Hematopoietic gene expression profile in zebrafish kidney marrow

Huai-Dong Song*†, Xiao-Jian Sun*†‡, Min Deng†§, Guo-Wei Zhang*†, Yi Zhou†¶, Xin-Yan Wu*†, Yan Sheng*, Yi Chen*, Zheng Ruan*, Chun-Lei Jiang*, Hui-Yong Fan*, Leonard I. Zon¶‖, John P. Kanki§, Ting Xi Liu§, A. Thomas Look§, and Zhu Chen*‡***

*State Key Lab for Medical Genomics, Shanghai Institute of Hematology, Ruijin Hospital Affiliated to Shanghai Second Medical University, Shanghai 200025, China; ‡Health Science Center, Shanghai Second Medical University and the Shanghai Institute of Biological Science, Chinese Academy of Sciences, Shanghai 200025, China; §Department of Pediatric Oncology, Dana–Farber Cancer Institute, Boston, MA 02115; and ¶Division of Hematology and Oncology and ‖Howard Hughes Medical Institute, Children's Hospital and Harvard Medical School, Boston, MA 02115

Contributed by Zhu Chen, October 1, 2004

The zebrafish kidney marrow is considered to be the organ of definitive hematopoiesis, analogous to the mammalian bone marrow. We have sequenced 26,143 ESTs and isolated 304 cDNAs with putative full-length ORF from a zebrafish kidney marrow cDNA library. The ESTs formed 7,742 assemblies, representing both previously identified zebrafish ESTs (56%) and recently discovered zebrafish ESTs (44%). About 30% of these EST assemblies have orthologues in humans, including 1,282 disease-associated genes in the Online Mendelian Inheritance in Man (OMIM) database. Comparison of the effective and regulatory molecules related to erythroid functions across species suggests a good conservation from zebrafish to human. Interestingly, both embryonic and adult zebrafish globin genes showed higher homology to the human embryonic globin genes than to the human fetal/adult ones, consistent with evo-devo correlation hypothesis. In addition, conservation of a whole set of transcription factors involved in globin gene switch suggests the regulatory network for such remodeling mechanism existed before the divergence of the teleost and the ancestor of mammals. We also carried out whole-mount mRNA *in situ* hybridization assays for 493 cDNAs and identified 80 genes (16%) with tissue-specific expression during the first five days of zebrafish development. Twenty-six of these genes were specifically expressed in hematopoietic or vascular tissues, including three previously unidentified zebrafish genes: *coro1a*, *nephrosin*, and *dab2*. Our results indicate that conserved genetic programs regulate vertebrate hematopoiesis and vasculogenesis, and support the role of the zebrafish as an important animal model for studying both normal development and the molecular pathogenesis of human blood diseases.

The zebrafish has proven to be a valuable model organism for genetic studies of human hematopoiesis (1, 2). In mammals, primitive hematopoiesis takes place in the yolk sac, later moving to the aorta–gonad–mesonephros region and the fetal liver (3), whereas definitive hematopoiesis in adults occurs in the bone marrow (1). In zebrafish, the earliest site of hematopoiesis is the intermediate cell mass (ICM), which is analogous to the blood island of the yolk sac in mammals, whereas definitive hematopoiesis appears to initiate in the dorsal aorta and maintains in the kidney marrow during later stages of maturation and throughout adulthood (4).

Cataloging ESTs and isolation of full-length cDNAs are essential means for gene discovery, as well as for the understanding of tissue/organ functions and the identification of tissue-specific markers. The comparative studies of transcriptomes among different species also shed light on the evolution of the molecular mechanism in physiology. Meanwhile, whole-mount mRNA *in situ* hybridization (WISH) in zebrafish has been used as a common assay to explore gene function and to find useful probes in mutagenesis screening (5, 6). To date, >26 complementation groups of mutations have been identified in zebrafish that affect hematopoietic development (1). The characterizations of these mutant lines have shown that most of them affect critical genes within the erythroid pathway (1, 7). However, few mutants that specifically affect granulopoiesis have been identified because of the paucity of zebrafish granulocyte-specific markers (2).

In this study, an analysis of zebrafish kidney marrow EST and full-length ORF-containing cDNA clones provided some insights into the evolutionary features of genetic regulation of hematopoiesis in vertebrates. The WISH assay uncovered the spatial and temporal expression patterns of 80 genes exhibiting tissue-specific expression during embryogenesis. In addition, we have focused on the developmental expression pattern of three recently discovered zebrafish genes in granulocytic and vascular tissues, *coro1a*, *nephrosin*, and *dab2*, whose roles in mammalian hematopoiesis and vasculature development are not yet well defined.

Materials and Methods

Library Construction and DNA Sequencing. Three hundred normal kidney marrows of mature zebrafish were collected for poly(A)$^+$ RNA extraction as described (8). The cDNAs were ligated into the λZAP expression vector, PBK-CMV (Stratagene), at the *Xho*I and *Eco*RI sites. The ligated products were packaged by using Gigapack Gold packaging extracts. Bacteria growth and plasmid extractions were performed in a 96-well format (Qiagen, Valencia, CA). Sequencing reactions were performed with T3 (5′ end) and T7 (3′ end) primers on 9600 Thermal Reactors (PerkinElmer). The reaction products were analyzed by using an ABI 377 or ABI 3700 DNA Sequencer (Applied Biosystems) as described (9).

Bioinformatic Analysis and Data Management. Quality assessment and base trimming of ESTs were performed by using PHRED2.1. The quality of the sequences was judged "good" if they contained <3% ambiguous bases and were >100 bp. EST sequences were clustered by CAT3.5 software (Pangea, Oakland, CA) to obtain EST assemblies, which were sequentially searched against the zebrafish UniGene and dbEST databases (www.ncbi.nlm.nih.gov). The EST assembly was considered identical to known entries (ESTs or genes) if they shared at least 95% identity over 100 bp of DNA sequences. The prediction of zebrafish–human orthologues relied on the "reciprocal best hit" search methods as

Abbreviations: ICM, intermediate cell mass; WISH, whole-mount mRNA *in situ* hybridization; hpf, hours postfertilization; pnd, pronephric ducts.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. CD580595–CD606737, AY391404–AY391473, AY394926–AY394976, AY398305–AY398433, AY422992–AY423041, and AY330221–AY330224).

†H.-D.S., X.-J.S., M.D., G.-W.Z., Y.Z., and X.-Y.W. contributed equally to this work.

**To whom correspondence may be addressed. E-mail: zchen@stn.sh.cn or thomas_look@dfci.harvard.edu.

described (10, 11). The criteria for a cDNA to be accepted as having an entire ORF have been described (12). The programs SIGNALP and TMHMM were used to predict the signal peptide and the transmembrane domains of full-length cDNAs.

Fish Maintenance and Embryonic Staging. AB zebrafish were maintained and staged as described (13). Embryos raised to time points beyond 24 h postfertilization (hpf) were treated with 0.003% phenylthiourea to prevent melanization. Embryos were removed from their outer chorions with 0.001% pronase and fixed overnight at 18, 24, 72, and 120 hpf in 4% paraformaldehyde at 4°C. Fixed embryos were washed in PBST (PBS supplemented with 0.1% Tween 20) and dehydrated in graded PBST/methanol solutions (3:1, 1:1, 1:3) for 10 min each and stored in absolute methanol at −20°C.

WISH and Histology. Fixed embryos (stored at −20°C in methanol) were rehydrated in graded PBST/methanol solutions (1:3, 1:1, 3:1) for 10 min each, followed by PBST rinse twice for 10 min each at room temperature (RT). Embryos at 72 and 120 hpf were further permeabilized with proteinase K solution (100 μg/ml) at RT for 20–30 min, rinsed in PBST twice, and refixed in 4% paraformaldehyde at RT for 30 min. Embryos at 18 and 24 hpf were not treated with proteinase K. Ten to 15 embryos from each time points (18, 24, 72, and 120 hpf) were combined and hybridized with digoxigenin-labeled antisense RNA probes at 68°C. Synthesis of RNA probes and hybridizations were performed as described (2). Double *in situ* hybridization assays were performed by using digoxigenin and fluorescein-labeled probes, and developed with the chromagenic substrates 5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium and Fast Red. After *in situ* hybridization, embryos were rehydrated and embedded in 3% sucrose overnight at 4°C before cryostat sectioning. Sections were counterstained with eosin and photographed.

Results

Identification and Functional Categorization of Zebrafish Orthologues of Human Genes. Among the 26,143 good quality EST sequences, 24,195 ESTs sequenced from 5′ end were clustered into 8,062 EST assemblies, whereas 1,948 corresponded to 3′ ends of cDNAs. After searching against the Washington University EST database (http://zfish.wustl.edu), the number of unique transcripts was estimated to be 7,742. Of these, by the date of submission (June 16, 2003), 47% corresponded to known zebrafish ESTs, 9% corresponded to known zebrafish genes deposited in the public database, and 44% represented previously undescribed ESTs.

Sequence-based "reciprocal best hit" analysis provides a reliable method to identify zebrafish orthologues of human genes (10). Of 7,742 unique transcripts, 2,225 (28.7%) were considered to be zebrafish orthologues of human genes, of which 272 were previously entered into zebrafish public databases and 1,953 were not. The 2,225 orthologues were grouped into seven categories based on the general cellular function, including cell signaling/communication (19.4%), gene/protein expression (18.7%), metabolism (14.6%), cell structure/motility (5.5%), cell division (4.7%), and cell organism/defense (4.2%), with the remaining 33% lacking enough structural and functional information for classification. Of note, 1,282 of these orthologues corresponded to human disease genes as registered in the Online Mendelian Inheritance in Man (OMIM) database. In addition, 304 previously undescribed cDNA clones with complete ORF were characterized based on our EST source, with a range of ORF length of 68–1,235 aa (median, 373 ± 10.6 aa). These genes might be involved in critical biological process according to their homology to known genes with established functions, such as secretory proteins, transmembrane proteins, signal transduction molecules, and enzymes, etc. (see Table 2, which is published as supporting information on the PNAS web site, and www.chgc.org.cn/danio).

Hematopoietic and Vascular Genes. Ninety genes (Table 3, which is published as supporting information on the PNAS web site), whose mammalian homologues have been shown to be either associated functionally with hematopoiesis/angiogenesis or hematopoietic disorder, were divided into three subgroups: (*i*) genes regulating the processes of hematopoiesis or angiogenesis (regulators); (*ii*) genes related to mature blood cell function (effectors), and (*iii*) genes involved in human hematopoietic disorders. These genes include cytokines, receptors, signal transduction molecules, and transcription factors that regulate the processes of blood or blood vessel development in different stages. It is worth noting that some genes involved in human hematopoietic diseases were uncovered in the present work. For example, the human *CRKL* gene has been implicated as a substrate for the BCR-ABL tyrosine kinase in chronic myelogenous leukemia (14), and six leukemia-associated genes, *MSF*, *SEPT6*, *FNBP1*, *GAS7*, *ARHGEF12* and *ELL*, have been identified as components of MLL fusion proteins in leukemia with 11q23 chromosomal translocation (15–20).

Many mammalian genes have been shown to play a role in the development and maturation of hematopoietic and vascular progenitors (Fig. 1) (7, 21). In this study, 21 orthologues of these genes were identified: two (*smad5* and *sdf1*) are involved in the induction of blood from ventral mesoderm (7, 22); nine (*c-myb*, *scl*, *fli1*, *flk1*, *pu.1*, *cebp*α, *cebp*β, *ikaros*, and *klf1*) are required for the maintenance and commitment of hematopoietic stem/progenitor cells; and six [*hemoglobin* (*hbaa1, ba1, ba2*), *alas2*, *myeloperoxidase* (*mpo*), and *phox47*] are expressed specifically in the erythroid or granulocytic lineages. In addition, a number of genes encoding homologues of human cytokines, *jak1*, *stat3*, *raf*, *erk1*, *pim1*, and *ikb*α (data not shown), which act as signaling transducers required for granulopoeisis, are present.

Evolutionary Features of Zebrafish Genes Related to Erythropoiesis. The availability of a large amount of EST and genome sequence data from zebrafish allowed an evolutionary analysis in comparison with other vertebrates. Here, we focused on some of the representative effectors and regulators of hematopoietic system, particular on erythropoiesis. As in other fish species and in mammals, zebrafish globin genes could be divided into α- and β-like families (Fig. 2*A*), and, in phylogenetic analysis, they formed distinct clusters with those of other species. Nevertheless, within each family, although the same set of the embryonic/fetal/adult globin genes from mammals constituted monophyletic branch, the zebrafish embryonic and adult globin subtypes found no such kind of close relationship with mammals and could only form an independent branch. These data suggest that the ancestors of α- or β-like globins should exist before the divergence between the teleost and the ancestor of mammals ≈450 million years ago, whereas the differentiation of subtype globin genes with regard to developmental stages might occur after this divergence. A pair-wise analysis of the similarity between zebrafish and human globins was further examined. The result revealed that the zebrafish globins consistently maintained highest homology to these human globin genes that were early expressed during embryogenesis. For example, the zebrafish α-like globins (hbael, hbae3, hbaa1) were most similar to the human ζ globin, whereas zebrafish β-like globins (hbbel, hbbe2, hbbe3, bal, ba2) displayed strongest homology to human ε globin (Table 1). It may be postulated, therefore, that embryonic globin genes represent the evolutionarily most conserved ones and the zebrafish globins may be have functionally more like the human embryonic globins, a point of view consistent with the evo–devo correlation hypothesis (23) and in strong support of the previous speculation that the individual member genes of each globin

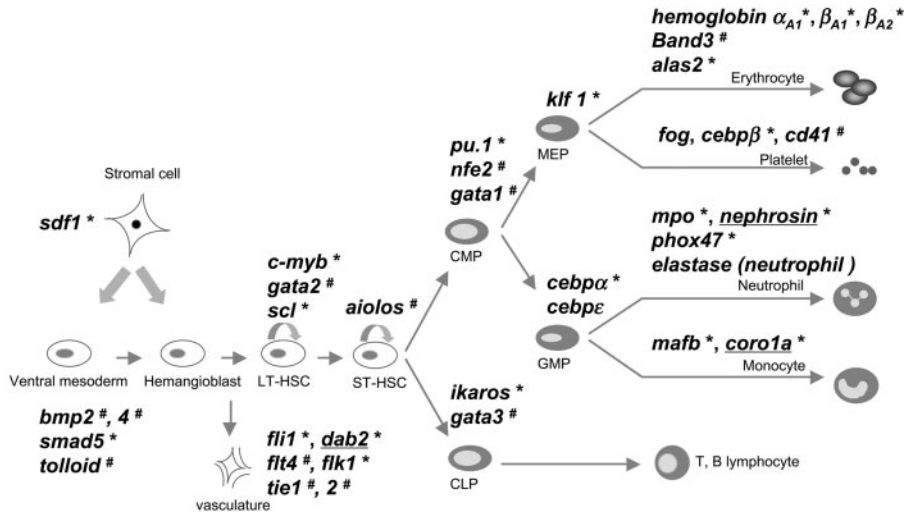

Fig. 1. Conservation of genetic program of vertebrate hematopoiesis. Conceptual schematic diagram of the blood development and the different stages of hematopoietic stem cell development are shown. Thirty-seven mammalian genes known to be involved in hematopoiesis/vasculogenesis are italicized and highlighted in bold. Asterisks indicate the zebrafish orthologues identified in this study, whereas # indicates those from the public zebrafish EST database. Two recently discovered myeloid-specific genes (*coro1a* and *nephrosin*) and a vascular gene (*dab2*) are underlined. LT-HSC, long-term hematopoietic stem cell; ST-HSC, short-term hematopoietic stem cell; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/macrophage progenitor.

family have arisen through a series of tandem duplications of ancestral α- and β-like globin genes during the evolution of higher vertebrates (24, 25).

A next and related question was the conservation between zebrafish and other vertebrates of the erythroid regulators, namely transcription factors associated with globin switch such as *GATA1*, *YY1*, *NR2F2* (*COUP-TFII*), *NFE2, TIEG2*, and *KLF1* (*EKLF*) (Fig. 2*B*). It was well known that, in mammals, *GATA1* is an activator of γ globin, and it also inhibits the expression of ε globin gene in the presence of *YY1*, whereas *KLF1* enhances the expression of β globin gene. Our analysis showed that all these regulators are well conserved in zebrafish (Fig. 2 *C–H*), indicating an early existence of regulatory network for remodeling mechanism such as the globin switch. It also may be interesting to note that, according to their position in the phylogenetic trees, regulators for early switch (*GATA1*, *YY1*, *NR2F2*, *NFE2*, and *TIEG2*) have higher homology through evolution than that of *KLF1*, a regulator of second subtype switch (Fig. 2 *C–H*). Interestingly, like the recent report for another fish species, the four subtype of T cell receptor genes (α, β, γ, δ) were found in zebrafish and showed close relationship with the corresponding human genes (data not shown), suggesting that the major hematopoiesis-related molecules appeared in a very early stage of vertebrate evolution.

WISH Analysis of Gene Expression. Because the functions of many of the mammalian homologues of the zebrafish ESTs are poorly understood, the examination of their respective gene expression patterns during early development should help to establish their tissue-specific roles and to identify valuable reagents for dissecting hematopoietic regulatory pathways. A total of 493 unique EST assemblies were selected based on their potential relevance to human hematopoiesis and diseases for WISH analysis. Embryos at 18, 24, 72, and 120 hpf were used to detect the dynamic changes in gene expression that occur during embryonic organogenesis. Fig. 3*Aa* illustrates the three major spatial and temporal expression patterns that have been observed: (*i*) no signal at any of the four time points (148 of 493; 30%); (*ii*) ubiquitous expression at all four time points (265 of 493; 54%); and (*iii*) tissue-specific gene expression in at least one embryonic stage (80 of 493; 16%). For examples, the expression of *phenylalanine hydroxylase* (*pah*), whose deficiency causes phenylketonuria and mental retardation in human infants, was detected in the developing central nervous system (CNS) and pronephric ducts (pnd) at 18 (Fig. 3*A b–d*) and 24 (Fig. 3*A e* and *f*) hpf. A number of genes were found exclusively expressed in the developing embryonic kidney, such as zebrafish *renin* in the putative juxtaglomerular cells (Fig. 3*A g–j*) or a recently discovered EST (RK011A3G03) in the pnd (Fig. 3*A k–n*). Other genes exhibited specific expression patterns in other organs/tissues such as the central nervous system, thyroid, skin, liver, ears, and neuromasts (see Tables 2 and 3 and www.chgc.org.cn/danio).

Of the 80 (16%) genes demonstrating tissue-specific expression patterns, 26 genes (5%) showed hematopoietic and/or vascular cell specific expression. Some of these were well known hematopoietic and/or vascular-specific genes, including *scl*, *pu.1*, *fli1*, *hemoglobin*, *mpo*, and *cebp*α. However, the significance of several clones expressed in blood or vasculature, such as *coro1a* (GenBank accession no. AY330221), *nephrosin* (GenBank accession no. AY330222), and *dab2* (GenBank accession no. AY330223), remained to be established. We have thus examined the expression patterns of these genes in more detail.

Recently Discovered Zebrafish Myeloid-Restricted Genes *coro1a* and *nephrosin*. The zebrafish orthologue of human *CORO1A* gene, *coro1a*, was expressed in cells that migrated over the surface of the anterior head and yolk at 18 hpf (Fig. 3*B a* and *b*) and throughout the embryo by 24 hpf (Fig. 3*Bc*). Double WISH assays with both *mpo*, a specific marker for both mammalian and zebrafish neutrophils, and *coro1a* were performed on 24-hpf embryos. All *mpo*-positive cells were also *coro1a*-positive, whereas some cells expressed *coro1a* alone (Fig. 3*B d* and *e*).

The zebrafish *nephrosin*-positive cells were first found scattered over the surface of anterior yolk in $\approx$10% of 24-hpf embryos. By 30 hpf, all embryos contained cells expressing *nephrosin* in both anterior yolk and posterior ICM (data not shown). By 72 and 120 hpf, cells expressing zebrafish *nephrosin* were distributed throughout the embryo, but concentrated over the anterior yolk, in the posterior ICM and in the posterior blood island (Fig. 3*B f* and *g*). Double WISH assays indicated that most

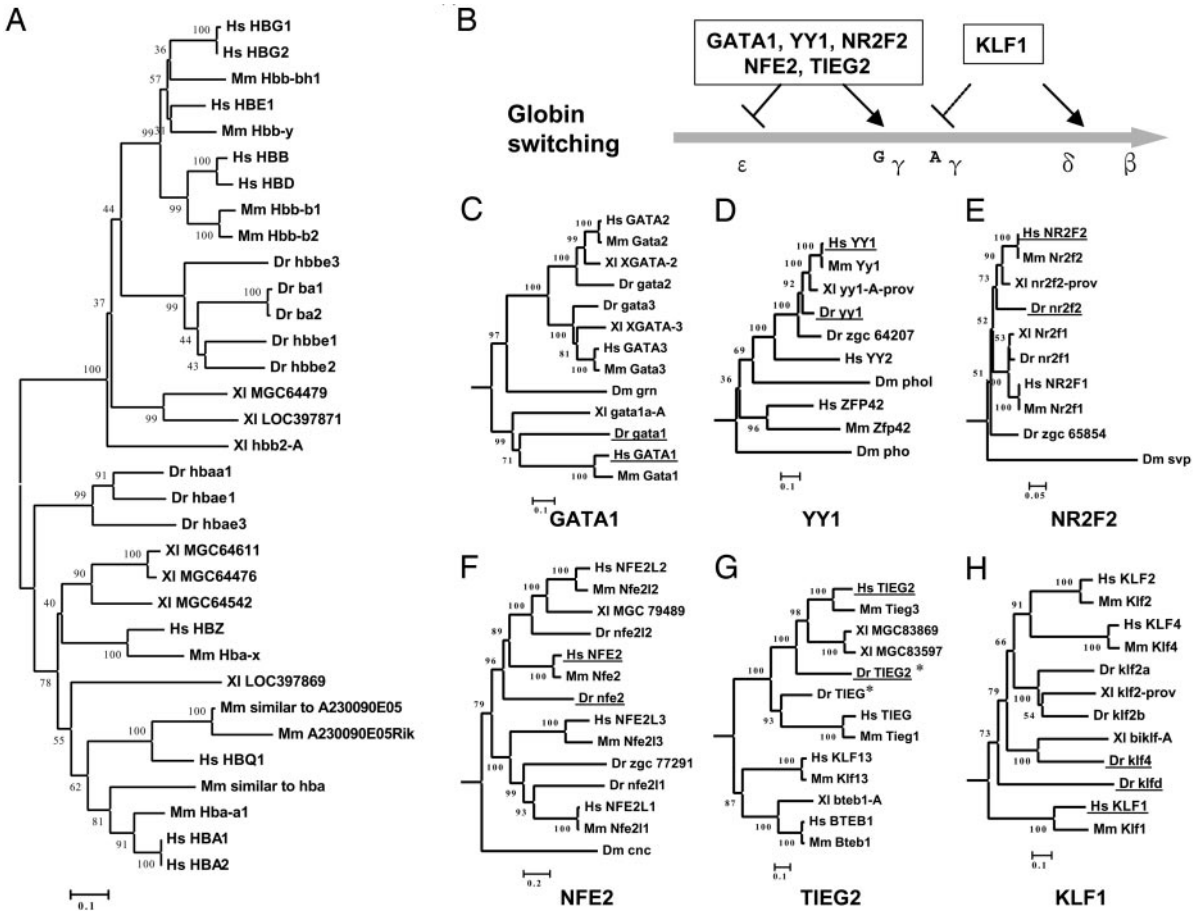

Fig. 2. Molecular phylogenetic analysis of hematopoietic lineage-specific factors. Amino acid sequences of α- and β-globins (*A*) and of erythropoiesis-related transcription factors (*C–H*) from human (Hs), mouse (Mm), *Xenopus* (Xl), and zebrafish (Dr) were aligned with CLUSTALX 1.83. Phylogenetic trees were constructed with the MEGA 2.1 program using the neighbor-joining method with 1,000 bootstrap replicates. The numbers indicate the bootstrap confidence level. In *C–H*, the human erythropoiesis-related transcription factors and their zebrafish counterparts are underlined. Asterisks indicate predicted zebrafish peptides according to EST assembly. The positions of transcription factors involved in globin gene expression switch are represented in *B*.

nephrosin-positive cells also expressed the granulocytic marker *mpo* (Fig. 3*Bh*), whereas a smaller number of these cells coexpressed the monocytic *lysozyme c* at 72 hpf (Fig. 3*Bi*). The *nephrosin*-positive cells did not coexpress *c-myb* transcripts in 72-hpf embryos (data not shown), suggesting that *nephrosin* was not expressed by myeloid progenitors. Compared to some granulocyte-specific genes such as *mpo* expressed as early as 18 hpf, zebrafish *nephrosin* could be a relatively late stage marker of granulocytic differentiation.

Recently Discovered Zebrafish Vascular-Specific Gene *dab2*. The expression of zebrafish *dab2* was restricted to cells of the pnd, ICM, dorsal aorta (da), and caudal vein (cv) of 18- and 24-hpf embryos (Fig. 3*C a* and *b*). By 3 days after fertilization, expression of *dab2* was detected mainly in cv (Fig. 3*Cc*) and pnd (Fig. 3*Cd*) as shown in cross section (Fig. 3*Ce*) at the level indicated in Fig. 3*Cd*. Double WISH studies on 24-hpf embryos with *fli1* confirmed that *dab2* was coexpressed with zebrafish *fli1* in da and cv cells forming in the trunk, but not in the most posterior

Table 1. The pairwise similarities between human and zebrafish globins

			α			β				
			E	F/A		E	F		A	
			Hs_HBZ	Hs_HBA1	Hs_HBA2	Hs_HBE1	Hs_HBG1	Hs_HBG2	Hs_HBD	Hs_HBB
α	E	Dr_hbae1	**6E − 40**	5E − 35	5E − 35	4E − 26	3E − 24	5E − 24	1E − 23	6E − 24
		Dr_hbae3	**2E − 23**	9E − 23	9E − 23	2E − 14	1E − 15	5E − 16	5E − 14	2E − 15
	A	Dr_hbaa1	**1E − 39**	2E − 37	2E − 37	2E − 25	4E − 26	1E − 26	5E − 25	5E − 26
		Dr_hbbe1	4E − 18	2E − 17	2E − 17	**3E − 39**	3E − 38	3E − 38	2E − 36	3E − 36
β	E	Dr_hbbe2	2E − 21	1E − 17	1E − 17	**9E − 41**	8E − 39	8E − 39	7E − 36	4E − 36
		Dr_hbbe3	8E − 20	9E − 17	9E − 17	**9E − 42**	1E − 39	1E − 39	1E − 35	2E − 35
	A	Dr_ba1	4E − 20	4E − 17	4E − 17	**4E − 42**	2E − 39	1E − 39	2E − 40	5E − 41
		Dr_ba2	2E − 20	2E − 17	2E − 17	**2E − 42**	9E − 40	5E − 40	3E − 40	1E − 40

The similarities are represented by expect value of BLASTP analysis. The highest similarities (lowest expected values) are indicated in bold.

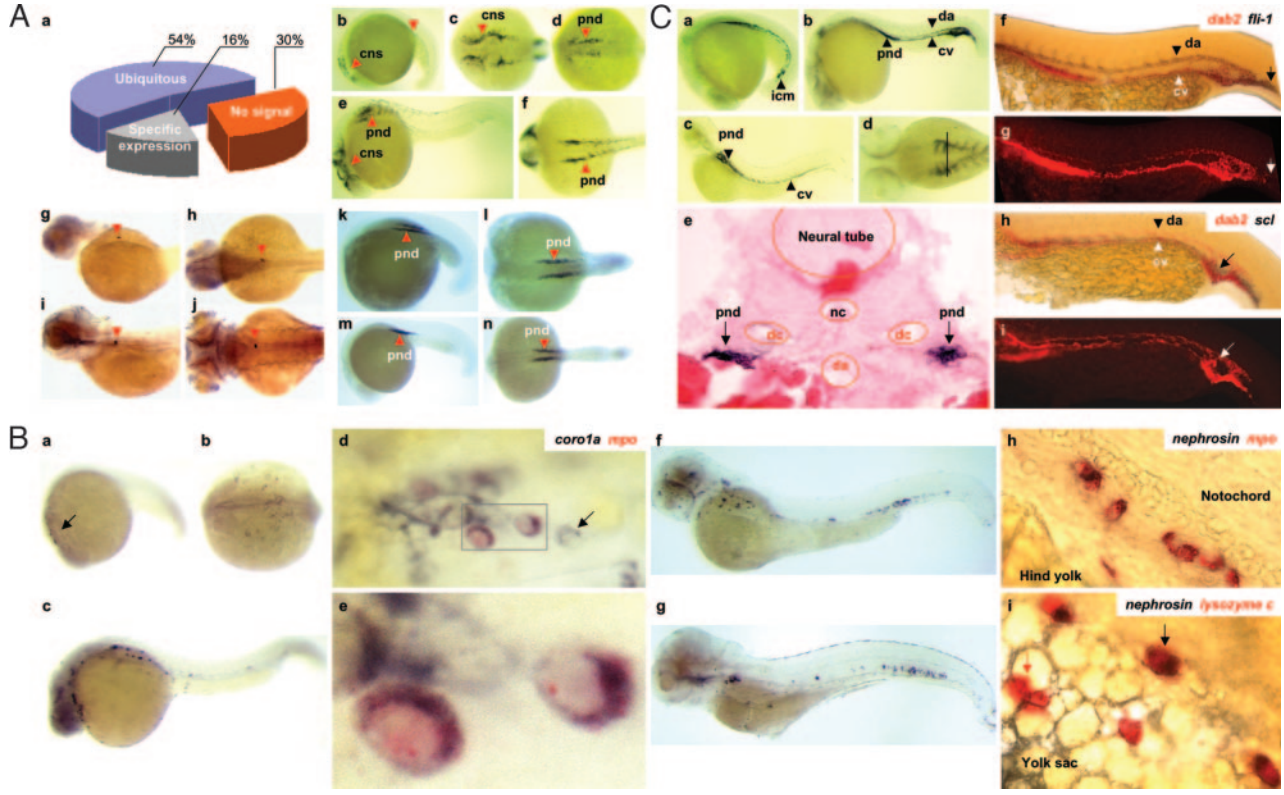

Fig. 3. Gene expression analysis with whole-mount mRNA *in situ* hybridization. (*A*) Representative examples of zebrafish genes exhibiting tissue-specific expression by WISH. (*a*) Categorization of gene expression patterns of 493 EST assemblies. The zebrafish orthologue of human phenylalanine hydroxylase (*PAH*) (Hs.325404) was expressed in the developing CNS and pnd at 18 (*b–d*) and 24 (*e* and *f*) hpf embryos (arrowheads). The zebrafish homologue of human *RENIN* (Hs.3210) was expressed in the site where the glomerulus of zebrafish kidney develops at 72 (*g* and *h*) and 120 (*i* and *j*, arrowheads) hpf embryos. The expression of a recently discovered zebrafish EST, RK011A3G03 in 18 (*k* and *l*) and 24 (*m* and *n*) hpf embryos is shown. RK011A3G03 transcripts are bilaterally expressed in the developing pnd (arrowheads). Lateral views in *b*, *e*, *g*, *i*, *k*, and *m* with dorsal upward and dorsal views in *c*, *d*, *f*, *h*, *j*, *l*, and *n* are shown. In all embryos, anterior is to the left. (*B*) Two recently discovered zebrafish genes expressed in myeloid cells. The *coro1a* is expressed at 18 (*a* and *b*) and 24 (*c*) hpf. Cells expressing both *coro1a* (blue) and *mpo* (red) are observed (*d*, box) and are magnified in *d*. The black arrow in *d* indicates a cell that only expresses *coro1a*. The *nephrosin* is expressed at 72 (*f*) and 120 (*g*) hpf embryos. Cells coexpressing *nephrosin* and *mpo* are observed in the posterior ICM (*h*). A cell coexpressing *nephrosin* and *lysozyme c* is indicated by the black arrow (*i*), whereas cells expressing only *lysozyme c* are indicated by red arrow. Dorsal view of the embryo shown in *b*, all others are viewed laterally with anterior to the left, dorsal upward. (Magnification, ×40 in *d–h*, ×63 in *i*, and ×100 in *e*.) (*C*) Expression patterns of *dab2*. *Dab2* is expressed at 18 (*a*) and 24 (*b*) hpf in pnd and ICM. Expression in the developing vasculature of the ICM includes the da and cv, as indicated. By 72 hpf, *dab2* is expressed in the pnd and cv (*c*). (*d*) A magnified dorsal view of the anterior half of the embryo in *c*. (*e*) A cross section through the region marked with a vertical bar in *d* shows the expression of *dab2* in the pnd. Two-color WISH showing the colocalization of *dab2* and *fli1* (*f* and *g*), but not *scl* (*h* and *i*) at 24 hpf (lateral views of the posterior half of the embryos, anterior to the left, dorsal upward). The expression of *dab2* is shown in red in *f* and *h* and red fluorescence in *g* and *i*. Cells coexpressing *dab2* and *fli1* can be seen in the da and cv (arrowheads). Arrows indicate cells that express *fli1* (*f* and *g*) or *scl* (*h* and *i*) alone. (Magnification, ×40 in *e*.)

cells of the developing posterior blood island (Fig. 3*C f* and *g*, arrows). In contrast, assays for the coexpression of *dab2* with the stem cell gene *scl* showed that these two genes were not coexpressed in cells of the ICM, or in the developing vasculature (Fig. 3*C h* and *i*, arrows).

Discussion

The complete sequencing of the human genome provides an important starting point for the functional characterizations of all human genes. Comparative genomic studies in mouse, fish and other model organisms will help to identify human gene transcription units, and gene "knockout" and "forward" genetic studies in model organisms will be of tremendous value in analyzing the function of these units. The completion of the zebrafish genome sequencing in conjunction with EST cataloging is a major step toward this goal. Until our study, the zebrafish public EST database contained ≈360,000 ESTs, whereas the zebrafish Unigene database had ≈17,000 unique gene entries. In our work, 7,742 EST assemblies of unique sequences and 304 full-length cDNAs were identified from a cDNA library of the zebrafish kidney marrow. Because 44% of these EST assemblies were previously undescribed ones, this study expands both the current zebrafish UniGene and EST databases. Furthermore, we identified 2,225 zebrafish orthologues of human genes, providing useful means to perform further functional analysis of these genes or to compare certain tissue-restricted pathways.

The gene expression profile of a given organ/tissue should reflect its physiological functions. As expected, many orthologues of mammalian genes required for blood and vasculature development are expressed in the zebrafish kidney marrow cDNA library. In general, these orthologues show a very high level of homology and are thought to be involved at different stages of differentiation along with distinct hematopoietic lineages. These findings have provided important molecular evidence for the conservation of genetic programs that regulate vertebrate hematopoiesis and confirmed that the adult kidney marrow is the major site of definitive hematopoiesis in the zebrafish. Of particular interest is the characterization of some evolutionary features of the embryonic/fetal/adult globin genes and related transcription factors between zebrafish and mammals. It seems that the ancestors for both α- and β-globin families have existed before the divergence of the common ancestor of teleost and mammals, whereas distinct globin gene

subtypes in each family might have evolved after this divergence, because no corresponding phylogenetic relationship could be seen among embryonic and adult globins between zebrafish and mammals. On the other hand, the fact that zebrafish globin genes in each family exhibited higher homology to embryonic, but not adult, counterparts in human suggests that these embryonic globin genes should be the evolutionarily most conserved ones. Moreover, the conservation of the transcription factors for globin gene switch across species from zebrafish to man is in support of the view that the regulatory mechanism underlying globin expression remodeling should exist at very early stage of vertebrate evolution.

Previous studies showed that WISH could provide important data to aid understanding gene function. As the first step in an effort to carry out systematic WISH analysis for zebrafish gene expression, we conducted an investigation of genes expressed in zebrafish kidney marrow. Detailed description was made in this study for three zebrafish homologues, *coro1a*, *nephrosin*, and *dab2*, that exhibited specific expression in hematopoietic/vascular tissues, but whose functions in mammals are poorly understood. Although there is currently no functional data on human CORO1A, mouse Coro1A has been shown to contain a WD (trp/asp) repeat and a leucine zipper motif (26) and to reside on phagosomal membranes in cells infected by mycobacteria, thus preventing endocytic transport of these bacteria to lysosomes and promoting their intracellular survival in cultured macrophages (27). This observation is consistent with our results that *coro1a* is expressed in myeloid cells coexpressing *mpo* and indicates that *coro1a* is a marker for mature granulocyte and monocytes. The *nephrosin* gene has not yet been identified in mammals, but has been found in carp (*Cyprinus carpio*). Carp nephrosin is a secreted zinc-coordinating endopeptidase of the astacin family (28). In zebrafish, *nephrosin* is expressed specifically in granulocytes. One could speculate that this protease may function by its secretion into the extracellular matrix, potentially playing a role in either the migration of myeloid cells or their ability to invade other tissues. In support of this hypothesis, carp *nephrosin* has been reported to be able to degrade both mammalian fibronectin and gelatin (28). The colocalization of zebrafish *dab2* with *fli1* expression in developing zebrafish vasculature is an intriguing finding. Human *DAB2* is a candidate tumor suppressor gene in epithelial ovarian cancer (29) and links transforming growth factor β receptors to the Smad pathway in a fibrosarcoma cell line (30). In zebrafish, *dab2* expression is observed in presumptive endothelial cells of the forming vasculature. Although colocalization of *dab2* with *fli1* in the dorsal aorta and axial vein of trunk in 18- to 24-hpf embryos was evident, several differences were also observed. The *fli1* gene is expressed in ventral mesoderm by 12 hpf, whereas *dab2* is not expressed until later (data not shown). In addition, the *fli1* gene is expressed in the head of 24-hpf embryos and the developing posterior blood island, whereas *dab2* expression appears to be restricted to the developing vasculature of the trunk. These results suggest that *dab2* is a bona fide trunk vascular marker during embryogenesis. Taken as a whole, our results demonstrated once again the value of WISH analysis of zebrafish genes as a unique approach in comparative functional genomics.

We thank Dr. Zi Long Wen (Institute of Molecular Agrobiology, Singapore) for providing zebrafish lysozyme C plasmids, George Kourkoulis, Jessica Vinokur, and Walter Saganic in the Dana–Farber Cancer Institute zebrafish facility for excellent fish care, and Pei Hao and Xun Chu for phylogenetic analysis. We thank all members of the Shanghai Institute of Hematology and State Key Lab for Medical Genomics for their support. This work was supported in part by the Chinese National Key Basic Research Project (973), the Chinese National High Tech Program (863), the National Natural Science Foundation of China, the Shanghai Municipal Commission for Science and Technology, the Shanghai Municipal Commission for Education, and National Institutes of Health Grant U01 DK063328.

1. Zon, L. I. (1995) *Blood* **86,** 2876–2891.
2. Bennett, C. M., Kanki, J. P., Rhodes, J., Liu, T. X., Paw, B. H., Kieran, M. W., Langenau, D. M., Delahaye-Brown, A., Zon, L. I., Fleming, M. D., *et al.* (2001) *Blood* **98,** 643–651.
3. Medvinsky, A. & Dzierzak, E. (1996) *Cell* **86,** 897–906.
4. Amatruda, J. F. & Zon, L. I. (1999) *Dev. Biol.* **216,** 1–15.
5. Kudoh, T., Tsang, M., Hukriede, N. A., Chen, X., Dedekian, M., Clarke, C. J., Kiang, A., Schultz, S., Epstein, J. A., Toyama, R., *et al*. (2001) *Genome Res.* **11,** 1979–1987.
6. Crosier, P. S., Bardsley, A., Horsfield, J. A., Krassowska, A. K., Lavallie, E. R., Collins-Racie, L. A., Postlethwait, J. H., Yan, Y. L., McCoy, J. M. & Crosier, K. E. (2001) *Dev. Dyn.* **222,** 637–644.
7. Thisse, C. & Zon, L. I. (2002) *Science* **295,** 457–462.
8. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
9. Mao, M., Fu, G., Wu, J. S., Zhang, Q. H., Zhou, J., Kan, L. X., Huang, Q. H., He, K. L., Gu, B. W., Han, Z. G., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8175–8180.
10. Barbazuk, W. B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J. A., McPherson, J. D. & Johnson, S. L. (2000) *Genome Res.* **10,** 1351–1358.
11. Liu, T. X., Zhou, Y., Kanki, J. P., Deng, M., Rhodes, J., Yang, H. W., Sheng, X. M., Zon, L. I. & Look, A. T. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6136–6141.
12. Zhang, Q. H., Ye, M., Wu, X. Y., Ren, S. X., Zhao, M., Zhao, C. J., Fu, G., Shen, Y., Fan, Y. H., Lu, G., *et al.* (2000) *Genome Res.* **10,** 1546–1560.
13. Westerfield, M. (1993) in *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish Danio (Brachydanio) rerio*, ed. Streisinger, G. (Institute of Neuroscience, Univ. of Oregon, Eugene), pp. 1.1–1.27.
14. Bhat, A., Kolibaba, K., Oda, T., Ohno-Jones, S., Heaney, C. & Druker, B. J. (1997) *J. Biol. Chem.* **272,** 16170–16175.
15. Yamamoto, K., Shibata, F., Yamaguchi, M. & Miura, O. (2002) *Int. J. Hematol.* **75,** 503–507.
16. Slater, D. J., Hilgenfeld, E., Rappaport, E. F., Shah, N., Meek, R. G., Williams, W. R., Lovett, B. D., Osheroff, N., Autar, R. S., Ried, T., *et al.* (2002) *Oncogene* **21,** 4706–4714.
17. Fuchs, U., Rehkamp, G., Haas, O. A., Slany, R., Konig, M., Bojesen, S., Bohle, R. M., Damm-Welk, C., Ludwig, W. D., Harbott, J., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 8756–8761.
18. Megonigal, M. D., Cheung, N. K., Rappaport, E. F., Nowell, P. C., Wilson, R. B., Jones, D. H., Addya, K., Leonard, D. G., Kushner, B. H., Williams, T. M., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2814–2819.
19. Kourlas, P. J., Strout, M. P., Becknell, B., Veronese, M. L., Croce, C. M., Theil, K. S., Krahe, R., Ruutu, T., Knuutila, S., Bloomfield, C. D., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2145–2150.
20. Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C. & Rowley, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12110–12114.
21. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. (2000) *Nature* **404,** 193–197.
22. Ma, Q., Jones, D., Borghesani, P. R., Segal, R. A., Nagasawa, T., Kishimoto, T., Bronson, R. T. & Springer, T. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 9448–9453.
23. Baguna, J. & Garcia-Fernandez, J. (2003) *Int. J. Dev. Biol.* **47,** 705–713.
24. Cooper, S. J., Murphy, R., Dolman, G., Hussey, D. & Hope, R. M. (1996) *Mol. Biol. Evol.* **13,** 1012–1022.
25. Hardison, R. (1998) *J. Exp. Biol.* **201,** 1099–1117.
26. Suzuki, K., Nishihata, J., Arai, Y., Honma, N., Yamamoto, K., Irimura, T. & Toyoshima, S. (1995) *FEBS Lett.* **364,** 283–388.
27. Ferrari, G., Langen, H., Naito, M. & Pieters, J. (1999) *Cell* **97,** 435–447.
28. Hung, C. H., Huang, H. R., Huang, C. J., Huang, F. L. & Chang, G. D. (1997) *J. Biol. Chem.* **272,** 13772–13778.
29. Mok, S. C., Chan, W. Y., Wong, K. K., Cheung, K. K., Lau, C. C., Ng, S. W., Baldini, A., Colitti, C. V., Rock, C. O. & Berkowitz, R. S. (1998) *Oncogene* **16,** 2381–2387.
30. Hocevar, B. A., Smine, A., Xu, X. X. & Howe, P. H. (2001) *EMBO J.* **20,** 2789–2801.