Identification, Characterization, and Mapping of Expressed Sequence Tags from an Embryonic Zebrafish Heart cDNA Library

Christopher Ton,[1] David M. Hwang,[1] Adam A. Dempsey,[1] Hong-Chang Tang,[1] Jennifer Yoon,[1] Mindy Lim,[1] John D. Mably,[2] Mark C. Fishman,[2] and Choong-Chin Liew[1,3,4]

[1]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1L5, Canada;* [2]*Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA;* [3]*The Cardiovascular Genomic Unit, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA*

The generation of expressed sequence tags (ESTs) has proven to be a rapid and economical approach by which to identify and characterize expressed genes. We generated 5102 ESTs from a 3-d-old embryonic zebrafish heart cDNA library. Of these, 57.6% matched to known genes, 14.2% matched only to other ESTs, and 27.8% showed no match to any ESTs or known genes. Clustering of all ESTs identified 359 unique clusters comprising 1771 ESTs, whereas the remaining 3331 ESTs did not cluster. This estimates the number of unique genes identified in the data set to be approximately 3690. A total of 1242 unique known genes were used to analyze the gene expression patterns in the zebrafish embryonic heart. These were categorized into seven categories on the basis of gene function. The largest class of genes represented those involved in gene/protein expression (25.9% of known transcripts). This class was followed by genes involved in metabolism (18.7%), cell structure/motility (16.4%), cell signaling and communication (9.6%), cell/organism defense (7.1%), and cell division (4.4%). Unclassified genes constituted the remaining 17.91%. Radiation hybrid mapping was performed for 102 ESTs and comparison of map positions between zebrafish and human identified new synteny groups. Continued comparative analysis will be useful in defining the boundaries of conserved chromosome segments between zebrafish and humans, which will facilitate the transfer of genetic information between the two organisms and improve our understanding of vertebrate evolution.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. BE693120–BE693210 and BE704450.]

The Human Genome Project (HGP) has amassed a vast quantity of sequencing data; over 90% of the human genes have been deposited into GenBank (June 2000). However, functional interpretation of this sequence data has proven more challenging. Much of this work has involved the study of model organisms because functional inferences based on interspecies comparison of sequences have identified implied function of many orthologous human sequences (Makalowski and Boguski 1998).

Recently the zebrafish, *Danio rerio*, has been recognized as a useful model for the study of development biology and genetics (for review, see Driever and Fishman 1996). One significant advantage of using the zebrafish as a model organism for developmental study is the external development and transparency of the zebrafish embryo. This permits the study of subtle developmental phenotypes in vivo. The zebrafish is also well suited for studies in cardiovascular development because a beating heart is formed and functions within 1 d of fertilization. In addition, the zebrafish embryo does not require blood flow for survival during the first 2 d of development. Thus, zebrafish mutants lacking a circulatory system can still develop normally in the first 2 d (Warren and Fishman 1998) and this allows for studies of mutations that affect the development of the zebrafish heart. Despite all these advantages, the zebrafish suffers from the major drawback of being a new model organism. For example, the number of genes that have been characterized from this species is small compared with other model organisms such as mouse, *Drosophila*, and *Caenorhabditis elegans*.

Expressed sequence tags (ESTs) have proven to be a powerful and rapid approach to identify new genes that are preferentially expressed in certain tissue or cell types (Hwang et al. 1997; Liew et al. 1994; Adams et al. 1995). ESTs have also been used for physical mapping,

[4]Corresponding author.
E-MAIL cliew@rics.bwh.harvard.edu; FAX (617) 975-0995.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.154000.

as has been demonstrated in the development of the human and mouse gene maps (Hayes et al. 1996; McCarthy et al. 1997; Deloukas et al. 1998). Currently, the number of zebrafish ESTs in the public databases is still small compared with mammalian sequences, and there are relatively few tissue-specific cDNA libraries.

Mutational screens in the zebrafish have identified several thousand mutations that affect normal development of the embryo (*Development*, Dec. 1996), including many with essential functions during embryonic heart development (Chen et al. 1996; Stainier et al. 1996). However, the usefulness of these mutations remains limited until the genes responsible for the observed phenotypes are cloned. This is limited in part by a paucity of ordered genes on the zebrafish gene map. Linkage maps based on rapid amplified polymorphic DNAs and microsatellite markers have been produced (Postlethwait et al. 1994, 1999; Johnson et al. 1996; Knapik et al. 1996, 1998; Shimoda et al. 1999). Because linkage mapping requires polymorphic markers for map construction, radiation hybrid (RH) mapping provides a complementary approach to rapidly assign genes and ESTs on the zebrafish map because RH mapping is able to map virtually any marker. Two recent RH maps (Geisler et al. 1999; Hukriede et al. 1999) of more than 3000 markers, genes, and ESTs have dramatically increased the density of the zebrafish gene map and should facilitate the cloning of many identified mutants.

Given the potential and importance of the zebrafish as a model organism for the studies of cardiac development, there is a need for development of EST resources from zebrafish heart cDNA libraries. Here we report the characterization of 5102 ESTs from a 3-d-old zebrafish embryonic heart cDNA library. We also report new map positions for 98 zebrafish ESTs identified in this cDNA library by RH mapping (Table 1) and identification of new synteny groups between zebrafish and human. This EST database represents a new genomic tool for studying aspects of cardiovascular development and disease in the zebrafish and a resource of genes for novel candidate gene discovery.

A total of 5102 EST sequences were processed with the TIGR Assembler to estimate the number of unique transcripts represented in the EST set. A total of 359 clusters composed of 1771 ESTs were generated, whereas the remaining 3331 ESTs did not cluster. The number of unique transcripts identified from the zebrafish embryonic heart EST set was therefore estimated at up to 3690.

RESULTS

Overview of ESTs from the Zebrafish Embryonic Heart cDNA Library

A unidirectional cDNA library was constructed from 3-d-old zebrafish embryonic hearts. A total of 5102 random clones were partially sequenced from this cDNA library to generate ESTs. In total, 2937 (57.6%) showed significant identity to known sequences in the nonredundant nucleotide and peptide databases; of these, 946 were zebrafish entries. Another 722 (14.1%) ESTs matched to other ESTs in dbEST but not to any known sequences. The remaining 1418 (27.8%) showed no match to any known sequences and were designated as novel genes (Table 2).

A total of 5102 EST sequences were processed with the TIGR Assembler to estimate the number of unique transcripts represented in the EST set. A total of 359 clusters composed of 1771 ESTs were generated, whereas the remaining 3331 ESTs did not cluster. The number of unique transcripts identified from the zebrafish embryonic heart EST set was therefore estimated at up to 3690.

Known Gene Expression Profile in Zebrafish Embryonic Heart

ESTs matching to known genes were categorized into seven categories on the basis of general functions of the genes (cell division, cell signaling/communication, cell structure/motility, cell organism/defense, gene/protein expression, metabolism, and unclassified) (Adams et al. 1995; Hwang et al. 1997). In total, 1242 unique known genes were represented and the percentage of transcripts in each category was calculated. The largest class of genes represented those involved in gene/protein expression (25.9%). This class was followed by genes involved in metabolism (18.7%), cell structure/motility (16.4%), cell signaling and communication (9.6%), cell/organism defense (7.1%), and cell division (4.4%). Genes lacking enough information to be classified constituted the remaining 17.9% (Table 3).

Consistent with the high proportions of ESTs involved in gene/protein expression, ribosomal proteins were some of the most abundantly expressed (Table 4). Among other abundantly expressed genes, nine copies of the bone morphogenetic protein 4 (BMP4) were identified. Within the category cell structure/motility, the largest groups of ESTs represented contractile proteins, cytoskeletal proteins, and components of extracellular matrix. The high frequency of these transcripts was not unexpected for the heart, on the basis of our previous experience. However, an unusually high number of keratin proteins (75 clones) and cytokeratin proteins (77 clones) were identified, perhaps due to inclusion of some noncardiac tissues during the isolation of the embryonic hearts.

Comparative Analysis of Gene Expression Profile between Human Fetal Heart and Zebrafish Embryonic Heart

To determine similarities and differences between the two-chambered zebrafish and the four-chambered hu-

Table 1. List of Mapped Zebrafish ESTs

EST identity	Clone names	Accession no.	Primers	Product size (bp)	LG
SDF5	Zeh0225	BE693123	F-ACGTGTAGTTAATGCAGCCG	R-GCTGCACTGTTACAGCAATG	182	1
Actin, alpha cardiac	Zeh0293	BE693127	F-GAACGTATGCCACTGGAATC	R-GACAGCAGAATTACAAGCG	138	1
Eph-related receptor Tyrosine kinase ligand 5 (HTK-L)	Zeh0344	BE693134	F-CTTGTCAGCCATCTGGAATG	R-ATGAATCTGGACATTGCTCG	200	1
Myosin light chain 1, alkali; skeletal fast muscle (MLC1SA)	Zeh0637	BE693148	F-ACAGCTAAGAGGTGCTGTCG	R-AAGTCACATCGTCCTCATGC	170	1
Zinc finger DNA binding protein (fZic)	Zeh0655	BE693149	F-TTCTGTACACATTTTCGTCG	R-GCCCAAGTCAAATGTTGTAC	204	1
E2F-related transcription factor (DP-1)	Zeh1183	BE693163	F-GCTTGCTCAGAGCTGTGAAG	R-GTGCTACTGATTCACGCCAG	122	1
Zinc finger factor (cysteine-rich protein)	Zeh1201	BE693166	F-TAAACTGCCAGCACATTTCC	R-TTATGGGATGTTGATCTCGG	163	1
Prolifin II	Zeh0341	BE693133	F-CGCAAATGGAGCTGAATATC	R-TGTTGCTACTGTGAGATGGG	184	2
Collagen pro alpha-I (II)	Zeh0853	BE693156	F-ACACAGACATTGCATTCCAC	R-TTGATTACGCCGTAGCTATG	170	2
Natural killer cell enhancer factor	Zeh10637	BE693184	F-ATGCTGCAGAGTCTAGTGCC	R-GTTCACCGATAAGCATGGAG	139	2
GTP-binding protein	Zehn0822	BE693205	F-CTGCACTGACGTTACACTGC	R-TCACGTATTGCATGCATCTG	106	2
RAG cohort 1 (RCH1)	Zeh0389	BE693142	F-GTACAACTTGGAGCACGAGG	R-GAATGTGTCGGCACTTGAAGC	185	3
Homeobox transcription factor (hoxb2a)	Zehn0229	BE693187	F-TATTCAATAGGGACAACGCC	R-TGCCCATGTCGAAGTATCAG	194	3
GTP-binding protein (GST1-HS)	Zehn1143	BE693210	F-GTGAACACGCTCATGCACTT	R-ATAATGGCAGGCGGATACAG	201	3
Rabin 3	Zehn0379	BE693189	F-GTGTTTCATCCGACAAGAACG	R-GAACAGAGCCGTCACAGATG	182	4
Carnitine acetyltransferase	Zeh0248	BE693124	F-ACAAAAGCATTCCAGGTGAC	R-ACTGCCACAATCACCAGTTC	112	5
Myosin heavy chain, beta	Zeh0269	BE693126	F-TGTGACTCTGCAATGTCAGC	R-TCTGGTTGACAAGCTTCAGC	150	5
Myosin alkali light chain, atrial	Zeh0374	BE693138	F-ACGTGTAGTTAATGCAGCCG	R-GAACAGTGATGGGTGCTGAG	123	5
Phosphoribosylpyrophosphate synthetase isoform (PRSP1)	Zeh0682	BE693153	F-GCTCCAGTGTAAGCTGTTGG	R-ACACAGGTCTGTGAAGTGCC	237	5
Ran/TC4 binding protein (RANBP1)	Zeh1094	BE693158	F-TGATGACTGACGACTGGTCC	R-CTTCACAAGACCTTGGTGCC	136	5
LIM domain transcription factor	Zeh10169	BE693181	F-TGGCAGACACTGAATAGCAG	R-GTTGGTCTCATGAGGAAACG	115	5
Apolipoprotein A-I protein precursor	Zehn0309	BE693188	F-ACATCTGTGCGAATGTGGTC	R-TTGAGGACTTGAGGACCATG	165	5
DEAD-box protein 72 (P72)	Zeh0176	BE693119	F-CACTTAATCGGTCCGTGATC	R-GTTGTGTCAATCTGCCAACA	196	6
High density lipoprotein binding protein (HDLBP)	Zeh0409	BE693144	F-ACCAACCTGCATAGCAACTG	R-GGCAGCAGAAGTCCTAAGGT	245	6
Gap modifying protein 1 (GMP1)	Zeh0670	BE693152	F-TCGATGGCAGAGGTATGTTC	R-TTCCGATTTGAGTGAACTGC	142	6
S-adenine homocystein hydrolase	Zeh1173	BE693162	F-TTCATCAAATGCTTTCCTCG	R-GTAAATGGCGCATTGAATTG	118	6
S-adenosylhomocystein hydrolase (AHCY)	Zeh1364	BE693175	F-CTACCCAGACTCACAGCCTG	R-TCGGCTCTTCCATGTCTTAC	153	6
PB1 (Polybromo)	Zewp0130	BE693203	F-GCGTGTCTTTCATCATCAGG	R-AAGACTGCCGGCTTAGTAGG	126	6
Myosin regulatory light chain, smooth muscle (MLCB)	Zeh0157	BE693117	F-CAAATGAGATCGAATGCATG	R-CTCGTGGATCATGTGTTCAC	155	7
Kruppel related zinc finger protein (HTF10)	Zeh0353	BE693137	F-AGTCAACATGAAACACCAGG	R-TTTGAACATATGCATGTTGG	159	7
Ferritin heavy subunit (FTH1)	Zeh1145	BE693159	F-ATATCCAGCCACACGTGATG	R-ACATGTTCCAGCAAGCTCACG	158	7
Troponin-T fast muscle	Zeh1249	BE693169	F-AACAGATAAAGCTGGCGAGC	R-TCACTCGTGGTCAAGACATG	249	7
Homeobox transcription factor iriquois 3 (Xiro3)	Zehn0543	BE693191	F-ATATCGTATCGACGCATTGG	R-AATGTTCATGCAATGGCTGTC	129	7
Myosin binding protein C, cardiac (MYBP-C)	Zehn0716	BE693192	F-CGAACTTCCAGTTTGCATTC	R-ACGAAGCCAAGTACAGGATG	146	7
Tumor necrosis factor receptor type I associated protein (TRADD)	Zehn0873	BE693193	F-TTAATCTCGTGGCTGGATCC	R-ACAGGCCTATCAACTGCTGG	138	7
Novel	Zehn1157	BE693198	F-CACATCTGGCAGACATCAGA	R-TGGTTCATGCACTGACTGAC	150	7
Arrestin TRCarr (ARRB2)	Zeh0294	BE693128	F-GGACGACTGAAGGATTCATG	R-ATCATCATCGCCTGACTGTGG	156	8
Y box protein 1	Zeh0308	BE693131	F-TCTGCATAGAGTCTGCAGGC	R-CAACATCCAACATCTGAGCA	109	8
Mitogen-activated protein kinase 14 (CSBP1)	Zeh1243	BE693168	F-GATGCTAAAGCCGGACAGATG	R-CCTGAGGTTGCTACTGTGAA	174	8
Atrial natriuretic factor (ANF)	Zeh1304	BE693172	F-CGGGATATGCTGTATGTATTTCAAC	R-TCGAATGTATATTGACACTGCGTAG	165	8
Death-Associated Protein 1 (DAP)	Zeh0189	BE693120	F-TCATGGCCATCACTTACTCG	R-CAAATGCCAAGCACATTCAG	179	9

Table 1. (*Continued*)

EST identity	Clone names	Accession no.	Primers		Product size (bp)	LG
PINCH protein (PINCH)	Zeh0381	BE693141	F-GTTTCCTTGTCCTCACAGGC	R-ACACTGCTATGAGCGAATGC	109	9
Transcription repressor (GCF2)	Zeh1367	BE693176	FACACGTCTCCAGCAACATTC	R-GACATGACATCCCCATCTTC	179	9
Parvalbumin beta (PVALB)	Zehn1044	BE693194	F-ACGATACAGTGCCACGACTG	R-ATCTGATGCCATCGCTGTC	144	9
Ws-3	Zeh0038	BE693112	F-ATCCCTCATAGAGCCAATGG	R-GCAAGGTTTCGAGGTAGAGG	113	10
Rac protein kinase beta	Zeh0582	BE693147	F-GCCCATGTCTGACTGTGATC	R-TTCGAGAGTGACGCCTTATC	181	10
Nonhistone Chromosomal Protein (HMG17)	Zeh0767	BE693154	F-ATCCCTCATAGAGCCAATGG	R-ATCGTAAATGTTGACAGGCG	170	10
Actin-related protein	Zehn1110	BE693209	F-AGGCGGATCTTAGTCAGGAC	R-TTCTGAGCTCTTCTGGCACT	99	10
Collagen type I, alpha-I (COL1A1)	Zeh0348	BE693135	F-AGAGATGTGCATTGCATTCG	R-TTGCCAGTTCGTCTAACGTC	115	12
Autoantigen annexin XI (ANX11)	Zeh0376	BE693139	F-GATGAACAGGCTGAACCTCC	R-TTCACTGAGGTTTGACCCTG	135	12
Creatine Kinase M	Zeh0657	BE693151	F-GAAACGAGCCAACAGTAGCC	R-TTGAAATGATTCTGCACGTG	165	12
Novel	Zeh0008	BE693110	F-TCAATTATTGCATGCAGCAC	R-TATCCTCATGAAGCCTGGAC	145	13
Hypothetical protein (K04G7.12)	Zeh0031	BE693111	F-GGTTCTGCTTGATCTCTGCC	R-ACAATGACGACGCTGACATC	105	13
Calcium-Binding protein (EF-Hand)	Aeh1186	BE693164	F-TTGAAATGCACAACAGACCC	R-TCATTGACCTGTGCATGTTC	152	13
BMP5	Zeh10669	BE693185	F-GCATATCCACCCACTGACAT	R-ATCAATTCATCAGCGACCAC	258	13
Vinculin (VCL)	Zehn2160	BE693199	F-AACTTTCACAACCAGGCACT	R-ACCTTTAGCTGAGATCCGTG	160	13
CArG box binding factor	Zeh1271	BE693171	F-ACACGATGGGAGGAAGTCTC	R-TGAAATCTGTTAGCGGCAAG	103	14
Receptor for activated protein kinase C (RACK1)	Zehp0047	BE693201	F-GCCACACTCTGATCAGGTTG	R-CATTGTTGATGAGCTGAGGC	137	14
Nonhistone Chromosomal Protein (HMG-14A)	Zeh0993	BE693157	F-ACTGCTGGCATGTTCACAAG	R-AAGCTAATGGCAGAGCTGTG	102	15
TBX2 Protein (T-Box protein 2)	Zeh1581	BE693179	F-CACTCTAATCATCCATGCGC	R-AGTAAGCGGCCTAGAGAGCC	164	15
neurofibromatosis protein type 1 (NF1)	Zehn0874	BE693206	F-TCAGACGAACACGCATCTTC	R-GAAGGCACAGTCTTGACTGC	155	15
Notch homologue 2	Zehs0146	BE693202	F-TGCATGTCGGATAGTTACCG	R-GCCATGTGATTGGCTAATTG	211	15
pregnancy-specific beta 1-glycoprotein 4 precursor (PSBG4)	Zeh0068		F-CAGTGAGGCACAAAGGTAGC	R-TGAACTTTAGAGAGGCTGGC	123	16
Novel	Zeh0082	BE693114	F-TGCCATTGCTGTATCTCACA	R-CGTCTGAATCTGTTGCATTG	181	16
Novel	Zeh0312	BE693132	F-TCAGCTGATGAAGTTCCAGA	R-ACATGTGTGCTTGTAGCAGG	122	16
Peanut (pnut)	Zeh0351	BE693136	F-AGATCTGCCTGTGTCCGAAC	R-ATGTTCATCCAGCAGACTGG	111	16
Novel	Zeh0402	BE693143	F-GAGTTGCAGAGCTGGAGAAC	GTATTGTTGCCTAGTGGCCA	217	16
Rab 13	Zeh0455	BE693145	F-CTCACACCACTCATCTGACC	R-TACATTCCAGTCTGTCAGCC	129	16
Plectin	Zeh0535	BE693146	F-ATCAAGCTTGCCAGATGAAG	R-GCACAAGCAAGACATGAGC	172	16
Apolipoprotein E precursor (APOE)	Zeh1311	BE693174	F-TTCATTTCAGCAGCTGAAGG	R-AATGCCATGTACTCACCACG	199	16
Protein-tyrosine-phosphotase nonreceptor type 2	Zeh1546	BE693177	F-ACTCGCTGAGCTTTAACCTG	R-ACCGTCGTGGTAAGTTGTTG	187	16
S-100 Protein	Zehn1116	BE693208	F-TGCATTGTAACTGCAGTTGC	R-CCTGCGAACAACTTTACCAG	169	16
Novel	Zeh0377	BE693140	F-TGCATGTCTGTGAGTGTTGA	R-CGCAGTGAGTGTTTATGCTC	223	17
IL-13 receptor alpha chain	Zewp0171	BE693204	F-GCTCGGATAGAAAGCAGACA	R-AGTACGTGATTGCGGTTCTG	111	17
Serine/Theonine protein kinase	Zeh1150	BE693161	F-GCTTGTGAAGCGAGTCTCAG	R-CTTGTGCACCAGGTCACTGT	184	18
Death-Associated Protein 5	Zeh1307	BE693173	F-GGCAAATGCAAGTCAGGTAC	R-ATCTGGTCCCATTGATCTGC	203	18
Frizzled protein	Zeh10603	BE693183	F-CTGATCGATGCCAACTCTTG	R-GCAATTGCTCTAGCATGGAG	151	18
Tropomyosin, alpha non-muscle	Zeh0298	BE693129	F-CAGTGCCACTGCTTTGAACT	R-GAGCAGAATGAGCCCAAGTC	138	19
hCDC10 (CDC10 homolog)	Zeh0656	BE693150	F-GTGGTATTGGAGAAGGCCAG	R-CCAGTTCACTGCTTGCTGAA	319	19
Titin	Zeh1256	BE693170	F-AAGAGCTGGCACAGTTTCTG	R-GGCTTGCACACTGAGTTCAT	146	19
TGF-beta receptor interacting protein 1	Zehn0464	BE693190	F-CTCCGTGCAGCTGAGTTAGG	R-GTTACAGCAGCGTTGGAGAG	143	19
Zinc finger protein 45 (BRC1744)	Zehn1068	BE693195	F-CTCTGTAAGCTGACCGATCC	R-GGCAGCAGTCTCAGTAATGC	245	19
Rab5c-like protein	Zehn1144	BE693197	F-AGTGCAAGGCATGGAGTAAG	R-CTAAGTGAATATGCGGCTGC	151	19

Table 1. (*Continued*)

EST identity	Clone names	Accession no.	Primers		Product size (bp)	LG
Regulator of G-protein signaling 7(RGS7)	Zeh0300	BE693130	F-GCAGTGATCACAATACCCTG	R-TCCTTCAGAACGCAGATAGA	175	20
Apolipoprotein B (APOB)	Zehl207	BE693167	F-GGATGACAATAGGTTGCAGG	R-GAAGCCAATGGACACTTCAC	167	20
Connective tissue growth factor XCTGF	Zehl559	BE693178	F-TGACAGGGATACTGGCTCTT	R-ACAGGACCTAGTCGAGTTAG	112	20
Deep Orange protein	Zehl0587	BE693182	F-ATGCACATCCGGTTACATGT	R-CGCAGAAGTTCGATCAAGAG	120	20
Novel	Zeh0115	BE693115	FATAGGCTATTGGCGTTGACA	R-GACGCGTGAATGAAGTGAGT	167	22
Zinc finger protein 37 (DNA binding protein) (ZFP37)	Zeh0174	BE693118	F-CTACATGCTGAATCTGGCCA	R-CACGAGAGGACTCACACTGG	164	22
Similar to yeast SSU72	Zeh1122	BE693160	F-GGCTGCGTCAGGTACAATTA	R-TACTGACCGCAGCAGAGTGT	263	22
Novel	Zeh0124	BE693116	F-GCCACTCTCAGTGCTGTAGC	R-GAGGATCATGGTCACCTGTG	140	23
Twist	Zeh0190	BE693121	F-GTTACCCGTCACTGAAGCAG	R-CTGACCTGATGGATCAAGGC	123	23
ARD-1 N-acetyltransferase homolog (TE2)	Zeh0223	BE693122	F-TAACTCCATGGGTGAGAACC	R-ACGGACGTCAAAGACTCATC	148	23
Neural cell adhesion molecule	Zeh0266	BE693125	F-AGAACGGATTCCTGGACTCA	R-CACAAGTGTAACCGCTCTGT	132	23
Carboxyl terminal LIM domain protein (CLIM1)	Zeh1190	BE693165	F-TACAGGGCTGTGAACTCCAC	R-AATACAGTTTCGCACATGCC	241	23
TGB-b superfamily receptor 1	Zehn1109	BE693196	F-ACTTGGTGCGAGCTGTAATG	R-TTGTGGACTTCCTAACTGCG	171	23
P1-Cdc21	Zeh1616	BE693180	F-CCTGCAGGATAATACGCAGT	R-TATGCAAAGCATGTGCTCTC	159	24
Eph-like receptor tyrosine kinase hEphB1b (EphB1)	Zehn0206	BE693186	F-CATGAGCCTCAGGAGTGAAG	R-AACACGGCAAGACTGTGATG	198	3,12
RanBP7	Zeh0048	BE693113	F-GTTGCGATATCCTGAAGCTG	R-CACGACCTTAGTGGACGATG	156	
Prostaglandin D Synthase	Zeh0800	BE693155	F-ACACATCGGTCCAGAACATG	R-TGAACAGTCATGGTGTGCTC	143	
Calpain 2	Zehn1036	BE693207	F-GTCTTCATCCAGGTCTGCTG	R-TCGAACTGGATATCCTGCAG	127	22
p47	Zehn2383	BE693200	F-TCTCCAACTCCAGAGTGCAG	R-AGCCTGACACTGAAGGAAGC	101	

Listed are the putative identities of mapped ESTs as determined by matches to known sequences in GenBank, the accession no. of the ESTs, the names of the ESTs, primer sequences, PCR product sizes, and linkage group assignment.

Table 2. Summary of ESTs from the Zebrafish Embryonic Heart

Unmatched–novel	1418 (27.8%)
ESTs matching to known sequences	
Matched to other ESTs	722 (14.1%)
Matched to known genes	2242 (43.9%)
Mitochondrial DNA	237 (4.6%)
Ribosomal proteins & RNA	447 (8.8%)
Repetitive elements	11 (0.2%)
Vector	25 (0.5%)
Total	5102 (100.0%)

man heart, we compared proportions of genes in each functional category by using human fetal data from Hwang et al. (1997). Significant differences were detected in five different functional categories. It was found that in the zebrafish embryonic heart, there were significantly fewer transcripts encoding proteins that function in cell division ($P < .005$), cell signaling/communication ($P < .001$), and gene/protein expression ($P < .001$), whereas those involved in cell structure/motility and cell/organism defense were significantly increased ($P < .001$) relative to human fetal heart (Fig. 1; Table 5). Detailed analysis of subcategories found that the decrease in cell division-related transcripts in zebrafish was due to a lower proportion of transcripts representing the general factors of cell division, whereas the decrease in cell/signaling communication was a result of the relative scarcity of identifiable growth factors and hormones in the zebrafish (Table 6). However, the number of transcripts representing effectors/modulators was significantly higher in the zebrafish. This increase could be attributed to a large number transcripts for parvalbumin, a calcium sequesterer detected in fish cardiac muscle (Laforet et al. 1991). Analysis of the cell structure/motility category revealed that extracellular matrix was the only subcategory that showed a significant decrease. However, the number of transcripts representing cytoskeletal proteins was much higher in the zebrafish. This increase was due to the large number of keratin and cytokeratin transcripts present. In the gene/protein expression category, the transcription factors, postranslational modification, ribosomal proteins, and translation factors subcategories all decreased significantly in the zebrafish.

Table 3. Functional Distribution of Known Genes

Functional Category	No. of Unique Genes, %
Cell division	55 (4.4%)
Cell signaling/communication	119 (9.6%)
Cell structure/motility	204 (16.4%)
Cell/organism defense	88 (7.1%)
Gene/protein expression	322 (25.9%)
Metabolism	232 (18.7%)
Unclassified	222 (17.9%)
Total	1242 (100%)

Significantly more ESTs were detected in the cell/organism defense category in the zebrafish, due largely to increases in three subcategories: general homeostasis, carrier proteins, and stress response. Although significant change was not detected in overall levels of transcripts devoted to metabolism, some subcategories exhibited significant changes. Specifically, the nucleotide and transport subcategories showed significant increases, but the sugar/glycolysis subcategory showed decreases. There were also significantly more ADP/ATP carrier proteins and ion-transporting ATPases identified in the zebrafish than in the human heart.

RH Mapping of Embryonic Heart ESTs

Primers were designed for 127 selected ESTs. Of these, 101 (79%) successfully amplified a zebrafish PCR product. Eleven of the primer pairs (9%) failed to amplify a detectable PCR product from zebrafish DNA, and primers for another 8 (6%) ESTs produced Hamster PCR products that could not be clearly distinguished from Zebrafish PCR products. Two primer pairs (2%) were designed for ESTs that are not covered in the hybrid panel (retentions frequency 0%) and primers for 5 (3%) other ESTs produced wrong size PCR products and were discarded. In total, mapping reactions were reproducibly scored for 102 genes represented in the EST set. Of these, 98 (96%) were successfully assigned to single linkage groups (LG), with 23 of 25 groups represented (Table 1). Linkage group 16 contained the most genes ($n = 10$), followed by LG 7 ($n = 8$), LG1 ($n = 7$), and LG6 ($n = 6$). No genes demonstrated significant linkage to LG21 or LG25 in this analysis (Table 1).

Synteny Analysis

To further analyze the conservation of synteny between zebrafish and humans, we compared positions of the mapped zebrafish ESTs and their human counterparts. Following the method described by Gates et al. (1999), we have identified one new conserved syntenic group between zebrafish and human and added more genes to the previously identified groups. Comparing map positions of zebrafish ESTs and human orthologs identified a new syntenic group belongs to linkage group 16 in zebrafish and chromosome 19 in human and added one to two extra genes to each of five previously identified groups (Table 6).

DISCUSSION

The generation of ESTs has proven to be a useful and rapid means to identify and isolate large numbers of expressed sequences (Adams et al. 1992, 1993; Hwang et al. 1994, 1995; Liew et al. 1994). Although extensive

Table 4. Most Abundant Genes Expressed in the Embryonic Zebrafish Heart

Identity	Frequency (%)	Identity	Frequency (%)
Cell division (*n* = 2)		Ribosomal protein S8	10
Nonhistone chromosomal protein HMG-17	6	Ribosomal protein L17	10
Prothymosin alpha	6	Ribosomal protein L8	10
Cell signaling/communication (*n* = 4)		Ribosomal protein L41	10
Parvalbumin, beta	26	Ribosomal protein L19	9
Calmodulin	11	Ribosomal protein L6	9
Bone morphogenetic protein 4 precursor (BMP4)	9	Ribosomal protein L11	8
Receptor for activated protein kinase C (RACK1)	6	Elongation factor 2	8
Cell structure/motility (*n* = 22)		Ribosomal protein L27	7
Myosin heavy chain, fast skeletal muscle	62	Ribosomal protein L3	7
Actin, alpha skeletal	53	Ribosomal protein S2	7
Actin, beta	42	Ribosomal protein S18	7
Keratin	37	Ribosomal protein L13	7
Cytokeratin S	35	Ribosomal protein L13A	7
Myosin light chain 2, fast skeletal muscle (mlc2f)	16	Ribosomal protein S3	7
Tropomyosin, alpha skeletal muscle	14	Homeobox protein LIM-3	6
Cytokeratin II	14	Ubiquitin	6
Cytokeratin 8	11	Ribosomal protein L18a	6
Myosin light chain 1a, fast skeletal	10	Ribosomal RNA large subunit	6
Cytokeratin type I (cytl)	10	Ribosomal protein L10	6
Collagen alpha-2 type I	9	Ribosomal protein S17	6
Myosin light chain 3, fast skeletal	9	Ribosomal protein S19	6
Actin, alpha cardiac	9	Acidic ribosomal protein P2	6
Tubulin, alpha	8	Ribosomal protein SA (P40)	6
Keratin, type II	7	Ribosomal protein S20	6
Myosin regulartory light chain 2A, atrial muscle	6	Ribosomal protein L22	6
Desmin	5	Ribosomal protein S9	6
Fibronectin	5	Ribosomal protein S10	6
Keratin, type II (58 kD)	5	Ribosomal protein L32	6
Myosin heavy chain, alpha cardiac	5	Ribosomal protein L1a	6
Myosin light chain 20-kD (MLC-2)	5	Ribosomal protein S11	6
Cell/organism defense (*n* = 10)		Ribosomal large subunit 26S	6
Globin, beta embryonic 1 (bE1)	31	Ribosomal protein L18	5
Heat shock cognate (hsc70)	31	Ribosomal protein S14	5
Globin 2, alpha-type embryonic	15	Ribosomal protein L1 (L4)	5
zfY1–A cold shock protein	11	Ribosomal protein L5	5
Heat shock protein hsp90beta	10	Ribosomal protein L14	5
Creatine kinase M2-CK	6	Ribosomal protein L9	5
Globin, alpha	6	Ribosomal protein S12	5
Globin, alpha-type embryonic	6	**Metabolism (n = 9)**	
Globin, beta	6	ADP/ATP carrier protein	19
Glutathione S-transferase	5	Cytochrome b	18
Gene/protein expression (*n* = 50)		NADH ubiquinone oxidoreductase subunit 4L	12
Elongation factor 1 alpha	43	Apolipoprotein A-I precursor protein	11
Acidic ribosomal phosphoprotein P0	16	Cytochrome C oxidase subunit III	8
Cathepsin L	15	Apolipoprotein E precursor protein	7
Elongation factor l-gamma	14	NADH dehydrogenase subunit I	7
Ribosomal protein S7	13	ATP synthetase beta-subunit	5
Ribosomal protein L7A	13	ATPase, calcium, sarcoplasmic/endoplasmic reticulum 1 B	5
Polyadenylate-binding protein	12	Isocitrate dehydrogenase	5
Ribosomal protein S4 isoform	11	**Unclassified (n = 3)**	
Ribosomal protein S6	11	Translationally controlled tumor protein P23 (TCTP)	12
Ribosomal protein L30	11	Ependymin beta and gamma chains (Epd)	7
Ribosomal protein S3A	10	SMT3A protein	7
Ribosomal protein L4	10		

Genes are categorized in seven different functional categories and are listed in descending order according to their frequencies.

EST-based resources exist for human and other mammalian models such as mouse and rat, the EST database for the zebrafish presently contains approximately 100,000 ESTs and is still being developed (Gong et al. 1997; Gong 1999). In this report, we characterized the transcriptional profile of 3-d-old embryonic zebrafish hearts by generation of 5102 ESTs. Clustering of 5102 ESTs estimated the maximum number of unique genes

Table 5. Relative Levels of Gene Expression in the Embryonic Zebrafish and Fetal Human Hearts

	No. of ESTs		Proportion of ESTs				
	Z	H	Z	H	EXP	OBS/EXP	χ^2
Cell division							
General	17	154	0.65%	1.42%	36.97	0.46	10.95†
DNA synthesis/replication	8	24	0.31%	0.22%	5.76	1.39	0.87
Apoptosis	6	11	0.23%	0.10%	2.64	2.27	4.28
Cell cycle	20	92	0.77%	0.85%	22.09	0.91	0.20
Chromosome structure	23	149	0.88%	1.37%	35.77	0.64	4.63
Category subtotal	74	430	2.84%	3.96%	103.24	0.72	8.63*
Cell signalling/communication							
Cell adhesion	11	93	0.42%	0.86%	22.33	0.49	5.80
Channel/transport proteins	10	78	0.38%	0.72%	18.73	0.53	4.10
Effectors/modulators	60	156	2.30%	1.44%	37.45	1.60	13.77†
Hormones/growth factors	27	297	1.04%	2.74%	71.31	0.38	28.32†
Intracellular transducers	27	242	1.04%	2.23%	58.10	0.46	17.04†
Metabolism	0	28	0.00%	0.26%	6.72	0.00	6.74
Protein modification	25	166	0.96%	1.53%	39.86	0.63	5.62
Receptors	29	97	1.11%	0.89%	23.29	1.25	1.41
Category subtotal	189	1157	7.25%	10.66%	277.79	0.68	31.83
Cell structure/motility							
General	12	48	0.46%	0.44%	11.52	1.04	0.02
Contractile proteins	229	868	8.79%	8.00%	208.40	1.10	2.22
Cytoskeletal	324	537	12.43%	4.95%	128.93	2.51	310.75†
Extracellular matrix	69	410	2.65%	3.78%	98.44	0.70	9.16*
Microtubule-associated/motors	3	0	0.12%	0.00%	0.00	n/a	n/a
Vesicular transport	4	33	0.15%	0.30%	7.92	0.50	1.95
Category subtotal	641	1896	24.60%	17.47%	455.22	1.41	92.17
Cell/organism defense							
General	52	100	2.00%	0.92%	24.01	2.17	30.63*
DNA repair	18	64	0.69%	0.59%	15.37	1.17	0.45
Carrier protein/membrane transport	96	303	3.68%	2.79%	72.75	1.32	7.65
Stress response	62	146	2.38%	1.35%	35.05	1.77	21.00†
Immunology	7	54	0.27%	0.50%	12.97	0.54	2.76
Category subtotal	235	667	9.02%	6.15%	160.14	1.47	38.74†
Gene/protein expression							
RNA synthesis							
RNA polymerases	3	28	0.12%	0.26%	6.72	0.45	2.07
RNA processing	61	335	2.34%	3.09%	80.43	0.76	4.85
Transcription factors	79	458	3.03%	4.22%	109.96	0.72	8.33*
Protein synthesis							
Posttranslational modification/targetting	56	341	2.15%	3.14%	81.87	0.68	8.45*
Protein turnover	54	151	2.07%	1.39%	36.25	1.49	8.81*
Ribosomal proteins	449	2232	17.23%	20.56%	535.89	0.84	17.81†
tRNA synthesis/metabolism	6	33	0.23%	0.30%	7.92	0.76	0.47
Translation factors	103	685	3.95%	6.31%	164.47	0.63	24.54†
Category subtotal	811	4263	31.12%	39.28%	1023.53	0.79	73.41

(*Continues*)

represented in this set at 3690. Because this analysis was performed on 5′ end sequences that may arise from multiple nonoverlapping segments of the same gene, the true number of unique genes is almost certainly lower.

Of known gene matches, a number of genes thought to be involved in cardiogenesis were identified in the data set. These included nine copies of BMP4, which has been found to be involved in the regulation of left-right asymmetries of the zebrafish heart (Chen et al. 1997; Schilling et al. 1999). Other important factors known to regulate cardiogenesis were also identi-

Table 5. (*Continued*)

	No. of ESTs		Proportion of ESTs				
	Z	H	Z	H	EXP	OBS/EXP	χ^2
Metabolism							
General	10	28	0.38%	0.26%	6.72	1.52	1.6
Amino acid	22	79	0.84%	0.73%	18.97	1.14	0.49
Cofactors	0	12	0.00%	0.11%	2.88	0.00	2.88
Energy/TCA cycle	144	556	5.53%	5.12%	133.49	1.10	0.87
Lipid	51	177	1.96%	1.63%	42.50	1.23	1.73
Nucleotide	32	78	1.23%	0.72%	18.73	1.70	9.48*
Protein modification	9	64	0.35%	0.59%	15.37	0.60	2.65
Sugar/glycolysis	50	363	1.92%	3.34%	87.15	0.56	16.40*
Transport	75	146	2.88%	1.35%	35.05	2.16	46.15†
Category subtotal	393	1503	15.08%	13.85%	360.86	1.10	3.3
Unclassified	263	936	10.09%	8.62%	224.73	1.15	7.14
Total	2606	10854					

*P = .005; †P = .001
(Z) Embryonic zebrafish; (H) Fetal human; (EXP) expected no. of transcripts; (OBS) observed no. of transcripts; (χ^2) chi square result.

fied, including homeobox transcription factors Nkx2.3/2.5, Mef2A/2C, and atrial natriuretic factor.

Although comparative analyses of DNA sequences have been performed between model organisms and humans (Koop 1995; Makalowski et al. 1996; Makalowski and Boguski 1998), little attention has been paid to studying the patterns of gene expression variations between model organisms and humans on a global scale. Understanding similarities and differences between identical tissues in different species is essential in establishing "synexpression" data sets, defining groups of genes that share a similar functional pathway (Niehrs and Pollet 1999). To investigate similarities and differences in gene expression profiles in the developing heart between zebrafish and humans, we analyzed relative levels of expression of genes with related functions. Despite limitations of comparing these two data sets at different stages of development, these findings provide us with a first look at global differences in overall physiological status between the two-chambered zebrafish and the four-chambered human heart, though for the most part, the analysis was too small to reliably reveal differences in the transcription of specific genes. Nevertheless, the results of this analysis suggest several interesting differences in patterns of expression. For example, the high frequency of transcripts detected in the cell/organism defense category in the zebrafish may indicate differences in homeostatic requirements between zebrafish and human hearts. A proportionally high number of heat shock cognate 70 transcripts (hsc70) was detected in the zebrafish heart, with 31 ESTs representing this gene (0.6% of all ESTs). This represents a significant increase in proportion of hsc70 expression over human fetal heart (0.1% of all ESTs; Hwang et al. 1997). Heat shock cognate 70 functions as a chaperone and is known to protect cells against apoptosis (Hohfeld 1998). Heat shock proteins can also be in-

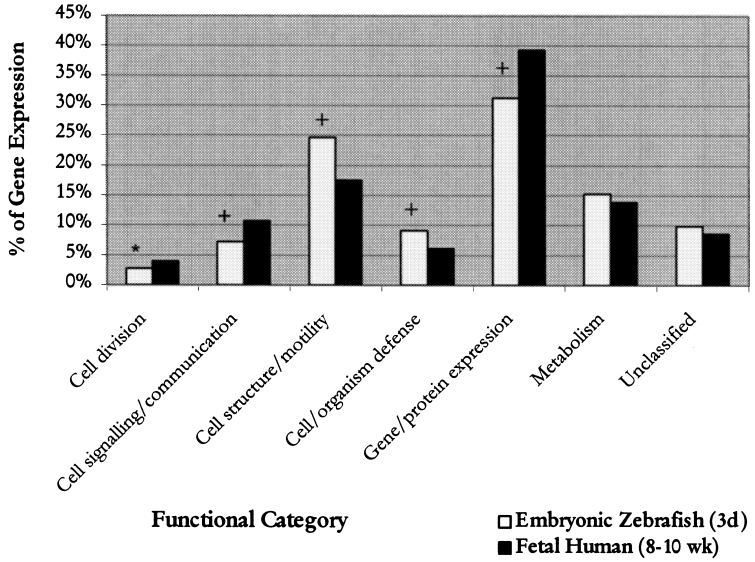

Figure 1 Comparison of relative levels of gene expression between embryonic zebrafish and human fetal hearts. Represented are levels of gene expression in the embryonic zebrafish and fetal human hearts in seven different functional categories. The χ^2 test was used to determine statistical significance (*P = .005; ^+P = .001).

Table 6. Zebrafish-Human Syntenies

LG	EST name	Gene	Reference	Human location
1	Zeh0637	MLC1SA	a	2q33–34
		OTX3	b	2p13
		DLX5	b	2q32
3	Zehn0229	HOXB2A	a	17q21–q22
	Zeh0389	RCH1	a	17q23.1–q23.3
		PARA2B	b	17q12
		CDC27	b	17q12–q23.2
		HOXB	b	17q21–q22
7	Zehn0716	MYBPC2	a	11p11.2
		CCND1	b	11q13
		FGF3	b	11q13
	Zehn873	TRADD	a	16q22
		VNC	b	16
		CK2A2	b	16q13
12	Zeh0348	COL1A1	a	17q21.31–q22.05
		HOXBB	b	17q21–q22
		RARA2A	b	17q12
		DLX3	b	17q21.3–q22
13	Zehn2160	VCL	a	10q22–q23
		RET	b	10q11.2
		PAX2	b	10q24.3–q25.1
16	Zeh0068	PSG4	a	19q13.2
	Zeh1311	APOE	a	19q13.2

[a]This paper.
[b]Gates et al. (1999).

duced by environmental stress. Unlike human fetuses that develop in a stable environment in utero, fish embryos develop externally and it is plausible that the increased levels of hsc70 in the zebrafish embryonic heart may serve a protective role during embryonic development in the face of a potentially changing environment.

Beyond analysis of expression profiles, one immediate application of this EST resource is as a substrate for RH mapping. Recent reports have dramatically increased the number of mapped zebrafish markers, genes, and ESTs (Geisler et al. 1999; Hukriede et al. 1999). Here, we present mapping results for an additional 102 ESTs identified from our library that should further facilitate the identification of zebrafish mutant genes with essential functions during zebrafish embryonic development (Chen et al. 1996; Stainier et al. 1996).

Comparative analysis of map positions between zebrafish and human has identified that gene orthologs that are syntenic in mammals are also syntenic in zebrafish (Postlethwait et al. 1998). This discovery of extensive sharing of chromosome segments between zebrafish and humans has practical significance to the HGP. For example, synteny between zebrafish and humans will enable researchers to identify human ortholog from a gene's position in the zebrafish genome. Reciprocally, and more importantly, the phenotype of a zebrafish mutation can suggest function for the human gene (Postlethwait and Talbot 1997). However, before any conclusive characterization can be made about this conservation, more detailed analyses of these conservations are needed to further define the boundaries of conserved chromosome segments and the extent to which gene order is maintained between zebrafish and human. This information would be particularly useful in identifying candidate genes for positional cloning analyses. It is anticipated that the continuing development of a dense zebrafish map will markedly increase its utility and facilitate the transfer of genetics information between the zebrafish and human.

This collection of 5102 ESTs provides us with a preliminary view into the gene expression profile of the zebrafish embryonic heart. The identification of many genes known to be involved in cardiogenesis suggests that the generation of ESTs is an excellent method for identifying additional genes with essential roles in heart development. Further integration with mapping data of these zebrafish ESTs will provide a richer resource for identifying candidate genes for the several thousand mutants that affect zebrafish development. Construction and characterization of cDNA libraries from additional stages of development, with comparison of gene expression profiles between libraries, should provide further valuable insights into the molecular mechanisms of heart development and disease.

METHODS

RNA Isolation

Total RNA was isolated from 3-d-old zebrafish embryonic heart samples by the method described by Chomczynski and Sacchi (1987). Tissues were homogenized and extracted twice with acidic guanidinium isothiocyanate-phenol-chloroform. The poly(A)+ RNA fraction was isolated by oligo-dT cellulose chromatography (Pharmacia). Purity and RNA integrity were assessed by absorbance at 260/280 nm and agarose gel electrophoresis.

cDNA Library Construction

Libraries were constructed in the λZAP Express vector (Stratagene) according to the manufacturer's protocols. First-strand cDNA was synthesized with an *Xho*I-oligo(dT) adapter-primer. After second-strand synthesis and ligation of *Eco*RI adapters, cDNA was digested with *Xho*I, generating cDNA flanked by *Eco*RI sites at 5′ ends and *Xho*I sites at the 3′ ends. Digested cDNAs were size-fractionated with Sephacryl S-500 spin columns and ligated into the λZAP Express vector predigested with *Eco*RI and *Xho*I. The resulting concatomers were packaged by using Gigapack Gold packaging extracts. After titration, aliquots of primary packaging mix were stored in 7% DMSO at –80°C as primary library stocks, and the remainder was amplified to establish stable library stocks.

Partial Sequencing of 5′ Ends of cDNA Inserts

Plaques were picked randomly and eluted into SM buffer. Phage eluates (5 μL) were directly used for PCR reactions (50-

μL final volume). Reaction mixtures contain 5 μL of 10X Taq buffer, 125 μL of each dNTP, 10 pmol each of forward primer (5′-GCCAAGCTCGAAATTAACCCTCACTAAAGGG-3′) and reverse primer (5′-CCAGTGAATTGTAATACGACTCACTAT AGGGCG-3′) and 1 U of Taq polymerase. The thermal cycle profile consisted of an initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 45 sec, 57°C for 30 sec, and 72°C for 3 min, and a final extension step of 72°C for 3 min. After agarose gel electrophoresis to determine the purity and concentration, 2 μL of PCR products were used directly for cycle sequencing by using the AmpliCycle Sequencing Kit (Perkin-Elmer) and 5 pmol of Cy5 labeled modified T3 primer (5′-GAAATTAACCCTCACTAAAGG-3′). The conditions for cycle sequencing were as follows: 94°C for 2 min, followed by 35 cycles of linear amplification (94°C, 30 sec; 50°C, 15 sec; 72°C, 1 min for 20 cycles and 94°C, 30 sec; 72°C, 1 min for 15 cycles). The reactions were stopped by addition of 0.5 v/v loading buffer (95% formamide, 20 mmol/L EDTA, 10 mg/mL blue dextran). Sequencing reactions were loaded onto 6% acrylamide gels and electrophoresed with A.L.F. and A.L.F. Express DNA sequencers (Pharmacia) (Hwang et al. 1995, 1997).

Bioinformatics

Sequence search analysis of all ESTs against the nonredundant GenBank/EMBL/DDBJ nucleotide, nonredundant GenBank CDS translation/PDB/SwissProt/PIR/PRF peptide, and dbEST databases were performed with the BLAST algorithm (Altschul et al. 1990; Gish and States 1993) on a Unix platform (Sun Microsystems). Assignment of putative identities for ESTs required a minimum P value of 10^{-10}. ESTs with known gene matches were categorized into different functional groups according to categories described in Hwang et al. (1997). Relative levels of gene expression were computed by summing the number of ESTs matching to that particular gene and dividing the sum by the total of ESTs that match to known genes (Hwang et al. 1997). The combined 5102 ESTs were clustered on the basis of sequence similarity by using TIGR Assembler (Fleischmann et al. 1995). Parameters were set so that ESTs were connected together only with a minimum of 95% nucleotide identity in an overlap region of 40 nucleotides. GenBank accession nos. of the Zebrafish Embryonic heart ESTs are AI353073-AI354214; AI616386-AI618739; AI618836-AI618858; AW453485-AW455194. Further clone information can be found on the Internet at URL www.tcgu.med.utoronto.ca.

Preparation of DNA Templates for 3′ End Sequencing

The cDNA clones were excised in vivo from the λZAP Express vector by using ExAssist/XLOLR helper phage system (Stratagene) before sequencing. Phagemid particles were excised by coinfecting *Escherichia coli* XL1-BLUE MRF′ cells with ExAssist helper phage. Excised pBluescript phagemids were used to infect *E. coli* XLOLR cells and selected by using kanamycin resistance. Single colonies were grown overnight in LB-kanamycin and DNA purified by using Qiagen plasmid purification kits. Purified DNA was then used for sequencing of 3′ ends.

Radiation Hybrid (RH) Mapping of cDNA Clones

A 94-hybrid zebrafish RH panel was purchased from Research Genetics. 3′-end sequences of each EST were used to design PCR primers with the assistance of the Williamstone Enterprises Primer Design program (http://www.williamstone.com). Primers were generally 20-bp long and were chosen to generate PCR products of 100–300 bp and a T_m range of 58–60°C. Primer pairs that showed high complementarity to each other or similarity to repeat sequences were discarded. ESTs for which no satisfactory primer pair was found were not used. Names, symbols, and primer sequences are summarized in Table 1. Each primer pair was pretested for specificity with zebrafish and hamster genomic DNA (Research Genetics). Primer pairs that gave a specific zebrafish product were used to screen the RH panel.

PCR amplification was performed in 10-μL reaction mixtures containing reaction buffer, 2mM each dNTP, 0.05 U Taq polymerase, 4 pM each primer, and 5 ng each hybrid. The thermal cycle profile consisted of an initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min, and a final extension step of 72°C for 10 min. PCR products were separated by gel electrophoresis in 2% agarose with 0.5X TBE, and photographed on a UV transilluminator.

Each primer pair was tested in duplicate and positive products were scored. In case of discrepancies (positive on one plate but negative on the other), the band(s) were rescored. Retention profiles were submitted to the Max Planck Institute (Tubingen, Germany) for analysis by SAMapper 1.0 (Geisler et al. 1999).

Statistical Analysis

Analysis of differences in expression levels between zebrafish and human genes was performed by using 2606 and 10,854 unique genes respectively, with ESTs from the mitochondrial genome excluded from calculations. The expected number of zebrafish ESTs present in each functional category/subcategory was calculated based on the frequency of the observed number of ESTs in the fetal human heart cDNA library. By using the same method for identifying differentially expressed genes from EST-based expression profiles as described in Hwang et al. (2000), the statistical significance of the deviation of observed EST profiles from expected was tested with the χ^2 test. For each category, the χ^2 value was calculated by summing the χ^2 value for that category with the χ^2 value calculated from the sum of the remaining category/subcategories. Statistical significance of the deviation from expectations was tested by the χ^2 value with one d.f. The thresholds of significance were established at $^*P = .005$ and $^+P = .001$. The statistical significance of deviation between the two sample sizes was confirmed by using another method for assessing significance of gene expression profiles as described in Audic and Claverie (1997) (http://igs-server.cnrs-mrs.fr).

Phylogenetic Analysis

Following the method described by Gates et al. (1999), each EST sequence was searched against the protein database at NCBI by using the BLASTX program (Altschul et al. 1990). Mammalian sequences that showed significant similarity to the zebrafish EST were retrieved. These sequences were then multiply aligned and neighbor-joining trees were constructed by using CLUSTALX (Thompson et al. 1997). A zebrafish EST is orthologous to a human gene if it appears as a sister group on the dendrogram. The locations of human gene loci were taken from Online Mendelian Inheritance in Man (OMIM) (http://www.ncbi.nlm.nih.gov/omim/); the Genome Data-

base (http://www.gdb.org/gdb), and The Human Gene Map (http://www.ncbi.nlm.nih.gov/genemap99/).

ACKNOWLEDGMENTS

We are grateful to Jack Liew for oligonucleotide synthesis, Wei Wei for assistance with automated sequencing, Robert Geisler and Gerd-Jörg Rauch for calculating map positions on the RH map and to everyone at the Cardiac Gene Unit for technical assistance. This work was supported by the Medical Research Council of Canada. C.T. is a recipient of a Heart and Stroke Foundation of Canada Traineeship. D.M.H. is a recipient of a Hunt Estate M.D./Ph.D. Studentship. A.A.D. is a recipient of a Heart and Stroke Foundation of Canada Traineeship. J. Y. is a recipient of a Heart and Stroke Foundation of Ontario Summer Student Scholarship.

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

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., and Venter, J.C. 1993. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* **4:** 373–380.

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

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Audic, S. and Claverie, J-M. 1997. The significance of digital gene expression profiles. *Genome Res.* **7:** 986–995.

Chen, J.N., Haffter, P., Odenthal, J., Vogelsang, E., Brand, M., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C.P., et al. 1996. Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development* **123:** 293–302.

Chen, J.N., van Eeden, F.J., Warren, K.S., Chin, A., Nusslein-Volhard, C., Haffter, P., and Fishman, M.C. 1997. Left-right pattern of cardiac BMP4 may drive asymmetry of the heart in zebrafish. *Development* **21:** 4373–4382.

Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Deloukas, P., Schuler, G.D., Gyapay, G., Beasley, E.M., Soderlund, C., Rodriguez-Tomé, P., Hui, L., Matise, T.C., McKusick, K.B., Beckmann, J.S., et al. 1998. A physical map of 30,000 human genes. *Science* **282:** 744–746.

Driever, W. and Fishman, M.C. 1996. Heritable disorders in transparent embryos. *J. Clin. Invest.* **97:** 1788–1794.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

Gates, M.A., Kim, L., Egan, E.S., Cardozo, T., Sirtokin, H.I., Dougan, S.T., Lashkari, D., Abagyan, R., Schier, A.F., and Talbot, W.S. 1999. A genetic linkage map for zebrafish: Comparative analysis and localization of genes and expressed sequences. *Genome Res.* **9:** 334–347.

Geisler, R., Rauch, G.J., Baier, H., van Bebber, F., Brobeta, L., Dekens, M.P., Finger, K., Fricke, C., Gates, M.A., Geiger, H., et al. 1999. A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23:** 86–89.

Gish, W. and States, D.J. 1993. Identification of protein coding regions by database similarity search. *Nat. Genet.* **3:** 266–272.

Gong, Z. 1999. Zebrafish expressed sequence tags and their applications. *Methods Cell. Biol.* **60:** 213–233.

Gong, Z., Yan, T., Liao, J., Lee, S.E., He, J., and Hew, C.L. 1997. Rapid identification and isolation of zebrafish cDNA clones. *Gene* **201:** 87–98.

Hayes, P.D., Schmitt, K., Jones, H.B., Gyapay, G., Weissenbach, J., and Goodfellow, P.N. 1996. Regional assignment of human ESTs by whole-genome radiation hybrid mapping. *Mamm. Genome* **7:** 446–450.

Hohfeld, J. 1998. Regulation of the heat shock conjugate Hsc70 in the mammalian cell: The characterization of the anti-apoptotic protein BAG-1 provides novel insights. *Biol. Chem.* **3:** 269–274.

Hukriede, N.A., Joly, L., Tsang, M., Miles, J., Tellis, P., Epstein, J.A., Barbazuk, W.B., Li, F.N., Paw, B., Postlethwait, J.H., et al. 1999. Radiation hybrid mapping of the zebrafish genome. *Proc. Natl. Acad. Sci.* **96:** 9745–9750.

Hwang, D.M., Hwang, W.S., and Liew, C.C. 1994. Single pass sequencing of a unidirectional human fetal heart cDNA library to discover novel genes of the cardiovascular system. *J. Mol. Cell. Cardiol.* **26:** 1329–1333.

Hwang, D.M., Fung, Y.W., Wang, R.X., Laurenssen, C., Cukerman, E., Tsui, S., Fung, K.P., Waye, M., Lee, C.Y., and Liew, C.C. 1995. Analysis of expressed sequence tags from a fetal heart cDNA library. *Genomics* **30:** 293–298.

Hwang, D.M., Dempsey, A.A., Wang, R.X., Rezvani, M., Barrans, J.D., Dai, K.S., Wang, H.Y., Ma, H., Cukerman, E., Liu, Y.Q., et al. 1997. A genome-based resource for molecular cardiovascular medicine: Towards a compendium of cardiovascular genes. *Circulation* **96:** 4146–4203.

Hwang, D.M., Dempsey, A.A., Lee C.Y., and Liew, C.C. 2000. Identification of differentially expressed genes in cardiac hypertrophy by analysis of expressed sequence tags. *Genomics* **66:** 1–14.

Johnson, S.L., Gates, S.L., Johnson, M., Talbot, W.S., Horne, S., Baik, K., Rude, S., Wong, J.R., and Postlethwait, J.H. 1996. Centromere-linkage analysis and consolidation of the zebrafish genetic map. *Genetics* **142:** 1277–1288.

Knapik, E.W., Goodman, A., Atkinson, O.S., Roberts, C.T., Shiozawa, M., Sim, C.U., Weksler-Zangen, S., Trolliet, M.R., Futrell, C., Innes, B.A., et al. 1996. A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphisms. *Development* **123:** 451–460.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C., and Jacob, H.J. 1998. A microsatellite genetic linkage map for zebrafish. *Nat. Genet.* **18:** 338–343.

Koop, B.F. 1995. Human and rodent DNA sequence comparisons: A mosaic model of genomic evolution. *Trends Genet.* **11:** 367–371.

Laforet, C., Feller, G., Narinx, E., and Gerday, C. 1991. Parvalbumin in the cardiac muscle of normal and haemoglobin-myoglobin-free antarctic fish. *J. Muscle Res. Cell Motil.* **5:** 472–478.

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 (ESTs). *Proc. Natl. Acad. Sci.* **91:** 10645–10649.

Makalowski, W. and Boguski, M. 1998. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci.* **95:** 9407–9421.

Makalowski, W., Zhang, J., and Boguski, M.S. 1996. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res.* **6:** 846–857.

McCarthy, L.C., Terrett, J., Davis, M.E., Knights, C.J., Smith, A.L., Critcher, R., Schmitt, K., Hudson, J., Spurr, N.K., and Goodfellow, P.N. 1997. A first-generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res.* **7:** 1153–1161.

Niehrs, C. and Pollet, N. 1999. Synexpression groups in eukaryotes. *Nature* **402:** 483–487.

Postlethwait, J.H. and Talbot, W.S. 1997. Zebrafish genomics: From mutants to genes. *TIG* **13:** 183–190.

Postlethwait, J.H., Johnson, S.L., Midson, C.N., Talbot, W.S., Gates, M., Ballinger, E.W., Africa, D., Andrews, R., Carl, T., Eisen, J.S., et al. 1994. A genetic linkage map for the zebrafish. *Science* **264:** 699–703.

Postlethwait, J.H., Yan, Y.-L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., et al. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* **18:** 345–349.

Postlethwait, J.H., Yan, Y.-L., and Gates, M.A. 1999. Using random amplified polymorphic DNAs in zebrafish genomic analysis. *Methods Cell Biol.* **60:** 165–179.

Schilling, T.F., Concordet, J.P., and Ingham, P.W. 1999. Regulation of left-right asymmetries in the zebrafish by Shh and BMP4. *Dev. Biol.* **2:** 277–287.

Shimoda, N., Knapik, E.W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., and Fishman, M.C. 1999. Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58:** 219–232.

Stainier, D.Y., Fouquet, B., Chen, J.N., Warren, K.S., Weinstein, B.M., Meiler, S.E., Mohideen, M.A., Neuhauss, S.C., Solnica-Krezel, L., Schier, A.F., et al. 1996. Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* **123:** 285–292.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25:** 4876–4882.

Warren, K.S. and Fishman, M.C. 1998. "Physiological genomics": Mutants screens in zebrafish. *Am. J. Physiol.* **275:** H1–H7.

Received June 30, 2000; accepted in revised form October 10, 2000.