Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development

Gregory Golling[1,3*], Adam Amsterdam[1*], Zhaoxia Sun[1], Marcelo Antonelli[1], Ernesto Maldonado[1], Wenbiao Chen[1], Shawn Burgess[1], Maryann Haldi[1], Karen Artzt[2], Sarah Farrington[1], Shuh-Yow Lin[1], Robert M. Nissen[1] & Nancy Hopkins[1]

*These authors contributed equally to this work.

Published online: 13 May 2002, DOI: 10.1038/ng896

To rapidly identify genes required for early vertebrate development, we are carrying out a large-scale, insertional mutagenesis screen in zebrafish, using mouse retroviral vectors as the mutagen. We will obtain mutations in 450 to 500 different genes—roughly 20% of the genes that can be mutated to produce a visible embryonic phenotype in this species—and will clone the majority of the mutated alleles. So far, we have isolated more than 500 insertional mutants. Here we describe the first 75 insertional mutants for which the disrupted genes have been identified. In agreement with chemical mutagenesis screens, approximately one-third of the mutants have developmental defects that affect primarily one or a small number of organs, body shape or swimming behavior; the rest of the mutants show more widespread or pleiotropic abnormalities. Many of the genes we identified have not been previously assigned a biological role *in vivo*. Roughly 20% of the mutants result from lesions in genes for which the biochemical and cellular function of the proteins they encode cannot be deduced with confidence, if at all, from their predicted amino-acid sequences. All of the genes have either orthologs or clearly related genes in human. These results provide an unbiased view of the genetic construction kit for a vertebrate embryo, reveal the diversity of genes required for vertebrate development and suggest that hundreds of genes of unknown biochemical function essential for vertebrate development have yet to be identified.

Introduction

The development of a multicellular organism requires the integration of diverse biological processes, including growth, differentiation and cellular physiology, as well as cell-type specification and patterning. Analysis of the genetic basis of development in vertebrate animals is following the approach taken in *Drosophila melanogaster* and *Caenorhabditis elegans* of using forward genetic screens to identify the genes necessary for embryonic development. Such genetic screens in zebrafish have shown that it is possible to obtain mutations that are involved in all of the aforementioned processes[1,2]. In addition, the isolation of mutants with defects primarily in the development of one or a few embryonic organs, such as kidney or heart, reveals that the formation of functional organs in zebrafish can be analyzed using a forward genetic approach[1,2].

It has been estimated from large-scale chemical mutagenesis screens in zebrafish that roughly 800 genes can be mutated to yield relatively specific or localized defects during development that can be identified by a visual screen of embryos up to five days post-fertilization[2]. These genes may be involved in the range of processes needed for the development of specific organs and structures. Approximately 1,600 additional genes can be mutated to yield less specific phenotypes or recurring syndromes; these may be genes whose products are required in many cell types. Identification of these 2,400 genes would contribute significantly to understanding vertebrate development. However, the cloning of chemically mutated genes requires either arduous positional cloning or a candidate gene approach. Since the initial description, five and a half years ago, of hundreds of mutants, representing about half or more of the genes that yield specific types of defects, the genes underlying only about 50 mutants have been reported. Most were cloned by the candidate-gene approach and may thus represent only a fraction of the types of genes important for specific developmental processes in zebrafish. Many of these cloned genes encode transcription factors, ligands and their receptors, and many were previously identified as important developmental genes in other species.

[1]*Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.* [2]*The University of Texas at Austin, Institute for Cellular and Molecular Biology, Department of Microbiology, Austin, Texas.* [3]*Current address: Znomics Inc., 2611 SW 3rd Avenue, Suite 200, Portland, Oregon 97201, USA. Correspondence should be addressed to N.H. (e-mail: nhopkins@mit.edu).*

hi#	mutant name	brain	eye	jaw/arches	midline	ear	fins	liver	gut	kidney	muscle	pigment	body shape	motility alone	motility with other	cns necrosis	small head/eyes	retardation	multiple defects	swim bladder not inflated	gene product	phenotypic notes
904	*hi904*																				similar to predicted protein KIAA1323 [*Homo sapiens*]	disorganized brain/CNS, brain hemorrage, bent tail
349	*spg* hi349																				POU domain gene 2 (*spiel ohne grenzen*[a])	1 otolith, mid-brain/hind-brain defects
548	*tcf2* hi548																				vHnf1/transcription factor 2	cystic kidney, pancreas smaller; strong alleles mispattern hindbrain
1463	*cd36l2* hi1463																				lysosome membrane protein II	variable mispatterning of hinbdbrain, notochord defects
528	*copz1* hi528																				nonclathrin coat protein copZ1	degeneration of eye after day 4
215	*hi215*																				similar to ATP-dependent RNA helicases	bent ceratohyal cartilage
307	*b3gat3* hi307																				similar to β 1, 3 glucuronyltransferase	branchial arches and jaw misshapen
572	*hi572*																				similar to predicted protein FLJ20508 [*Homo sapiens*]	branchial arches short, ceratohyal bent posteriorly
1116a	*hi1116a*																				similar to predicted protein DT1P1A10 [*Homo sapiens*]	bent ceratohyal cartilage, misshapen arches
1548	*hi1548*																				similar to protein product DKFZP434B168 [*Homo sapiens*]	jaw and neurocranium abnormal shape
1002	*csnk1a1* hi1002																				casein kinase 1α	pectoral fins and jaw misshapen, cartilage appears "wrinkled"
954	*uxs1* hi954																				UDP-glucuronic acid decarboxylase	cartillage present but does not stain with Alcian blue
975	*sqt* hi975																				nodal-related-1 (*squint*[b])	mild eye cyclopia, u-shaped somites
229	*smu* hi229																				smoothened (*smooth muscle omitted*[c])	mild eye cyclopia, u-shaped somites
429	*hi429*																				similar to predicted protein DJ434O14.5 [*Homo sapiens*]	stomach epithelium abnormal, small liver
2092	*cad1* hi2092																				caudal homeobox 1	short body, no yolk extension
428	*hi428*																				similar to protein product DKFZp434H247 [*Homo sapiens*]	disorganized muscle striations, reduced circulation in tail
577a	*atp6e* hi577a																				vacuolar ATP synthase subunit E	reduced pigment, small otoliths, arches/jaw misshapen, touch insensitive
923	*hi923*																				vacuolar ATP synthase subunit H	reduced pigmentation, day 3 brain necrosis
1207	*atp6l* hi1207																				vacuolar ATP synthase 16 kd proteolipid subunit	reduced pigment in eyes and body, ruffled pectoral fins, mild brain necrosis
2499a	*vps18* hi2499a																				deep orange/vacuolar protein sorting protein 18	reduced melanophores and iridophores
318	*rrm1* hi318																				ribonucleotide reductase protein r1 class I	s-curved body, misshapen head
459	*hi459*																				similar to ADP-ribosylation factor-related protein (ARF2-related)	ventrally curved tail
486	*knsl1* hi486																				kinesin-related motor protein EG5	curved body
1688	*kny* hi1688																				glypican-6 (*knypek*[d])	shortened tail, somites U-shaped
1780b	*ppt* hi1780b																				wnt5 (*pipetail*[e])	small, kinked tail, fused somites, undulating notochord, broad head
1059	*neurod3* hi1059																				neurogenin related protein-1 ([f])	gaping jaw, bent body, touch insensitive
472	*hi472*																				similar to DKFZp564L2423 [*Homo sapiens*] and VIP36 [*Canis familiaris*]	touch insensitive
1715	*arnt2* hi1715																				aryl hydrocarbon receptor nuclear translocator 2A	touch insensitive
199	*u2af1* hi199																				splicing factor U2AF, 35 KD subunit	inflated hindbrain ventricle
258	*rpl35* hi258																				60S ribosomal protein L35	inflated hindbrain ventricle
297	*hi297*																				similar to protein product FLJ10498 [*Homo sapiens*]	
383a	*cct3* hi383a																				chaperonin containing T-complex protein-1, γ subunit	thin/no yolk extension
447	*dtl* hi447																				denticleless homolog	dorsally bent body
601	*snrpd1* hi601																				small nuclear ribonucleoprotein D1	
688	*rrm2* hi688																				ribonucleotide reductase protein r2 class I	dorsally bent body
783	*pole2* hi783																				DNA polymerase ε subunit B	
821a	*hi821a*																				similar to predicted protein KIAA0007 [*Homo sapiens*]	
1026	*rps18* hi1026																				40S ribosomal protein S18	inflated hindbrain ventricle
1045	*stka* hi1045																				serine/threonine kinase a, aurora-related	severe brain necrosis at 24 hpf
1055a	*mak16l* hi1055a																				MAK16-like homolog [*Schizosaccharomyces pombe*]	thin/no yolk extension
1159	*sf3b4* hi1159																				splicing factor 3b subunit 4	
1371	*snrpc* hi1371																				U1small nuclear ribonucleoprotein C	dorsally bent body, swims in circles
1373	*hi1373*																				similar to predicted protein MGC1346 [*Homo sapiens*]	curved body, constricted yolk sac extension
1411	*mcm7* hi1411																				DNA replication licensing factor MCM7	
1464	*ddx19* hi1464																				DEAD-box RNA helicase ddx19	severe brain necrosis at 24 hpf
1581	*hi1581*																				nucleolar phosphoprotein Nopp34 homolog [*Homo sapiens*]	
319	*mcm3* hi319																				DNA replication licensing factor MCM3	
491	*tcp1* hi491																				chaperonin containing T-complex protein-1, α subunit	
550	*hi550*																				SNF2-related protein	
558	*vars2* hi558																				valyl-tRNA synthetase	
577b	*rps5* hi577b																				40S ribosomal protein S5	
642	*cct2* hi642																				chaperonin containing T-complex protein-1, β subunit	
800a	*cct7* hi800a																				chaperonin containing T-complex protein-1, η subunit	
889	*taf7* hi889																				taf7 RNA polymerase II, TBP-associated factor 55kd	
994	*hi994*																				similar to protein product FLJ22611 [*Homo sapiens*]	underdeveloped gut, missing branchial arches
1116b	*ef1a* hi1116b																				translation elongation factor eEF1α	
1143	*smc4l1* hi1143																				structural maintenance of chromosomes family member SMC4-like 1	
1182	*terf2* hi1182																				telomeric repeat factor a	protruding jaw, very small eyes
1244	*mcm2* hi1244																				DNA replication licensing factor MCM2	small eyes, missing branchial arches
1257	*hi1257*																				similar to protein product AK027570 [*Homo sapiens*]	
1262	*sill* hi1262																				TAL1 (SCL) interrupting locus-like	dorsally bent body
1284	*rpl24* hi1284																				60S ribosomal protein L24	
1433	*adss* hi1433																				adenylosuccinate synthetase, non-muscle	
1579	*ncbp2* hi1579																				nuclear cap binding protein subunit 2	
1841b	*odc* hi1841b																				ornithine decarboxylase	
2696	*ccna2* hi2696																				cyclin A2	
694	*ppp1r10* hi694																				protein phosphatase 1 nuclear targeting subunit, PNUTS	
887	*timm23* hi887																				mitochondrial inner membrane translocating protein, TIMM23	
1447	*hi1447*																				similar to ATP-dependent RNA helicases	
591	*ars2* hi591																				arsenite-resistance protein 2 [*Homo sapiens*]	
1058	*sec61a* hi1058																				Sec61 α-subunit	cartilage doesn't stain with Alcian blue, small otoliths
1072	*baf53a* hi1072																				BRG1/brm-associated factor 53a	
1618	*hdac1* hi1618																				histone deacetylase 1	do not stain with Alcian blue
526a	*slc25a5* hi526a																				ADP/ATP carrier protein, fibroblast isoform	

Fig. 1 Phenotypic classification of 75 insertional zebrafish mutants. Each colored box represents one or more of the phenotypes of the corresponding mutant (hi#). The disrupted gene and brief phenotypic notes are also shown for each mutant. Gray boxes indicate the different types of general 'syndromes'. All genes of unknown biochemical function are shown in red. Chemically induced mutants that are known to be allelic are shown in parentheses in the gene product column; references are indicated by superscripted letters: [a]24,25; [b]26; [c]27,28; [d]29; [e]30; [f]unnamed mutant (M. Granato, pers. comm.).

We designed a method of insertional mutagenesis for zebrafish using a Moloney murine leukemia–based retroviral vector as a mutagen[3,4]. We find that most retrovirally mutated genes can be tentatively identified in as little as two weeks. The ease of gene cloning has allowed us to take a relatively unbiased approach to identifying the genes required for early development as we do not have to select a small number of mutants for positional cloning and are not biased towards known genes, as is the case with a candidate-gene approach. Notably, the types of genes we have identified encode a wide range of proteins, including many without known biological or biochemical functions.

Results

Identification of retrovirus-induced mutations

We use previously described[3,4] methods for inducing insertions in the fish germ line using high-titer pseudotyped mouse retroviruses, for identifying embryonic mutations and for cloning the mutated genes. After identifying a proviral insert potentially linked to the mutant phenotype, we clone DNA flanking this insert using inverse PCR. In slightly less than 50% of the cases, sequencing this junction fragment yields a candidate gene by homology search against the National Center for Biotechnology Information (NCBI) database. In the remaining cases, we obtain and sequence additional genomic DNA adjacent to one or both sides of the original junction fragment using small chromosomal walks. In about two-thirds of all cases, this two-step procedure yields a candidate gene. We then use RT–PCR and RACE to obtain the rest of the cDNA. The position of the insertion within the gene for each mutant is shown in Web Fig. A online.

To confirm that the correct junction fragment (and gene) have been cloned, we carry out linkage analysis using primers designed to amplify different-sized products from chromosomes with or without the putative mutagenic insert in a PCR-based assay. We have shown that mutant embryos are invariably homozygous with respect to the putative mutagenic insertion, whereas their wildtype siblings are heterozygous or non-transgenic with respect to the insert. We genotype a minimum of 24 mutant and 24 wildtype embryos in most cases (see Web Fig. A online). We consider that if no recombinants are seen, the insert should lie no further than 3 cM (2 Mb) from the mutation responsible for the mutant phenotype. Thus, if no other insertion showing linkage to the phenotype is present, we consider the insert with tight linkage to be the probable cause of the mutation. A caveat is that during the course of the screen, we have identified mutants that are clearly not linked to a detectable proviral insert and probably result from spontaneous mutations. Thus, it is possible that a proviral insert could be linked to a mutation, but not be its cause. We estimate that this situation could escape detection in at most 1% of our mutants.

To obtain further evidence that the correct gene has been identified, we use RT–PCR or *in situ* hybridization to determine whether expression of the gene is disrupted by the appropriate insertion. In all cases reported here that were analyzed in this way (37 cases), gene expression was reduced or ablated, or the transcript was altered, as a result of the insertion shown to be linked to the mutant phenotype (see Web Fig. A online).

In this screen, we kept all embryos with mutant phenotypes visible in a dissecting microscope by five days post-fertilization. As seen in chemical mutagenesis screens, about one-third of the insertional mutants show highly specific developmental defects involving one or a few organ systems, whereas the majority have one of several more general, frequently recurring syndromes[5]. Similar to most embryonic mutants isolated in zebrafish by any method of screening, almost all of our mutants are recessive-lethals, and homozygous mutant embryos die between three and ten days of age.

Classification of mutant phenotypes

The first 75 embryonic-lethal mutants in this screen for which the disrupted gene has been identified are listed in Fig. 1 (for an expanded version, see Web Fig. A online). Mutants are grouped according to their phenotypic defects, following a previously described classification system[1,2]. Most designations were made

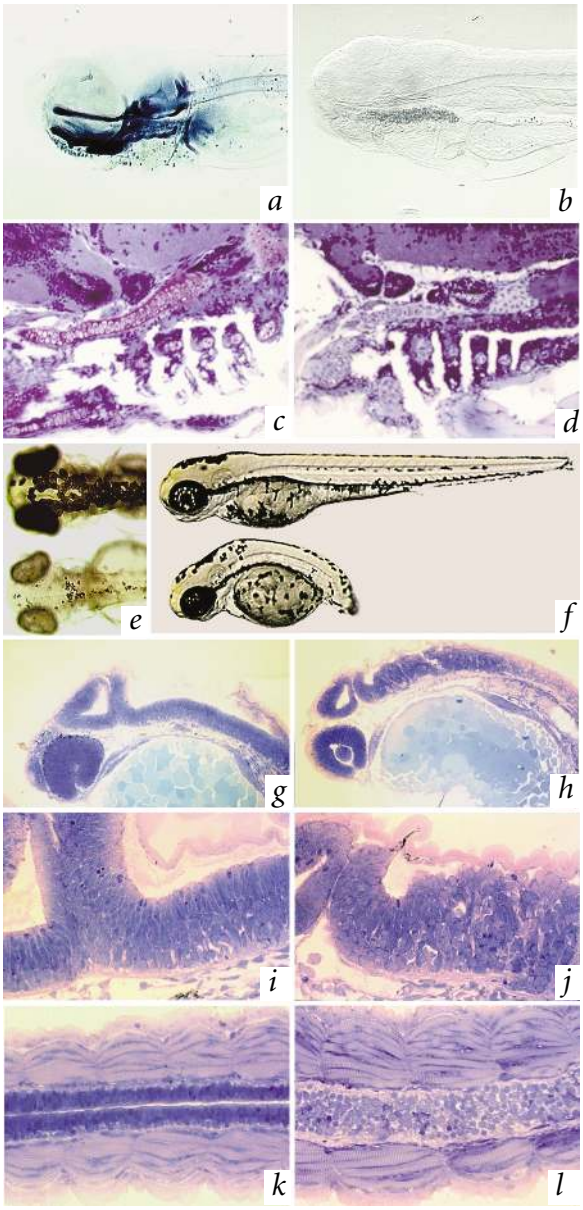

Fig. 2 Examples of mutant phenotypes isolated by insertional mutagenesis. ***a–d***, Hi954 mutant embryo showed cartilage defects, in that it did not stain with Alcian blue, as seen in these lateral views of wildtype (*a*) and hi954 (*b*) embryos at 5 d. This lack of staining did not correspond entirely to a loss of either jaw/arch structures or cartilage cells, as these features were present in sagittal sections of mutant embryos (compare wild type (*c*) with hi954 (*d*), cartilage cells stained in pink with Fuscian). ***e***, A mutation affecting pigment is represented by hi923. These mutants (bottom) had melanophores, yet they appeared much smaller than those in wildtype embryos (above). ***f***, In hi2092 mutants (bottom), the posterior portion of the body was significantly reduced, as shown in these day-3 embryos. ***g–l***, 33h hi904 mutant embryo had a severe morphological defect in the brain and CNS (wildtype in *g*,*i*,*k*; hi904 in *h*,*j*,*l*). Higher magnifications of sagittal sections showed a high degree of disorganization of neuronal cells in the brain (compare *i* and *j*) and neural tube (compare *k* and *l*).

Table 1 • Types of genes bearing insertional and chemically induced mutants

Protein category	Phenotype:	Type of mutant: Insertional Specific	Insertional General	Chemical Specific[a]
DNA replication and cell cycle		1	6	0
Chromatin/transcription associated factors		0	7	0
Transcription factors		6	0	21
RNA processing		0	6	0
Helicases		1	2	0
Translation/post-translation		2	10	0
Protein folding		0	4	0
Kinases		1	1	0
Organelle		7	2	0
Receptors and ligands		3	0	22
Metabolic enzymes		2	3	3
Other		2	2	4
Unknown function		7	9	0
Total		32	52	50

[a]All of the chemical mutants for which the mutated gene has been cloned have specific developmental phenotypes. The numbers indicate the total mutants for each category. The numbers for the chemical mutants include all those published or reported as of 1 October 2001.

solely on the basis of examination by low-power microscopy. In some cases, mutant embryos were sectioned, stained with Alcian blue to visualize cartilage or stained with acridine orange to identify those mutants with apoptotic tissues. As considerable study is needed to accurately describe the phenotype of any mutant, the classifications given here must be considered preliminary. In some cases, further study may identify additional defects. For example, a mutation in *vhnf1* was initially found to result in a kidney mutant. However, further analysis, including study of additional alleles, showed that the gene is required for development of kidney, pancreas, liver and hindbrain[6].

As in large-scale chemical mutagenesis screens, insertional mutants with unique, relatively specific developmental defects include those in which phenotypes are observed primarily in the brain, eyes, jaw, arches or cartilages, midline, ear, fins, liver, gut, kidney, muscle, pigment, body shape, motility or touch insensitivity in the absence of any visible structural abnormality, and motility or altered touch sensitivity in combination with some degree of visible morphological defects. Mutants classified as having more general developmental defects include those with extensive cell death in the central nervous system (CNS), those with a small head and eyes (some of which initially showed limited cell death in the head, some of which did not), embryos that show retardation, mutants that have several defects (meaning that many organs and structures are visibly affected while others seem grossly normal) and mutants in which the only apparent phenotype is failure to inflate the swim bladder (although this last trait is seen in almost all embryonic mutations of zebrafish, regardless of what other defects are also present). The large phenotypic class of mutants with these common and general defects was discarded in chemical mutagenesis screens[1,2]. As our method of mutagenesis allows rapid identification of mutated genes, we kept these mutants to generate a more complete picture of the genes required for the development of the five-day-old embryo.

Although the number of mutants described here is small and the characterization preliminary, the range of phenotypes we observed seems indistinguishable from the spectrum of phenotypes induced by ENU mutagenesis[1,2,5]. Of the 75 mutants, 6 harbor insertional alleles of chemically induced or radiation-induced mutations for which the mutated gene has been reported (Fig. 1); others may prove to be allelic when complementation studies are carried out.

Pictures of four insertional mutants with relatively specific developmental defects are shown in Fig. 2. Mutations in a variety of genes produce phenotypes involving cartilage. For example, hi954, a mutant in which the cartilage does not stain with Alcian blue, but in which cartilage cells can be seen in sections (Fig. 2*b*,*d* versus Fig. 2*a*,*c*), results from a disruption of the gene encoding UDP-glucuronic acid decarboxylase. The ENU-induced mutant *jekyll* shows a very similar Alcian blue staining pattern, although it has in addition a cardiac valve defect not seen in hi954 (refs 7,8). The gene disrupted in *jekyll* was recently found to encode UDP-glucose dehydrogenase[9].

Most mutants with pigmentation abnormalities that we have isolated so far have disruptions in genes encoding proteins associated with cytoplasmic organelles. The mutants hi577a, hi923 (Fig. *2e*) and hi1207, as well as the previously reported[4] hi112, have reduced pigmentation in both the body and eyes and all are the result of mutations in different subunits of vacuolar ATP synthase. The gene disrupted in another mutant with a pigmentation abnormality, hi1463, encodes a protein similar to lysosomal membrane protein II (LIMP II). Pigmentation mutant 2499a results from a lesion in a gene encoding a vacuolar protein sorting protein that is homologous to *D. melanogaster* deep orange. A number of mutations affecting eye color in *D. melanogaster* and coat color in mouse involve genes that encode membrane components of vesicles involved in pigmentation[10,11]. As many of these proteins are also expressed in lysosomes[12,13], it is not unexpected that strong alleles of these mutations would result in lethality.

A mutant (hi2092) with an insertion in the well known developmental gene *caudal* has a shortened trunk and tail and no yolk extension (Fig. 2*f*). The mutant hi904 has severe disorganization of the brain and CNS (compare Fig. 2*g* with Fig. 2*h*); in higher-magnification sections, the cells appear more loosely organized than those of wildtype embryos (Fig. 2*i–l*). The gene mutated in hi904 has some recognizable motifs but no known biochemical function and has orthologs in fly and human.

Genes required for early vertebrate development

The types of genes mutated in the two broad phenotopic classes of mutants, those with relatively specific and those with widespread defects, including insertional mutants previously reported from our pilot screen[3,4,14,15], are shown in Table 1. It has been suspected that because zebrafish embryos receive maternally encoded gene products, development can proceed to multicellular stages even if the embryo harbors a mutation in a cell-essential gene[16]. It has been thought that many mutants with so-called nonspecific developmental defects may result from mutations in cell-essential genes[1]. By contrast, mutants with more specific developmental phenotypes might result from lesions in genes required specifically for normal development and function of multicellular structures, or possibly from lesions in housekeeping genes with tissue-specific expression or with more than one biochemical function. Although the data are limited, our findings suggest that genes required for protein synthesis, RNA processing, DNA replication and chromatin assembly

often give rise to nonspecific developmental phenotypes when mutated. As expected, genes encoding transcription factors, receptors and ligands are well represented among mutants with specific developmental phenotypes; however, so are many genes encoding organelle proteins, a diverse array of other types of genes and many genes of unknown biochemical function.

One of the largest categories of genes responsible for both phenotypic classes of mutants is that of genes for which a biochemical function cannot be predicted. These account for 20% of the mutants we have identified. The proteins that these genes are predicted to encode differ in their degree of novelty. One protein (hi459) and its closest human homolog resemble a protein of known function in mammals (ADP-ribosylation factor–related), but the similarity is not sufficient to conclude that they would carry out the same biochemical or cellular function (see Web Fig. B online). Some proteins share recognizable motifs with proteins of known biochemical or cellular functions (hi904, zinc finger; hi447 and hi821a, WD40 repeats). The eleven others do not have any identifiable protein motifs (searched against Conserved Domain Database at NCBI). For all of the genes, there is either a clearly identifiable human ortholog, or at least a human gene with some similarity that can be identified (using protein-protein BLAST (blastp) or translated db (tblastn) searches against the nr and human genome databases, $P < 10^{-29}$). Some do not yet have recognizable orthologs in *D. melanogaster* or in yeast (hi572 and hi1262) and may thus be encoded by genes specific to or conserved only in vertebrate development. Others (hi428, 904, 1548, 591) do not have recognizable orthologs in yeast and could thus be specific to animals. However, divergence in amino-acid sequence could make it difficult to identify the orthologs of some genes in these species. Some of the genes of unknown biochemical function are highly conserved from yeast to mammals (see Web Fig. C online).

Discussion

Large-scale genetic screens that used ENU as a mutagen in zebrafish provided the first broad view of the types of embryonic and early larval mutants that can be obtained in this species, and allowed rough estimates of the number of genes required for early development[1,2]. However, the difficulty of cloning the mutated genes precluded obtaining an equally revealing picture of the genes required for early developmental processes in zebrafish. The results presented here provide the first large-scale, unbiased view of the genes required for the development of a vertebrate embryo. They reveal an array of genes, including genes probably involved in cellular biological processes and physiology. Although these functions have been thought of as distinct from development, they have recently been shown, in some cases, to be inseparable even from specific developmental processes[17]. Table 1 shows a comparison of the types of genes cloned in our lab with those cloned and reported by all other zebrafish labs. Other labs have used primarily candidate-gene cloning, and less frequently purely positional cloning, of chemical or radiation-induced mutants with specific developmental phenotypes. The bias towards genes encoding transcription factors, receptors or ligands presumably results from the bias imposed by selecting candidate genes in advance, and from a bias in choosing mutants with particular types of developmental defects for the laborious process of candidate or positional cloning.

Some of the genes we have identified are important in growth control, are involved in human diseases or both. For example, *pescadillo* (hi2), a BRCT motif–containing gene identified in our pilot screen and required for the normal size of some but not all embryonic organs[15], was recently identified as a gene whose expression is elevated in p53-deficient tumor cell lines and is thought to be involved in cell-cycle check points[18,19]. The transcription factor gene *vhnf1*, mutated in the kidney and pancreas mutant hi548, underlies a genetic form of human diabetes, maturity onset diabetes of the young type V, in which individuals have kidney abnormalities in addition to diabetes[20,21]. It seems likely that other genes we identify in our screen will also be involved in human diseases, possibly including diseases for which the genetic basis has not yet been determined.

A concern in using insertional mutagens is whether they are biased in their insertion sites and whether they can induce mutations in all genes. That retroviruses have some bias in integration sites in zebrafish is suggested by the fact that when more than one insertional allele of a gene is recovered, it is often the case that the insertions lie close together. So far, we have recovered 320 insertional mutants for which we have cloned genomic DNA flanking the mutagenic insert; these represent 254 loci: 205 loci (81%) are represented by a single allele, 37 loci (15%) by 2 alleles, 10 loci (4%) by 3 alleles, and 1 locus each by 4 and 7 alleles. Because the single allele class is so high and the data do not fit a Poisson distribution, it is impossible to estimate the total number of genes mutable by retroviral insertion, except that it must be at least several-fold higher than the number recovered so far. The allele distribution in large-scale ENU mutagenesis screens also did not fit a Poisson distribution[1,2], indicating that all genes are not mutated with equal frequency with either mutagen, and that retroviruses may be no more biased in the genes they mutate than are chemical mutagens.

Upon the completion of our current screen, we will have isolated insertional mutations in around 450–500 genes, roughly 20% of the 2,400 genes estimated to be required for the development of a zebrafish embryo. This will include the identification of approximately 100 genes of unknown biochemical function to which we will be able to assign a biological role. As we will clone all of these mutated genes without selecting for certain types of phenotypes, we should be able to estimate the number of different types of gene products—transcription factors, signaling molecules and vesicle trafficking proteins and so on—that are essential for embryonic development or for the development of any given embryonic structure. Finally, this large collection of mutants will be a valuable community resource, supplementing the large number of chemical mutants, with the benefit that the mutated gene will have already been identified.

Methods

Mutagenesis. We carried out the preparation of virus, injection of embryos and breeding and screening scheme as previously described[4]. Mutant lines will be made available upon request.

Genotyping embryos. For linkage analysis, we sorted embryos from heterozygous parents into phenotypically wildtype and mutant groups. In most cases, we genotyped 24 embryos of each group by PCR. We used a pair of genomic primers flanking the responsible mutagenic viral insertion and a viral-specific primer in a single reaction. A viral insertion leads to amplification between the viral-specific primer and one of the genomic primers, while disrupting amplification between the genomic pair. Thus, wildtype and mutated DNA yield distinctive PCR products. In a few cases, we determined genotypes by Southern blot using genomic sequence flanking the insertion as a probe, as previously described[3].

cDNA cloning. We identified putative exons by comparing the sequences of the genomic DNA adjacent to the mutagenic insert with the public databases, using BLAST[22]. We used either RT–PCR or RACE to obtain the rest of the cDNA. If homology was found to a zebrafish expressed sequence tag for which both 5′ and 3′ sequence was available, we used RT–PCR; otherwise, we carried out RACE using the SMART RACE kit (Clontech). The source of RNA for either approach was a pool of embryos at 1 d, 2 d, 3 d, 4 d and 5 d post-fertilization.

Alcian blue staining. We carried out cartilage-staining procedures as previously described[23], except that HCl was used in place of acetic acid and the pH of the final solution was 1.0.

Accession numbers. Accession numbers for mutated genes can be found in Web Fig. A online.

Note: Supplementary information is available on the Nature Genetics website.

Acknowledgments

We thank the following individuals for maintenance of our zebrafish colony: T. Such, S. Fuss, S. Farrington, M. Munsey and J. Whitehouse. We thank T. Angelini and A. Corcoran for raising of baby fish, J. Anderson, N. Anderson, E. Choi, J. Davenport, J. Montgomery and B. Sarkisian for assistance in screening for mutants; K. Anderson and M. Cunningham for identifying and maintaining the mutant lines of fish; B. Benfield for contributions in maintaining the laboratory and M. Lee for contributions to gene identification. We are grateful to B. Bosselman of Amgen for his continued support. This work was supported by grants from the National Center for Research Resources of the National Institutes of Health (to N.H.), Amgen (to N.H.) and the NIH (to the Center for Cancer Research at MIT). G.G. was supported by a fellowship from Amgen. A.A. was supported by a fellowship from the Ford Foundation. M.A. and E.M. were supported by fellowships from the Pew Foundation. R.N. was supported by the Anna Fuller Fund Fellowships in Molecular Oncology.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 18 January; accepted 24 April 2002.

1. Driever, W. *et al.* A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46 (1996).
2. Haffter, P. *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36 (1996).
3. Gaiano, N. *et al.* Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383**, 829–832 (1996).
4. Amsterdam, A. *et al.* A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* **13**, 2713–2724 (1999).
5. Mullins, M.C., Hammerschmidt, M., Haffter, P. & Nusslein-Volhard, C. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.* **4**, 189–202 (1994).
6. Sun, Z. & Hopkins, N. *vhnf1*, the MODY5 and familial GCKD-associated gene, regulates regional specification of the zebrafish gut, pronephros, and hindbrain. *Genes Dev.* **15**, 3217–3229 (2001).
7. Stainier, D.Y. *et al.* Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development* **123**, 285–292 (1996).
8. Neuhauss, S.C. *et al.* Mutations affecting craniofacial development in zebrafish. *Development* **123**, 357–367 (1996).
9. Walsh, E.C. & Stainier, D.Y. UDP-glucose dehydrogenase required for cardiac valve formation in zebrafish. *Science* **293**, 1670–1673 (2001).
10. Lloyd, V., Ramaswami, M. & Kramer, H. Not just pretty eyes: *Drosophila* eye-colour mutations and lysosomal delivery. *Trends Cell Biol.* **8**, 257–259 (1998).
11. Jordan, S. & Beermann, F. Nomenclature for identified pigmentation genes in the mouse. *Pigment Cell Res.* **13**, 70–71 (2000).
12. Diment, S., Eidelman, M., Rodriguez, G.M. & Orlow, S.J. Lysosomal hydrolases are present in melanosomes and are elevated in melanizing cells. *J. Biol. Chem.* **270**, 4213–4215 (1995).
13. Bhatnagar, V. & Ramalah, A. Characterization of Mg2+-ATPase activity in isolated B16 murine melanoma melanosomes. *Mol. Cell. Biochem.* **189**, 99–106 (1998).
14. Becker, T.S., Burgess, S.M., Amsterdam, A.H., Allende, M.L. & Hopkins, N. *not really finished* is crucial for development of the zebrafish outer retina and encodes a transcription factor highly homologous to human nuclear respiratory factor-1 and avian initiation binding repressor. *Development* **125**, 4369–4378 (1998).
15. Allende, M.L. *et al.* Insertional mutagenesis in zebrafish identifies two novel genes, *pescadillo* and *dead eye*, essential for embryonic development. *Genes Dev.* **10**, 3141–3155 (1996).
16. Kane, D.A. *et al.* The zebrafish early arrest mutants. *Development* **123**, 57–66 (1996).
17. Pearson, H. Two become one. *Nature* **413**, 244–246 (2001).
18. Charpentier, A.H. *et al.* Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Res.* **60**, 5977–5983 (2000).
19. Kinoshita, Y. *et al.* Pescadillo, a novel cell cycle regulatory protein abnormally expressed in malignant cells. *J. Biol. Chem.* **276**, 6656–6665 (2001).
20. Iwasaki, N. *et al.* Liver and kidney function in Japanese patients with maturity-onset diabetes of the young. *Diabetes Care* **21**, 2144–2148 (1998).
21. Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor-1 β gene (TCF2) associated with MODY. *Nature Genet.* **17**, 384–385 (1997).
22. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
23. Schilling, T.F. *et al.* Jaw and branchial arch mutants in zebrafish I: branchial arches. *Development* **123**, 329–344 (1996).
24. Belting, H.G. *et al. spiel ohne grenzen/pou2* is required during establishment of the zebrafish midbrain-hindbrain boundary organizer. *Development* **128**, 4165–4176 (2001).
25. Burgess, S., Reim, G., Chen, W., Hopkins, N. & Brand, M. The zebrafish *spiel-ohne-grenzen* (*spg*) gene encodes the POU domain protein Pou2 and is essential for formation of the midbrain, hindbrain and for pre-gastrula morphogenesis. *Development* **129**, 905–916 (2002).
26. Feldman, B. *et al.* Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* **395**, 181–185 (1998).
27. Chen, W., Burgess, S. & Hopkins, N. Analysis of the zebrafish smoothened mutant reveals conserved and divergent functions of hedgehog activity. *Development* **128**, 2385–2396 (2001).
28. Varga, Z.M. *et al.* Zebrafish smoothened functions in ventral neural tube specification and axon tract formation. *Development* **128**, 3497–3509 (2001).
29. Topczewski, J. *et al.* The zebrafish glypican *knypek* controls cell polarity during gastrulation movements of convergent extension. *Dev. Cell* **1**, 251–264 (2001).
30. Rauch, G.J. *et al.* Wnt5 is required for tail formation in the zebrafish embryo. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 227–234 (1997).