequences only (Supplementary Table S1). Among these 227 genes, 93 genes encode for metabolic or catabolic enzymes, 52 for blood components, 19 for carrier/transporter proteins, 15 for proteins involved in protein synthesis, 8 for structural proteins and 5 for receptors. In addition, five transcription factors, including *hnf3*β, *hnf3*γ and *hnf4*, are also among the liver-enriched genes identified (Supplementary Table S1).

Identification of 129 genuine liver-enriched genes

We blasted all 295 candidate liver-enriched genes against our own EST database containing 31,000 ESTs (EMBO accession numbers AL901610–AL928536, CR391781–CR391903, CR926498–CR931631, CT010409–CT010425, CT025916, CT025917), representing ~17,000 unigene clusters (Lo et al., 2003). 126/295 genes hit one or more of their corresponding EST clones in our collection (e.g. Table 1, genes 1–73). Representative EST clones corresponding to these matched genes were retrieved from our EST stocks (Lo et al., 2003) and probes derived from these clones were used in RNA gel blot hybridization to confirm our gene chip hybridization result. Seventy-three out of 90 genes examined (accounting for 81%) were confirmed as genuine liver-enriched genes (Table 1, genes 1–73). In order to identify more genuine liver-enriched genes, additional 22 genes from the list of 295 candidate liver-enriched genes but having no match in our EST collection were cloned via RT-PCR and 18 of them were identified as genuine liver-enriched genes via RNA gel blot hybridization (Table 1, genes 74–91). As a result, 91 genes were identified as genuine liver-enriched genes using Affymetrix zebrafish GeneChip (Table 1; Fig. 1A). In an independent experiment, cDNA microarrays carrying ~9000 unigene clusters from our own EST collection were prepared (Lo et al., 2003;Wen et al., 2005) and used to hybridize with probes derived from RNA samples extracted from the liver or liver-free body to identify

Table 1
Identification of 129 Zebrafish liver-enriched genes via RNA Gel-blot hybridization

Affymetrix ID	UniGene ID	Clone ID	EST ID	mRNA size (kb)	Description
Dr.1064.1.S1_at	*Dr.1064*	*158-B09-1*	*AL907071*	*1.5*	*Glycine amidinotransferase*
Dr.1128.1.S1_at	*Dr.1128*	*045-A12-2*	*AL918232*	*1.9*	*Carboxypeptidase N, polypeptide 1*
Dr.11540.1.A1_at	*Dr.11540*	*151-G12-2*	*CT010409*	*4.4*	*Sterol carrier protein x*
Dr.11725.1.S1_at	*Dr.11725*	*086-B09-2*	*AL920792*	*2.8*	*Estrogen receptor 2b*
Dr.11729.1.S1_at	*Dr.11729*	*040-H10-2*	*AL917966*	*2.2*	*Cytochrome P450 (cyp2ja)*
Dr.1202.1.S1_at	Dr.28455	199-E09-2	AL928027	1.4	Similar to apolipoprotein A-I
Dr.12399.1.S1_at	*Dr.29718*	*159-C09-2*	*AL914072*	*1.9*	*HGF-like*
Dr.12454.1.S1_at	Dr.12454	108-C06-2	AL922229	1.4	GTP cyclohydrolase I feedback regulator
Dr.12551.1.S1_at	Dr.12551	042-D02-2	AL918038	2	FXYD type ion transport regulator 1
Dr.12949.1.S1_at	*Dr.12949*	*178-D07-2*	*AL926798*	*2.4*	*Type II iodothyronine deiodinase*
Dr.1307.1.S1_at	*Dr.1307*	*147-D12-2*	*CT010410*	*2.8*	*Forkhead box A3*
Dr.13239.1.S1_at	*Dr.13239*	*029-C07-2*	*AL917251*	*1.9*	*Unknown*
Dr.1449.1.S1_at	*Dr.51020*	*167-D09-1*	*AL926262*	*0.7*	*Liver basic fatty acid binding protein 10*
Dr.15161.1.S1_at	Dr.15161	137-G05-2	AL924291	2.1	Alpha-2-antiplasmin precursor
Dr.15192.1.A1_at	*Dr.15192*	*084-A04-2*	*AL920696*	*2.2*	*Cytochrome P450 (cyp4)*
Dr.1605.1.S1_at	Dr.28864	097-B07-1	AL921513	1.4	Ser/Cys proteinase inhibitor, clade A-1
Dr.16397.1.A1_at	Dr.39228	059-E03-2	AL919156	5	Alpha-2-macroglobulin precursor
Dr.1699.1.A1_at	Dr.1699	023-G03-1	CT010411	2	Liver pyruvate kinase
Dr.17459.1.S1_a_at	Dr.25043	020-A11-2	AL916653	3	Inter-α-trypsin inhibitor heavy chain 3
Dr.17515.1.A1_at	*Dr.31399*	*042-E09-1*	*AL918054*	*1.8*	*L-lactate dehydrogenase*
Dr.17618.1.A1_at	*Dr.17618*	*203-F06-2*	*AL928303*	*1.7*	*Kininogen II precursor*
Dr.1835.1.S1_at	*Dr.1835*	*071-A08-2*	*AL919914*	*3.6*	*Hepatocyte nuclear factor 4, alpha*
Dr.190.1.S1_at	*Dr.190*	*073-D02-2*	*CT010412*	*2, 2.3, 3*	*Complement component factor B*
Dr.19784.1.A1_at	*Dr.19784*	*150-F06-2*	*AL925121*	*1.9*	*Diamine acetyltransferase*
Dr.20185.1.S1_at	Dr.20185	138-C08-2	AL924333	3.6, 3.1	Lipoprotein lipase
Dr.20966.1.S1_at	*Dr.20966*	*068-H10-2*	*AL919785*	*1.4*	*Alcohol dehydrogenase 8a*
Dr.21064.1.S1_at	*Dr.21064*	*132-A12-2*	*AL923897*	*2.3*	*4-hydroxyphenylpyruvate dioxygenase*
Dr.21929.1.A1_at	Dr.21929	143-C02-2	AL924582	3.7	Lipoprotein B100 precursor
Dr.22129.1.A1_at	Dr.22129	010-F07-2	AL916138	1.5	Unknown
Dr.23036.1.A1_at	*Dr.23036*	*043-H02-2*	*AL918150*	*2*	*Guanine deaminase*
Dr.23043.1.A1_at	Dr.23043	157-C03-2	AL925560	1.7	Cytochrome P450 (2D20)
Dr.23788.2.A1_at	Dr.23788	258-A11-2	CR931040	1.8	Glutathione S-transferase pi
Dr.24233.1.S1_at	*Dr.24233*	*014-G06-2*	*AL916390*	*8*	*Fibronectin 1b*
Dr.2426.1.S1_at	*Dr.2426*	*135-D07-2*	*AL924131*	*1.2*	*Inter-alpha-trypsin inhibitor light chain*
Dr.2433.1.A1_at	Dr.2433	025-H05-1	CT010414	1.8, 2.3	Liver carboxylesterase-2
Dr.2452.1.A1_at	Dr.45463	026-A11-2	AL910873	2.2	Complement component C9 precursor
Dr.24953.1.S1_at	Dr.29079	018-G04-2	AL916586	1	Apolipoprotein C-II precursor
Dr.25009.1.S1_at	*Dr.25009*	*170-E05-2*	*AL926438*	*m.b*	*Vitellogenin 1*
Dr.25009.4.A1_at	Dr.31982	179-B01-1	AL915046	6	Vitellogenin 2
Dr.25207.1.A1_at	*Dr.25207*	*031-D09-2*	*AL917381*	*1.1*	*Complement component C8 gamma chain*
Dr.25214.1.A1_at	Dr.25214	109-D10-2	AL922321	2.4	Vitamin D 24-hydroxylase
Dr.25679.1.S1_at	Dr.28289	098-E02-2	AL914012	1.5	Solute carrier family 1, member 3
Dr.25776.1.S1_at	Dr.25776	183-C03-2	AL927052	2.5	Vitamin-K dependent protein C precursor
Dr.2596.3.A1_at	Dr.2596	249-H11-2	CR930360	1.8	Betaine-homoCys S-methyltransferase
Dr.26467.1.S1_at	Dr.26467	150-C03-2	AL925094	3.6, 1.8	Hepatic lipase
Dr.2961.1.S1_at	Dr.2961	157-E02-2	AL925577	1.3	Retinol dehydrogenase 1-like
Dr.3004.1.A1_at	*Dr.29754*	*131-H05-2*	*AL923875*	*1.4*	*Unknown*
Dr.3498.1.S1_at	*Dr.3498*	*093-C08-2*	*AL913938*	*2.8*	*Methionine adenosyltransferase Iα*
Dr.3585.1.S1_at	*Dr.3585*	*090-E07-2*	*AL921089*	*1.5*	*Angiotensinogen*
Dr.3613.1.S1_at	*Dr.3613*	*123-E05-2*	*AL923291*	*3.7*	*Ceruloplasmin*
Dr.4111.1.S1_at	*Dr.4111*	*168-G06-2*	*AL926340*	*1.5*	*Fructose-1,6-bisphosphatase 1*
Dr.4199.1.S1_at	*Dr.4199*	*033-D10-2*	*AL917508*	*2.3, 3.5*	*Serine/cysteine proteinase inhibitor C1*
Dr.4797.1.S1_at	Dr.4797	052-D09-2	AL918727	1.4	26–29kD-proteinase protein
Dr.4867.1.A1_at	Dr.4867	102-G12-2	AL921891	1.1	Haptoglobin-related protein precursor
Dr.4907.1.S1_at	*Dr.4907*	*021-B02-2*	*AL916718*	*2*	*Fibrinogen gamma-A chain precursor*
Dr.4938.1.S1_at	Dr.4938	127-F03-2	AL914933	1.8	Fatty acid desaturase 2
Dr.5462.1.S1_at	*Dr.49798*	*120-C08-2*	*AL912228*	*1.2*	*RPE-retinal G protein-coupled receptor*
Dr.5467.1.A1_at	Dr.49103	072-B09-2	AL919988	2.2	Guanosine monophosphate reductase
Dr.5479.1.S1_at	*Dr.5479*	*137-E09-2*	*AL924278*	*1.1*	*Plasma retinol binding protein 4*
Dr.5674.2.S1_at	Dr.47298	147-C12-1	AL910993	1	Apolipoprotein C-I precursor
Dr.6787.1.S1_at	*Dr.6787*	*156-B09-1*	*AL913919*	*1.4*	*Noelin precursor*

(continued on next page)

Table 1 (*continued*)

Affymetrix ID	UniGene ID	Clone ID	EST ID	mRNA size (kb)	Description
Dr.6924.1.S1_at	*Dr.6924*	*055-F01-2*	*AL912909*	*1.9*	*S-adenosylhomocysteine hydrolase*
Dr.7171.1.S1_at	Dr.7171	188-A03-2	AL927325	2.4	Glutamic pyruvate transaminase 2
Dr.742.1.S1_at	*Dr.742*	*077-A02-2*	*AL920293*	*3.9*	*TGF β-induced protein BIGH3 precursor*
Dr.7559.1.S1_at	Dr.7559	033-B08-2	AL917493	1	Nuclear protein 1
Dr.7634.1.A1_at	Dr.31376	004-C10-2	AL915791	2.2	Hemopexin
Dr.7648.1.A1_at	Dr.49819	167-A10-2	AL926235	1.4	Zona pellucida glycoprotein 3
Dr.8149.1.A1_at	*Dr.8149*	*179-H04-1*	*AL908391*	*1*	*IGF binding protein 2*
Dr.845.1.A1_at	*Dr.845*	*044-E07-2*	*CT010417*	*2.5*	*Fibrinogen alpha chain*
Dr.8505.1.S1_at	Dr.29757	130-D01-2	AL923737	1.4	Alpha-2-HS-glycoprotein precursor
Dr.852.1.S1_at	*Dr.852*	*074-D02-2*	*AL920131*	*3*	*Inter-alpha (globulin) inhibitor H2*
Dr.876.1.S1_at	Dr.876	179-D11-2	CT010418	1.1	60S ribosomal protein L5-like
Dr.9478.2.S1_at	Dr.9478	074-D08-2	AL920136	3.5	Myeloid-specific peroxidase
DrAffx.2.68.S1_at	Dr.30460	067-D09-2	AL919697	4, 2.7	Cytochrome P450 monooxygenase (2j2)
Dr.13681.1.S1_at	Dr.21117	A1		1	Apolipoprotein M
Dr.3073.1.A1_at	*Dr.36665*	*A2*		*1.8*	*Serpin peptidase inhibitor A7*
Dr.17693.1.A1_at	Dr.33915	A3		2.1	40S ribosomal protein S6
Dr.10717.1.S1_at	*Dr.10717*	*A6*		*3.6, 2.0*	*Estrogen receptor 1*
Dr.5528.1.S1_at	Dr.52334	A8		2	Fetuin-B precursor
Dr.13466.1.A1_at	Dr.13466	A11		2	Vitronectin
Dr.4103.1.A1_at	*Dr.17536*	*A12*		*1.9, 1.4*	*Translationally controlled tumor protein*
Dr.16871.1.A1_x	Dr.43856	A13		2.3	Protein disulfide-isomerase
Dr.18834.1.S1_at	*Dr.18834*	*A14*		*2.1*	*Urate oxidase*
Dr.3645.1.S1_at	*Dr.3645*	*A15*		*4.4, 5.5*	*Plasminogen*
Dr.8491.1.A1_at	Dr.8491	A16		2.4	Plasma hyaluronan-binding protein
Dr.10788.1.S1_at	Dr.10788	A17		1.5	Nothepsin
Dr.22229.1.A1_at	Dr.22229	A19		2.6	Unknown
Dr.7516.1.A1_at	*Dr.7516*	*A20*		*1.6*	*Amino acid transporter system 3*
Dr.17437.1.S1_at	Dr.17437	A21		1.4	C1q-related factor precursor
Dr.2655.1.S1_at	Dr.31393	A22		2.7	Nucleoprotein TPR
Dr.17561.1.S1_at	Dr.17561	A23		0.8	Unknown
Dr.1889.1.S1_at	Dr.1889	A26		2.2	Transferrin-a
	Dr.19224	142-F03-1	AL908432	1.6	Aldolase b, fructose-bisphosphate
	Dr.1232	*012-G11-2*	*AL916272*	*1.3*	*Reticulocalbin*
	Dr.12340	*008-C12-1*	*CT010419*	*1.5*	*Zygote arrest 1*
	Dr.12889	116-A01-2	AL910894	1.4	Dynein light chain-A
	Dr.19527	142-C09-2	CT010420	2.1	Unknown
	Dr.20124	*053-H11-1*	*CT010421*	*1.5*	*SRY-box 4a*
	Dr.25683	098-D03-1	AL911218	1.2	Cathepsin Lb precursor
	Dr.26431	024-H05-1	CT025917	3.2	Retinal short-chain dehydrogenase
	Dr.26855	*127-A09-2*	*AL923515*	*3.1, 2.4*	*Coagulation factor II (thrombin)*
	Dr.27412	*127-D06-2*	*AL913602*	*1.4*	*Secreted immunoglobulin 4 precursor*
	Dr.28645	018-E05-1	AL916572	1.4	Purine nucleoside phosphorylase
	Dr.31382	049-B06-1	AL918506	1.4	Vestigial like 4
	Dr.32172	160-A01-2	CT025916	1.7	Acetyl-Coenzyme A acyltransferase 1
	Dr.33932	163-G06-2	AL926009	2	Complement component 8α
	Dr.34567	061-B02-2	AL919280	2.3, 4	NF-kappaB inhibitor alpha
	Dr.4338	*004-F08-2*	*AL915818*	*0.9*	*Versican precursor*
	Dr.4865	024-B03-2	AL916916	9.4, 3.4, 2	Ser/Cys proteinase inhibitor D1
	Dr.5628	027-A08-1	AL903660	1.4	Heterogeneous ribonucleoprotein A0
	Dr.6829	068-C01-1	AL908628	1.4	L-myc-1 proto-oncogene protein
	Dr.7551	076-E10-2	AL920263	1.5	Cytochrome P450 2d6
	Dr.8171	*046-A03-2*	*AL918291*	*1.6*	*Presenilin 2*
	Dr.9875	110-C02-2	AL922371	1.7	Unknown
		190-F09-2	AL911853	1.7	Component in TH receptor complex
		010-B12-2	CT010422	1	DRE1 protein
		045-D04-2	CT010423	2.4	Unknown
		090-A09-2	AL921055	0.6	Nedd4 binding protein 3
		093-D01-2	AL921269	1	Unknown
		130-D05-2	AL923741	1.3	Paired box protein
		134-F04-2	AL924067	3	Unknown
		144-G11-2	AL915157	1.6	Nicotinic acetylcholine receptor beta 7

Table 1 (*continued*)

Affymetrix ID	UniGene ID	Clone ID	EST ID	mRNA size (kb)	Description
		148-D12-2	AL924964	4.5	Apobec-1 complementation factor isoform
		159-H09-2	AL925766	2.1	Similar to alpha-2-antiplasmin precursor
		185-E11-2	AL927183	2.3	Unknown
		193-H07-2	AL927686	2	2-oxoisovalerate dehydrogenase β subunit
		202-G04-2	AL928237	2.5	Similar to protein kinase C, eta type
		149-A04-2	CT010425	2.4	Phosphatidic acid phosphatase type 2B
		122-F03-2	AL923223	2.2	Cytochrome P450 24a1

Genes 1–91 were identified based on microrray hybridization on Affymetrix GeneChip and genes 92–129 were obtained based on hybridization on our homemade cDNA microarray. These 129 genes were confirmed as liver-enriched genes via RNA gel blot hybridization with probes either derived from EST clones (1–73, 92–129) or cloned gene fragment (74–91). Affymetrix probe ID, UniGene accession number (ID), EST clone ID and corresponding EST accession number are provided. Transcript sizes were obtained based on mRNA migration on an RNA gel blot with reference to molecular size marker. Genes also enriched in the embryonic liver are shaded in grey. Genes subjected to promoter analysis are italicized. m.b: multiple bands.

candidate liver-enriched genes. Probes derived from these candidate genes were also used in RNA gel blot hybridization and 38 genes that were not represented in the Affymetrix GeneChip were identified as additional genuine liver-enriched genes (Table 1, genes 92–129). Combining these two sets of data, we identified a total of 129 genes as genuine liver-enriched genes in the adult zebrafish. Majority of these liver-enriched genes encode enzymes involved in metabolism and catabolism (40/129) while a significant number encode blood components (37/129) and 11 genes are represented by transcribed sequence only (Table 1). Analysis of RNA gel blot hybridization patterns revealed that 116 genes showed single hybridization band. The remaining 13 genes showed two or more hybridization bands, suggesting the presence of multiple splicing products or closely related homologues for these genes. Based on the ratio of their transcript levels in liver vs. liver-free body, these 129 genes can be further divided into three groups (Figs. 1A, B). The first group contains 54 genes whose expression is apparent in the liver but undetectable or barely detectable in the liver-free body and these genes are arbitrarily considered as liver-specific genes, including genes *liver fatty acid binding protein 10*, *vitellogenin 2*, *fibrinogen gamma-A chain*, *fibrinogen alpha-E chain*, *glutamic pyruvate transaminase 2*, *complement component factor B*, etc. (Figs. 1A, B). The second group contains 54 genes that displayed an expression ratio (liver vs. liver-free body) between 6.9:1 and 2:1 and these genes are considered to be highly liver-enriched genes, including genes *alcohol dehydrogenase 8a*, *carboxylesterase-2*, *lipoprotein lipase*, *apolipoprotein M*, *nothepsin*, *hnf4*, etc. (Figs. 1A, B). The third group contains 21 genes displaying an expression ratio less than 2:1 and these genes are considered as moderately liver-enriched genes, including genes *foxA3*, *TGF-β induced protein*, *vitronectin*, *diamine acetytransferase*, etc. (Figs. 1A, B).

Search for Hnf binding sites in the promoters of liver-enriched genes

Liver is responsible for the production of many metabolic and catabolic enzymes and of variety of blood components. Six families of liver-enriched transcription factors, namely HNF-1, HNF-3, HNF-4, HNF-6, C/EBP and D-binding protein (DBP), are known to control the expression of a vast number of liver-specific or -enriched genes (Cereghini, 1996; Odom et al., 2004; Schrem et al., 2002, 2004). To gain information about the possible mechanism controlling the expression of these 129 liver-enriched genes, we blasted the sequences of all 129 liver-enriched genes against the NCBI unigene database and non-redundant (nr) database and were able to obtain full-length cDNA sequences corresponding to 77 genes (Table 2). Full-length cDNA sequences were then used to blast the genomic sequence deposited in the NCBI non-redundant database (nr) and ESEMBL database and genomic sequences were found available for 51 genes (Supplementary Tables S2 and S3). Genomic sequence 5 kb upstream of the first base of each full-length cDNA sequence was retrieved (Supplementary Table S2) and used for searching for Hnf1, Hnf3, Hnf4 and Hnf6 binding sites. In total, promoters of 50 out of 51 genes were found to contain putative binding sites for one or more of these four transcription factors (Supplementary Table S2). Searching for Hnf1 binding sites using consensus sequence GTTAQTNNTY (Q: A/C; Y: T/C; N: any base) (Odom et al., 2004; Tronche et al., 1997) identified 48 putative sites in the promoter regions of 29 genes, including well known genes *α1-antitrypsin*, *insulin-like growth binding protein*, etc. (Schrem et al., 2002) (Table 2; Supplementary Table S2). Searching for Hnf3 binding sites using consensus sequence: AWTRTTURYTY (W: A/T; R: A/G; U: G/T) (Cereghini, 1996) identified a total of 74 putative sites in 34 genes, including genes *vitellogenin 1* and *fructose-1,6-bisphosphatase* (Schrem et al., 2002) (Table 2; Supplementary Table 2). Searching for Hnf4 binding sites using consensus sequence GNNYAAQGUNC (Odom et al., 2004) identified 26 putative sites in 22 genes, including plasma genes *retinol binding protein*, *complement component C8*, etc. (Schrem et al., 2002) (Table 2; Supplementary Table 2). Searching for Hnf6 binding sites using consensus sequence ATTGAYTW (Odom et al., 2004; Samadani and Costa, 1996) identified 90 putative sites in 44 genes, including genes *CYP2J* and *CYP4* (Schrem et al., 2002) (Table 2; Supplementary Table 2). Previous studies have shown that the expression of many liver-specific or -enriched genes is often co-regulated by more than one hepatocyte nuclear factors (Odom et al., 2004; Schrem et al., 2002). Cross-comparison of genes containing binding sites for each of these four transcription

factors showed that, with respect to genes regulated by a single factor, none (out of 22 gene) is found by Hnf4, only one (out of 29) by Hnf1 (*presenilin 2*), two (out of 34) by Hnf3 (*zygote arrest 1* and *kininogen II*) and 8 (out of 44) by Hnf6 (including *diamine acetyltransferase*, *urate oxidase*, etc.), suggesting that Hnf6 probably functions more independently

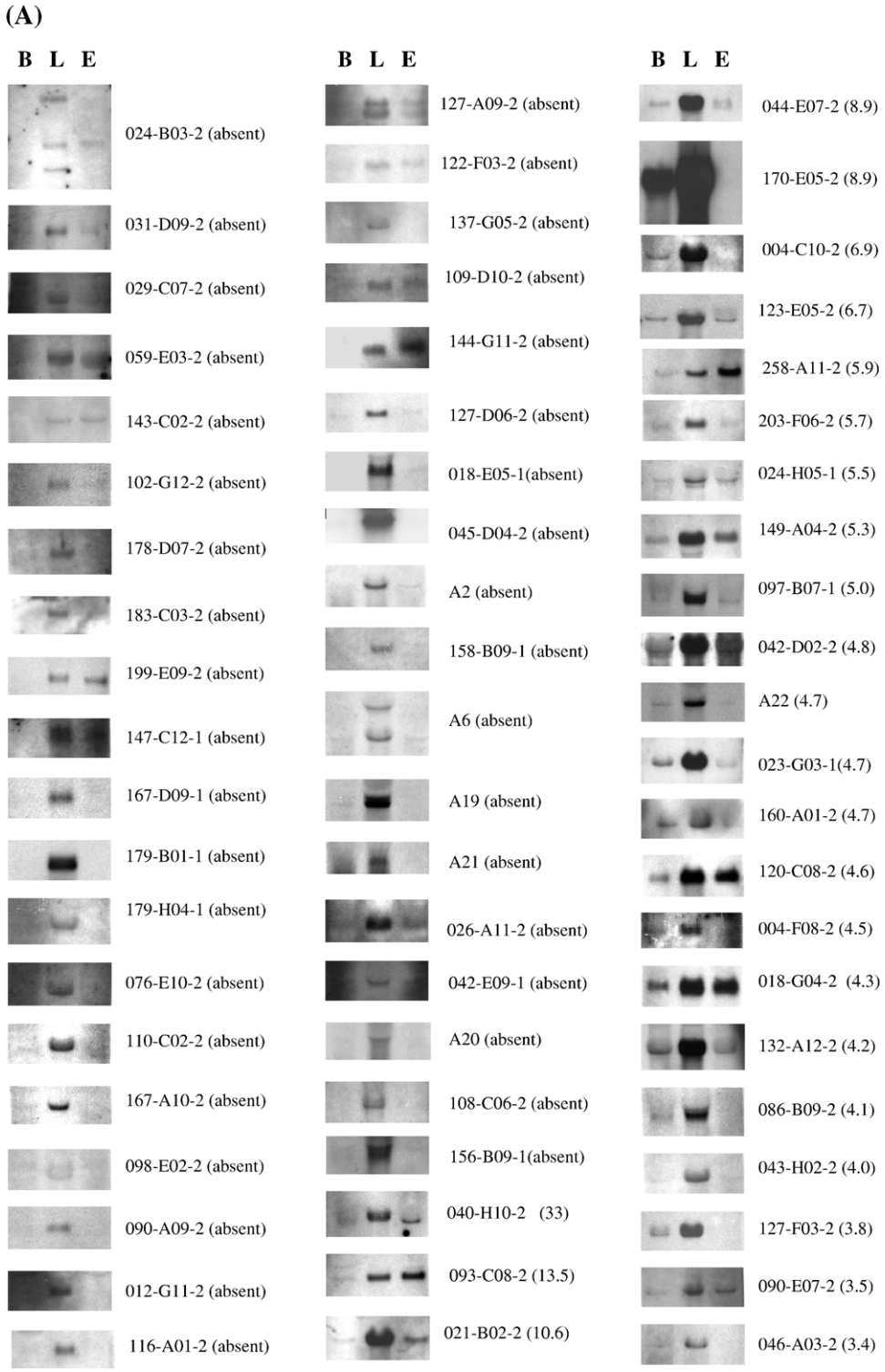

Fig. 1. Classification of liver-enriched genes in zebrafish. RNA samples were extracted from the dissected liver and liver-free body, respectively. In panel A, RNA samples from 3 dpf embryos were also included. Probes were derived from EST clones or gene fragments cloned in plasmid vector via RT-PCR using gene-specific primers (Table 1; Supplementary Table 4). RNA hybridization blots were arranged according to the descending ratio of expression levels of liver vs. liver-free body (as indicated in the bracket). We arbitrarily grouped this set of genes according to the ratio of expression level. Genes with ratios more than 6.9 are considered as liver-specific genes, ratios between 6.9 and 2 as highly-liver-enriched genes and ratios less than 2 as moderately-liver-enriched genes. Clone identification is shown on the right side. Hybridization intensity was quantified using software Molecular Analyst (Bio-rad). B, liver-free body; L, liver; E, 3 dpf embryos.

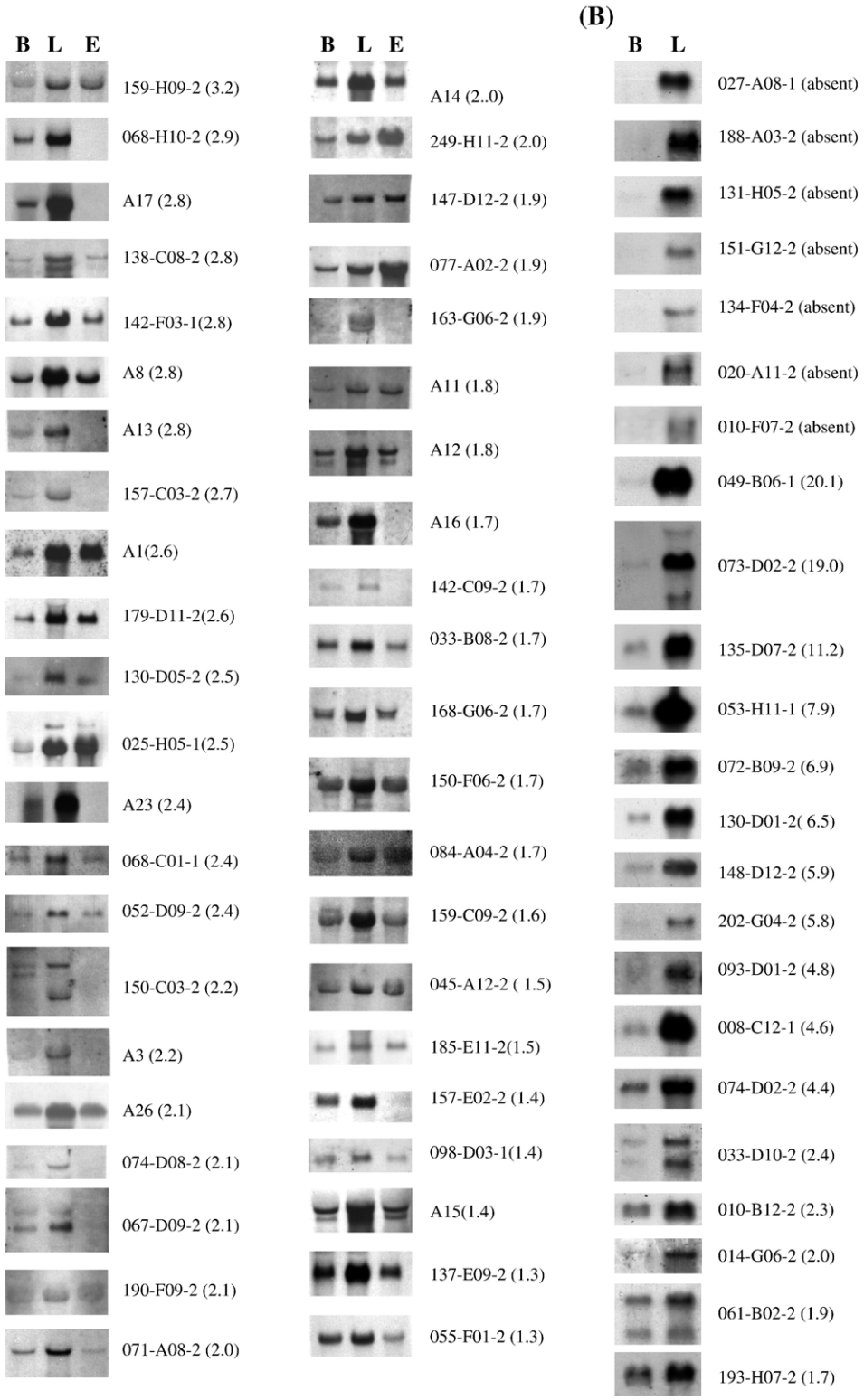

Fig. 1 (*continued*).

or is a more general transcription factor (Table 2). Further analysis showed that 8 out of 47 genes are co-regulated by pairing of these four factors, including one gene (*noelin*) by Hnf1 and Hnf3, one gene (*carboxypeptidase N*) by Hnf1 and Hnf6, three genes (*glycine amidinotransferase*, *angiotensinogen* and *sox4*) by Hnf3 and Hnf6, and three genes (*L-lactate dehydrogenase*, *cytochrome oxidase II* and *S-adenosylhomocysteine hydrolase*) by Hnf4 and Hnf6 (Table 2). We then examined genes co-regulated by any three of these four factors and found that 2 genes are co-regulated by Hnf1, Hnf3 and Hnf4, 3 genes by Hnf1, Hnf4 and Hnf6, 5 genes by Hnf3, Hnf4 and Hnf6 while 12 genes are co-regulated by Hnf1, Hnf3

Table 2
Liver genes putatively co-regulated by the network of four HNF factors

Factor(s)	Clone ID	EST ID	FLcDNA ID	Description
11 genes regulated by a single HNF factor				
HNF1	046-A03-2	AL918291	41107586	Presenilin 2
HNF3	203-F06-2	AL928303	53734154	Kininogen II precursor
HNF3	008-C12-1	CT010419	62531328	Zygote arrest 1
HNF6	151-G12-2	CT010409	38969930	Sterol carrier protein x
HNF6	147-D12-2	CT010410	2982340	Forkhead box A3
HNF6	084-A04-2	AL920696	34785184	Cytochrome P450 (cyp4)
HNF6	150-F06-2	AL925121	56269422	Diamine acetyltransferase
HNF6	033-D10-2	AL917508	51858526	Serine/cysteine proteinase inhibitor C1
HNF6	A14		49900518	Urate oxidase
HNF6	A20		53734166	Amino acid transporter system 3
HNF6	127-D06-2	AL913602	45739320	Secreted immunoglobulin 4 precursor
8 genes co-regulated by the pair combination of the four HNF factors				
HNF1,3	156-B09-1	AL913919	50418462	Noelin precursor
HNF1,6	045-A12-2	AL918232	44890357	Carboxypeptidase N, polypeptide 1
HNF3,6	158-B09-1	AL907071	34784848	Glycine amidinotransferase
HNF3,6	090-E07-2	AL921089	63102190	Angiotensinogen
HNF3,6	053-H11-1	CT010421	34452015	SRY (sex determining region Y)-box 4
HNF4,6	042-E09-1	AL918054	49900819	L-lactate dehydrogenase
HNF4,6	131-H05-2	AL923875	38328392	Unknown
HNF4,6	055-F01-2	AL912909	37681724	*S*-adenosylhomocysteine hydrolase
22 genes co-regulated by triple combination of the four HNF factors				
HNF1,3,4	135-D07-2	AL924131	33416402	Inter-alpha-trypsin inhibitor light chain
HNF1,3,4	A2		53734453	Serpin peptidase inhibitor A7
HNF1,3,6	029-C07-2	AL917251	49903797	Unknown
HNF1,3,6	167-D09-1	AL926262	51980431	Liver basic fatty acid binding protein 10
HNF1,3,6	071-A08-2	AL919914	34194037	Hepatocyte nuclear factor 4, alpha
HNF1,3,6	014-G06-2	AL916390	51949770	Fibronectin 1b
HNF1,3,6	170-E05-2	AL926438	15290748	Vitellogenin 1
HNF1,3,6	123-E05-2	AL923291	39795309	Ceruloplasmin
HNF1,3,6	120-C08-2	AL912228	42744550	RPE-retinal G protein-coupled receptor
HNF1,3,6	077-A02-2	AL920293	27127276	TGF beta-induced protein BIGH3 precursor
HNF1,3,6	074-D02-2	AL920131	28279171	Inter-alpha (globulin) inhibitor H2
HNF1,3,6	A6		10944301	Estrogen receptor 1
HNF1,3,6	A12		9802398	Translationally controlled tumor protein
HNF1,3,6	A15		37589179	Plasminogen
HNF1,4,6	178-D07-2	AL926798	37589727	Type II iodothyronine deiodinase
HNF1,4,6	068-H10-2	AL919785	41223379	Alcohol dehydrogenase 8a
HNF1,4,6	093-C08-2	AL913938	28278851	Methionine adenosyltransferase Iα
HNF3,4,6	159-C09-2	AL914072	54038374	HGF-like
HNF3,4,6	043-H02-2	AL918150	63102011	Guanine deaminase
HNF3,4,6	044-E07-2	CT010417	49900365	Fibrinogen alpha chain
HNF3,4,6	127-A09-2	AL923515	33416400	Coagulation factor II (thrombin)
HNF3,4,6	004-F08-2	AL915818	45259474	Versican precursor
9 genes co-regulated by all four HNF factors				
HNF1,3,4,6	086-B09-2	AL920792	23466358	Estrogen receptor 2b
HNF1,3,4,6	040-H10-2	AL917966	13435015	Cytochrome P450 (cyp2ja)
HNF1,3,4,6	073-D02-2	CT010412	1015969	Complement component factor B
HNF1,3,4,6	031-D09-2	AL917381	38541963	Complement component C8 gamma chain
HNF1,3,4,6	168-G06-2	AL926340	31419555	Fructose-1,6-bisphosphatase 1
HNF1,3,4,6	021-B02-2	AL916718	28279122	Fibrinogen gamma-A chain precursor
HNF1,3,4,6	137-E09-2	AL924278	49902661	Plasma retinol binding protein 4
HNF1,3,4,6	179-H04-1	AL908391	33604151	IGF binding protein 2
HNF1,3,4,6	012-G11-2	AL916272	47937869	Reticulocalbin

EST ID: EST accession number in GeneBank; FlcDNA ID: full-length cDNA accession number in GeneBank.

and Hnf6 (Table 2). Finally, we found that 9 genes, including *complement factor B* and *C8*, *fibrinogen γ-A*, *estrogen receptor 2b*, *plasma retinol binding protein 4*, etc., are co-regulated by all four transcription factors (Table 2). However, limiting the search for Hnf binding sites to 5 kb promoter region will no doubt miss out identifying Hnf binding sites that are contained in the enhancers, therefore, more number of genes are expected to be co-regulated by these four factors. In

addition, it is necessary to point out that the identified Hnf binding sites are only putative ones and experimental work has to be performed to decide which ones are the genuine binding sites in the future.

Identification of 69 genes enriched in the embryonic liver

The wild type zebrafish liver bud appears at ~44 h post-fertilization (hpf) and undergoes rapid expansion to form the left and right lobes between 3 and 5 days post-fertilization (dpf) (Field et al., 2003; Ober et al., 2003; Wallace and Pack, 2003). In mammals, some adult liver-enriched genes (such *albumin* and *fetolipoprotein*) are expressed even before hepatocytes are fully differentiated during embryogenesis (Lee et al., 2005; Zhao et al., 2005), probably reflecting the needs of some of the liver function during the early stage of liver development. During the course of confirming our microarray hybridization results via RNA gel bolt hybridization, we found that many adult liver-enriched genes are also expressed in the embryos at 3 dpf (Fig. 1A). To investigate if any mature liver-enriched genes are also expressed in the embryonic liver, probes derived from these 129 liver-enriched genes were used to perform whole-mount in situ hybridization on 3 dp, 4 dp and 5 dpf embryos. In total, 66 genes were found to be expressed in the embryonic liver from 3 dpf onwards (Supplementary Fig. S1). Three genes, namely *p450-like*, *retinol dehydrogenase 1-like* and *nucleoprotein TPR*, respectively, are not detectable at 3 dpf but are detected in the embryonic liver from 4 dpf onwards (Supplementary Fig. S1). Detailed analysis of the expression pattern showed that majority of these embryonic liver-enriched genes (57/69) was specifically enriched in the liver but not in other digestive organs (Fig. 2A). 6/69 genes, including *liver pyruvate kinase*, *HNF4α*, *glutathione S-transferase*, *retinol dehydrogenase 1-like*, *fructose-1,6-bisphosphatase 1* and *cytochrome P450 monooxygenase*, are enriched both in the liver and gut (Fig. 2B). Interestingly, 6/69 genes, including *glycine amidinotransferase*, *S-adenosylhomocysteine hydrolase*, *nuclear protein 1*, *60S ribosomal protein L5-like* and *aldolase b*, are enriched not only in the liver but also in pancreas and gut as well (Fig. 2C). In addition, expression of a few genes is also found in organs other than digestive organs. For example, as predicted, besides in the liver, two genes, namely *noelin* and *insulin growth factor-like binding protein*, are also expressed in the head region (Fig. 2A) (Barembaum et al., 2000; Wood et al., 2005).

Novel gene leg1 is essential for liver expansion growth

Among the 69 embryonic liver-enriched genes identified, we noticed that EST 131-H05-2 is clustered under the zebrafish Unigene number Dr.29754. Full-length cDNA sequence corresponding to the EST 131-H05-2 was obtained via 5′-race and 3′-race (EMBO accession number CT737258). This gene encodes a peptide of 361 amino acids and blast search and amino acid sequence alignment revealed that it has homologues in mouse (accession number gi12844359; 31% identity) and human (accession number gi60502446; 30% identity) (Fig. 3A). This group of proteins has never been characterized previously and is thus unknown of function. We designated the corresponding gene as *liver-enriched gene 1* (*leg1*). RNA gel blot hybridization showed that *leg1* is undetectable in embryos until 17 h post-fertilization (hpf). *leg1* starts to express at 24 hpf and remains at relative high levels thereafter (Fig. 2B). Whole-mount in situ hybridization showed that *leg1* starts to be enriched in the liver at 2 dpf (Fig. 3C). To study if *leg1* is involved in controlling liver development, an antisense morpholino oligos specifically targeting the translation start codon ATG (MO^{leg1}) was designed and injected into embryos at the single-cell-stage. Whole-mount in situ hybridization using the liver-specific marker *secreted immunoglobulin 4* (Fig. 2A, probe 127-D06-2) showed that majority of MO^{leg1} morphants (35 out of 42 morphants examined, accounting for 83%) 4 days post-injection displayed a smaller liver when compared with embryos injected with a miss-match control morpholino (MO^{mis}) (Supplementary Fig. S2). Co-injection of *leg1* mRNA with MO^{leg1} nullified the effect of MO^{leg1} and majority of injected embryos (26 out of 38 embryos examined, accounting for 68%) appeared relatively normal, suggesting that the smaller liver phenotype observed in MO^{leg1} morphants was likely due to the knock-down of Leg1 expression in these morphants (Fig. 3D). Thus, Leg1 is a novel protein that is important for liver development in zebrafish. Interestingly, whole-mount triple-marker in situ hybridization using a *trypsin* probe revealed that the MO^{leg1} morphant also conferred a smaller exocrine pancreas while the islet appeared normal when examined with an *insulin* probe, suggesting that Leg1 somehow also regulates the development of the exocrine pancreas in zebrafish (Fig. 3D).

Discussion

Successful progression from liver initiation and development to form a functional mature liver is a dynamic process and is accomplished by precise coordination of gene expression (Jochheim et al., 2003; Petkov et al., 2000, 2004). While many genes responsible for liver initiation and development are expressed temporally during liver organogenesis, some of them will continue to be expressed to maintain the status of liver stem cells or other cell types. The mature liver expresses a vast range of genes encoding metabolic enzymes and blood components to perform its physiological role in the body. It is well-known that defects in some individuals of these enzymes or blood components in mammals will cause diverse liver- and/or blood-related diseases (Hillebrandt et al., 2005; Wolfrum et al., 2005). It is our great interest to ask if mature liver-enriched genes including those encoding metabolic enzymes and blood components are also important for liver development. However, due to the fact that embryogenesis in mammals is completed intrauterinally, there is limited genetic data to show if this group of genes plays a role during liver organogenesis. We reported here the identification of 129 genes enriched in the adult liver in zebrafish. Based on their relative expression levels in the liver tissue vs. the liver-free body via RNA gel blot hybridization, these genes can be divided into three groups: 54 liver specific

genes, 54 highly-liver-enriched genes and 21 moderately-liver-enriched genes. Detailed analyses showed that a great number of these 129 genes are well known liver-enriched genes encoding a variety of enzymes for metabolism or blood components for maintaining homeostasis in human and other vertebrates. Because the early embryogenesis in zebrafish happens externally and because morpholino-mediated gene knock-down technique has been well established, these mature liver-enriched genes can be subjected to functional studies for their roles during liver organogenesis. Prior to using the morpholino-mediated gene knock-down strategy to address if genes encoding metabolic enzymes and blood components function dually during liver development, we first need to determine if these adult liver-enriched genes are also enriched in the embryonic

Fig. 2. Identification of 69 genes enriched in the embryonic liver. Probes were derived from all 129 adult liver-enriched genes and were used to perform whole-mount in situ hybridization on 3 dpf, 4 dpf and 5 dpf embryos, respectively. (A) Expression of 56 genes is predominantly enriched in the liver but not in other digestive organs. (B) Expression of six genes is enriched in the liver and intestine but not in the pancreas. (C) Expression of six genes is enriched in all three major digestive organs. Clone identification is shown on the right corner. The liver is indicated by black arrow, the intestine by red arrow and the pancreas by green arrow.

Fig. 2 (*continued*).

liver. In situ hybridization showed that the expression of 69 genes is enriched in the embryonic liver, including 22 genes encoding enzymes, 29 genes for blood components and 7 genes uncharacterized previously. Thus, these genes are candidate genes for functional studies of their roles during liver organogenesis via morpholino-mediated gene knock-down method (Nasevicius and Ekker, 2000) and eventually via TILLING (targeting induced local lesions in genomes) strategy for the target-selected mutagenesis of the interested genes (Wienholds et al., 2003) in our future work. As an initial effort, antisense morpholino specifically targeting the novel gene *leg1* was designed and injected into the one-cell-stage embryos and the resultants (morphants) displayed a much smaller liver when compared with that in a control morphant derived from injection of MO^{mis}, demonstrating that *leg1* is important for the liver expansion growth. This preliminary result will prompt us to identify *leg1* mutant via TILLING strategy in the future. No doubt, many other embryonic liver-enriched genes responsible

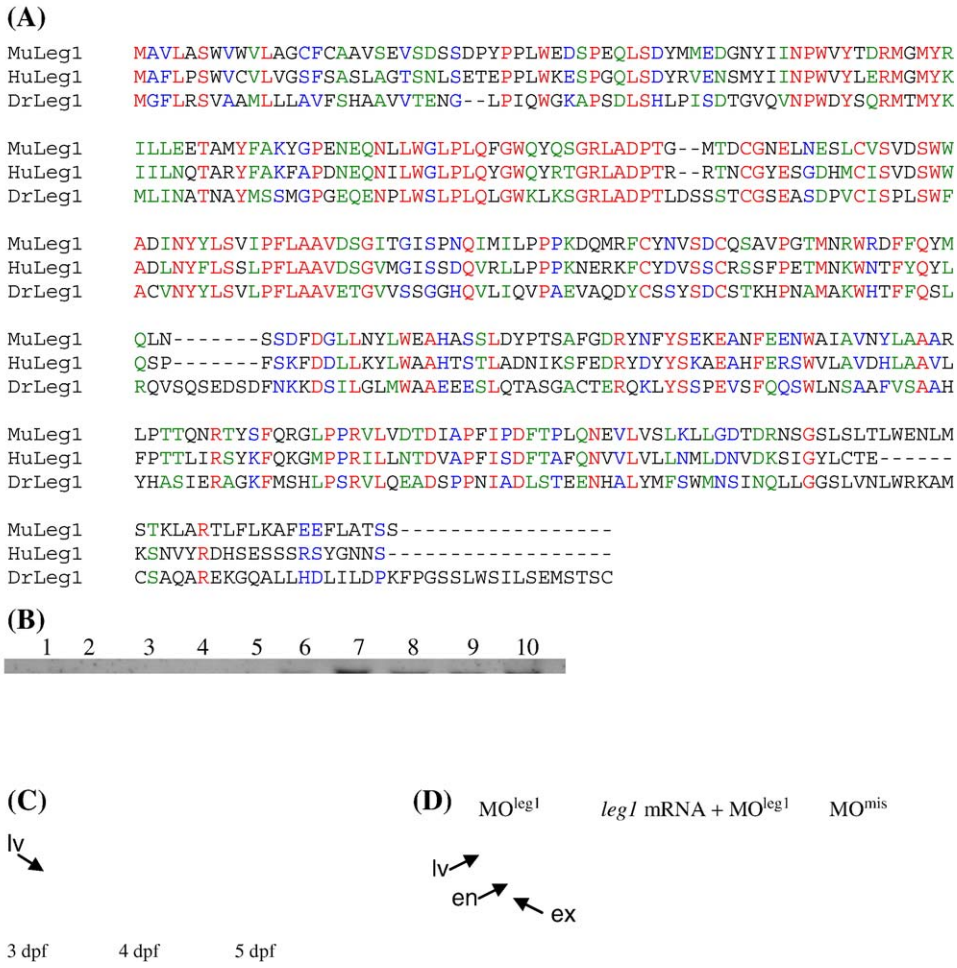

Fig. 3. Leg1 functions as a novel factor involved in controlling liver expansion growth. (A) Amino acid sequence alignment of Leg1 (DrLeg1, CT737258) with its homologues in human (HuLeg1, gi60502446) and mouse (MuLeg1) gi12844359). (B) RNA gel blot hybridization revealed that *leg1* is undetectable before 17 hpf (lanes 1–5: 1-cell-stage, 5 hpf, 9 hpf, 12 hpf and 17 hpf, respectively). *leg1* starts to express at 24 hpf and is maintained at relative high level through the rest of embryonic stages (lanes 6–10: 1 dpf, 2 dpf, 3 dpf, 4 dpf and 5 dpf, respectively). (C) Whole-mount in situ hybridization using a *leg1* probe revealed that *leg1* is enriched in the embryonic liver. (D) MO^{leg1} morphants conferred a smaller liver phenotype (35 out of 42 morphants examined; representative embryo shown on the left) when compared with that of a control MO^{mis} morphant (embryo on the right). Co-injection of the *leg1* mRNA with MO^{leg1} restored the morphant phenotype to normal (embryo in the middle; 26 out 38 embryos examined). Embryos were stained with triple markers: a *secreted immunoglobulin 4* probe as the liver (lv) marker, an *insulin* probe as the endocrine pancreas (en) marker and a *trypsin* probe as the exocrine pancreas (ex) marker.

for liver initiation and development are not included in our current list because many such genes are expressed temporally during liver organogenesis and are no longer expressed in the adult liver. More specific experiments have to be designed and performed to identify such genes, such as via dissecting the liver bud out at different developmental stages and then comparing the expression patterns to identify candidate genes. In any case, the work shown here is a complementary way to identify liver-enriched genes in the developing liver.

Six families of liver-enriched transcription factors, namely HNF-1, HNF-3, HNF-4, HNF-6, C/EBP and D-binding protein (DBP), are known to control the expression of a vast number of liver-specific or -enriched genes via binding to their binding sites within the promoter region or the enhancer region (Cereghini, 1996; Odom et al., 2004; Schrem et al., 2002, 2004). Previous reports have shown that, in many cases, these transcription factors target the promoters of many liver-enriched genes and often work together to control the expression of these liver-enriched genes (Cereghini, 1996; Odom et al., 2004; Schrem et al., 2002). This explains why loss-of-function of individual of these transcription factors often alters a large array of liver-enriched genes (Li et al., 2000; Shih et al., 2001; Wolfrum et al., 2004). Searching for binding sites for Hnf1, Hnf3, Hnf4 and Hnf6 in promoter sequences for 51 liver-enriched genes, we noticed that majority of these genes harbor binding sites for more than one transcription factor, suggesting that the expression of these liver-enriched genes is co-regulated by these factors. It is necessary to point out that the information obtained in this work is undermined due to the fact that our search for Hnf binding sites is limited to the 5 kb region of the promoter region thus preventing us from identifying binding sites contained in the enhancers outside the 5 kb region. Nevertheless, our data provide additional evidence to support the previous hypothesis that these liver-enriched transcription factors form a genetic network, often interacting with each

other, to control the expression of genes responsible for the physiological function of the liver (Odom et al., 2004; Schrem et al., 2002).

Acknowledgments

We thank Dr. Zilong Wen, Dr. Jun Chen, Ms. Weiping Ma, Mr. Honghui Huang and Ms Hua Ruan for their valuable advises and technical helps. This work is supported by the Agency for Science, Technology and Research (A-STAR) in Singapore.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2006.03.018.

References

Barembaum, M., Moreno, T.A., LaBonne, C., Sechrist, J., Bronner-Fraser, M., 2000. Noelin-1 is a secreted glycoprotein involved in generation of the neural crest. Nat. Cell Biol. 2, 219–225.

Blouin, A., Bolender, R.P., Weibel, E.R., 1977. Distribution of organelles and membranes between hepatocytes and nonhepatocytes in the rat liver parenchyma. A stereological study. J. Cell Biol. 72, 441–455.

Cereghini, S., 1996. Liver-enriched transcription factors and hepatocyte differentiation. FASEB J. 10, 267–282.

Chen, Y., Jurgens, K., Hollemann, T., Claussen, M., Ramadori, G., Pieler, T., 2003. Cell-autonomous and signal-dependent expression of liver and intestine marker genes in pluripotent precursor cells from *Xenopus* embryos. Mech. Dev. 120, 277–288.

Chen, J., Ruan, H., Ng, S.M., Gao, C., Soo, H.M., Wu, W., Zhang, Z., Wen, Z., Lane, D.P., Peng, J., 2005. Loss of function of def selectively up-regulates {Delta}113p53 expression to arrest expansion growth of digestive organs in zebrafish. Genes Dev. 19, 2900–2911.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., Malicki, J., Stemple, D.L., Stainier, D.Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., Boggs, C., 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. Development 123, 37–46.

Duncan, S.A., 2003. Mechanisms controlling early development of the liver. Mech. Dev. 120, 19–33.

Field, H.A., Ober, E.A., Roeser, T., Stainier, D.Y., 2003. Formation of the digestive system in zebrafish: I. Liver morphogenesis. Dev. Biol. 253, 279–290.

Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J., Jiang, Y.J., Heisenberg, C.P., Kelsh, R.N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., Nusslein-Volhard, C., 1996. The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. Development 123, 1–36.

Hillebrandt, S., Wasmuth, H.E., Weiskirchen, R., Hellerbrand, C., Keppeler, H., Werth, A., Schirin-Sokhan, R., Wilkens, G., Geier, A., Lorenzen, J., Kohl, J., Gressner, A.M., Matern, S., Lammert, F., 2005. Complement factor 5 is a quantitative trait gene that modifies liver fibrogenesis in mice and humans. Nat. Genet. 37, 835–843.

Jochheim, A., Cieslak, A., Hillemann, T., Cantz, T., Scharf, J., Manns, M.P., Ott, M., 2003. Multi-stage analysis of differential gene expression in BALB/C mouse liver development by high-density microarrays. Differentiation 71, 62–72.

Kawamoto, S., Matsumoto, Y., Mizuno, K., Okubo, K., Matsubara, K., 1996. Expression profiles of active genes in human and mouse livers. Gene 174, 151–158.

Lee, S., Cheng, H., King, K.E., Wang, W., He, Y., Hussain, A., Lo, J., Harberd, N.P., Peng, J., 2002. Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition. Genes Dev. 16, 646–658.

Lee, C.S., Friedman, J.R., Fulmer, J.T., Kaestner, K.H., 2005. The initiation of liver development is dependent on Foxa transcription factors. Nature 435, 944–947.

Li, J., Ning, G., Duncan, S.A., 2000. Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. Genes Dev. 14, 464–474.

Liu, Y.W., Gao, W., Teh, H.L., Tan, J.H., Chan, W.K., 2003. Prox1 is a novel coregulator of Ff1b and is involved in the embryonic development of the zebra fish interrenal primordium. Mol. Cell Biol. 23, 7243–7255.

Lo, J., Lee, S., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W., Wang, W., Wen, Z., Peng, J., 2003. 15,000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. Genome Res. 13, 455–466.

Matthews, R.P., Lorent, K., Russo, P., Pack, M., 2004. The zebrafish onecut gene hnf-6 functions in an evolutionarily conserved genetic pathway that regulates vertebrate biliary development. Dev. Biol. 274, 245–259.

Mayer, A.N., Fishman, M.C., 2003. Nil per os encodes a conserved RNA recognition motif protein required for morphogenesis and cytodifferentiation of digestive organs in zebrafish. Development 130, 3917–3928.

Nasevicius, A., Ekker, S.C., 2000. Effective targeted gene "knockdown" in zebrafish. Nat. Genet. 26, 216–220.

Ober, E.A., Field, H.A., Stainier, D.Y., 2003. From endoderm formation to liver and pancreas development in zebrafish. Mech. Dev. 120, 5–18.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., Fraenkel, E., Bell, G.I., Young, R.A., 2004. Control of pancreas and liver gene expression by HNF transcription factors. Science 303, 1378–1381.

Petkov, P.M., Kim, K., Sandhu, J., Shafritz, D.A., Dabeva, M.D., 2000. Identification of differentially expressed genes in epithelial stem/progenitor cells of fetal rat liver. Genomics 68, 197–209.

Petkov, P.M., Zavadil, J., Goetz, D., Chu, T., Carver, R., Rogler, C.E., Bottinger, E.P., Shafritz, D.A., Dabeva, M.D., 2004. Gene expression pattern in hepatic stem/progenitor cells during rat fetal development using complementary DNA microarrays. Hepatology 39, 617–627.

Reimold, A.M., Etkin, A., Clauss, I., Perkins, A., Friend, D.S., Zhang, J., Horton, H.F., Scott, A., Orkin, S.H., Byrne, M.C., Grusby, M.J., Glimcher, L.H., 2000. An essential role in liver development for transcription factor XBP-1. Genes Dev. 14, 152–157.

Reiter, J.F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N., Stainier, D.Y., 1999. Gata5 is required for the development of the heart and endoderm in zebrafish. Genes Dev. 13, 2983–2995.

Sadler, K.C., Amsterdam, A., Soroka, C., Boyer, J., Hopkins, N., 2005. A genetic screen in zebrafish identifies the mutants vps18, nf2 and foie gras as models of liver disease. Development 132, 3561–3572.

Samadani, U., Costa, R.H., 1996. The transcriptional activator hepatocyte nuclear factor 6 regulates liver gene expression. Mol. Cell Biol. 16, 6273–6284.

Schrem, H., Klempnauer, J., Borlak, J., 2002. Liver-enriched transcription factors in liver function and development: Part I. The hepatocyte nuclear factor network and liver-specific gene expression. Pharmacol. Rev. 54, 129–158.

Schrem, H., Klempnauer, J., Borlak, J., 2004. Liver-enriched transcription factors in liver function and development: Part II. The C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. Pharmacol. Rev. 56, 291–330.

Shih, D.Q., Bussen, M., Sehayek, E., Ananthanarayanan, M., Shneider, B.L., Suchy, F.J., Shefer, S., Bollileni, J.S., Gonzalez, F.J., Breslow, J.L., Stoffel, M., 2001. Hepatocyte nuclear factor-1alpha is an essential regulator of bile acid and plasma cholesterol metabolism. Nat. Genet. 27, 375–382.

Stainier, D.Y., 2001. Zebrafish genetics and vertebrate heart formation. Nat. Rev., Genet. 2, 39–48.

Su, A.I., Guidotti, L.G., Pezacki, J.P., Chisari, F.V., Schultz, P.G., 2002. Gene expression during the priming phase of liver regeneration after partial hepatectomy in mice. Proc. Natl. Acad. Sci. U. S. A. 99, 11181–11186.

Sun, Z., Hopkins, N., 2001. vhnf1, the MODY5 and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, pronephros, and hindbrain. Genes Dev. 15, 3217–3229.

Thisse, C., Zon, L.I., 2002. Organogenesis-heart and blood formation from the zebrafish point of view. Science 295, 457–462.

Tronche, F., Ringeisen, F., Blumenfeld, M., Yaniv, M., Pontoglio, M., 1997. Analysis of the distribution of binding sites for a tissue-specific transcription factor in the vertebrate genome. J. Mol. Biol. 266, 231–245.

Wallace, K.N., Pack, M., 2003. Unique and conserved aspects of gut development in zebrafish. Dev. Biol. 255, 12–29.

Wallace, K.N., Yusuff, S., Sonntag, J.M., Chin, A.J., Pack, M., 2001. Zebrafish hhex regulates liver development and digestive organ chirality. Genesis 30, 141–143.

Wen, C., Zhang, Z., Ma, W., Xu, M., Wen, Z., Peng, J., 2005. Genome-wide identification of female-enriched genes in zebrafish. Dev. Dyn. 232, 171–179.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R.H.A., Cuppen, E., 2003. Efficient target-selected mutagenesis in zebrafish. Genome Res. 13, 2708–2716.

Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J.M., Stoffel, M., 2004. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. Nature 432, 1027–1032.

Wolfrum, C., Poy, M.N., Stoffel, M., 2005. Apolipoprotein M is required for prebeta-HDL formation and cholesterol efflux to HDL and protects against atherosclerosis. Nat. Med. 11, 418–422.

Wood, A.W., Schlueter, P.J., Duan, C., 2005. Targeted knockdown of insulin-like growth factor binding protein-2 disrupts cardiovascular development in zebrafish embryos. Mol. Endocrinol. 19, 1024–1034.

Zaret, K.S., 2002. Regulatory phases of early liver development: paradigms of organogenesis. Nat. Rev. Genet. 3, 499–512.

Zhao, R., Watt, A.J., Li, J., Luebke-Wheeler, J., Morrisey, E.E., Duncan, S.A., 2005. GATA6 is essential for embryonic development of the liver but dispensable for early heart formation. Mol. Cell. Biol. 25, 2622–2631.

Zorn, A.M., Mason, J., 2001. Gene expression in the embryonic *Xenopus* liver. Mech. Dev. 103, 153–157.

Website references

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene&cmd; Unigene database, National Centre for Biotechnology Information (NCBI).

http://www.ncbi.nlm.nih.gov/BLAST/; non-redundant database for searching ESTs, full-length cDNA sequences and genome sequences described in this work, NCBI.

http://www.ensembl.org/Danio_rerio/index.html; a database jointly developed by EMBL, European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute (WTSI) for searching zebrafish genome sequences.