Original Articles

MARINE BIOTECHNOLOGY

Generation of Aberrant Transcripts of and Free DNA Ends in Zebrafish *no tail* Gene

Kimi Yamakoshi, Yuji Shishido, Nobuyoshi Shimoda*

Japan Science and Technology Corporation (JST), Institute for Genome Research, University of Tokushima, 3-18-15 Kuramoto, Tokushima, 770-8503, Japan

Received: 29 August 2003 / Accepted: 19 February 2004 / Online publication: 8 June 2005

Abstract

The zebrafish *no tail* gene (*ntl*) is indispensable for the formation of the notochord and the tail structure. In a wild-type zebrafish population, we occasionally observed adult zebrafish with a narrow or no tailfin. This led us to examine the hypothesis that the activity of *ntl* was somehow genetically unstable. Here we present two findings regarding the gene. First, approximately 3% of *ntl* transcripts were aberrant; most of them carried deletions at various positions. Second, free, DNA double-stranded ends (DSEs) were formed at an AT dinucleotide repeat in *ntl*. DSEs were also generated in another zebrafish gene, *noggin2* (*nog2*). DSEs in *ntl* and *nog2* had common characteristics, which suggested that the AT repeats in these genes elicited DSEs by blocking progression of the replication.

Key words: Zebrafish — *no tail* — double-stranded end — aberrant transcript — microsatellite repeat

Introduction

In some laboratories, including ours, it has been observed that regardless of their strains, zebrafish with an abnormal caudal fin sporadically appear in wild-type zebrafish colonies, albeit at very low frequency (on average, <0.1%) (Figure 1). As crosses between fish without tailfins produced only normal siblings, it was possible that the phenotype was caused by the instability of a gene required for the caudal region of the body. We have been interested in genetic instability, so by finding the gene, we hoped to unravel one of the mechanisms that causes it. Because the tailfin phenotype was reminiscent of mouse *Brachyury* (*T*) heterozygous mutants, we hypothesized that the gene was the *no tail* (*ntl*), a homologue of the mouse *Brachyury* (*T*) gene, and that altered tails could be associated with sporadic generation of aberrant *ntl* messenger RNA. Null and dominant mutations in *Brachyury* cause short-tail and no-tail phenotypes, respectively, in heterozygous mice (MacMurray and Shin, 1988; Herrmann et al., 1990). The phenotypes are due to insufficient notochord cells in the tail anlage, the most posterior structure of mouse embryos (Yanagisawa, 1990; Herrmann and Kispert, 1994). A short-tail phenotype has also been observed in *Brachyury* heterozygous dogs (Haworth et al., 2001). *Brachyury* encodes a DNA-binding protein (Herrmann and Kispert, 1993) and is specifically expressed in nascent mesoderm and in differentiating notochord (Herrmann and Kispert, 1994).

The zebrafish *Brachyury* homologue was named *no tail* (*ntl*). The expression of *ntl* starts just before gastrulation and persists during gastrulation in cells of the germ ring and of presumptive notochord in zebrafish embryos (Schulte-Merker et al., 1992). At the end of gastrulation, the expression remains in cells of the presumptive notochord and the tailbud, the most posterior cells (Schulte-Merker et al., 1992). Expression of *ntl* became almost undetectable by 48 hours after fertilization and was never observed in adult fish (Schulte-Merker et al., 1992). Embryos that are *ntl* homozygous lack a differentiated notochord and a proper posterior structure, as is the case for *Brachyury* homozygous embryos (Halpern et al., 1993; Odenthal et al., 1996; Stemple et al., 1996). Thus the expression pattern and homozygous mutant phenotypes of *ntl* resemble those of *Brachyury*.

The experiments revealed that on average 3% of *ntl* transcripts were aberrant. We then searched for

*Present address: Laboratory of Developmental Genetics, Department of Biological Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
Correspondence to: Nobuyoshi Shimoda; E-mail: shimoda@biol.s.u-tokyo.ac.jp

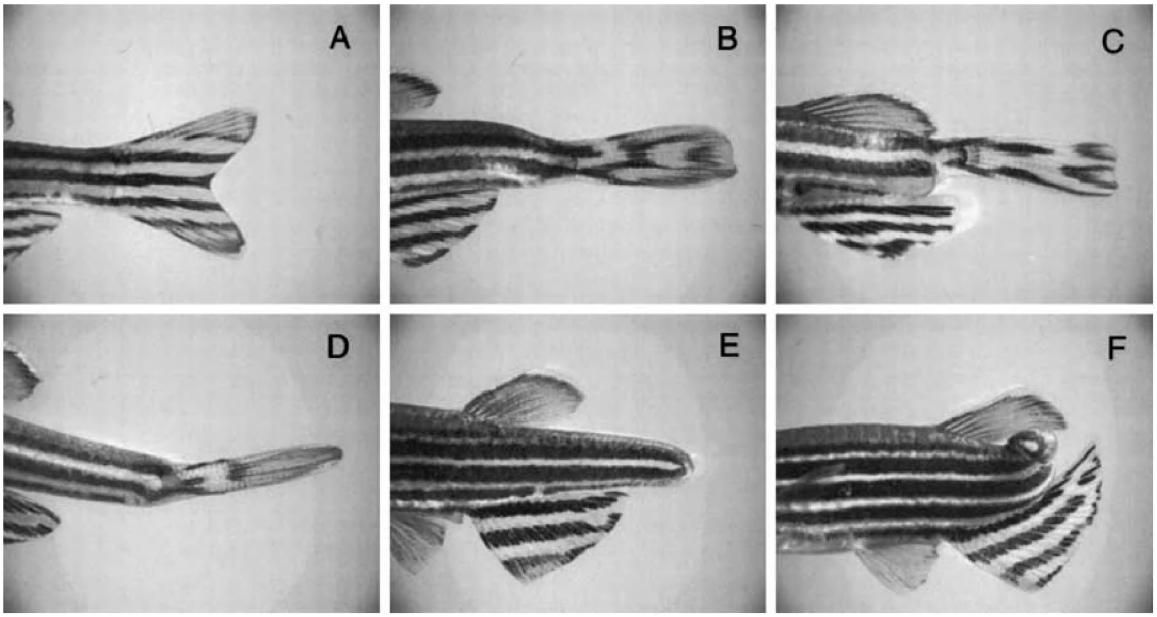

Fig. 1. Variations of zebrafish tailfins. **A**: Normal tailfin. **B–D**: Narrow tailfin. **E, F**: No tailfin.

the causative factor behind generation of the aberrant transcripts and found a long AT dinucleotide repeat in an intron of *ntl*, in which formation of DNA double-stranded ends (DSEs) could be detected.

Materials and Methods

Extraction of Zebrafish Genomic DNA. Adult fish (1 year old) were homogenized with small scissors in 500 μl of lysis buffer containing 100 mM NaCl, 20 mM Tris (8.0), 50 mM EDTA, 0.5% Sodium dodecylsulfate (SDS), and more than 6 U of Proteinase K (Roche), and then incubated for 2 hours at 55°C with occasional vigorous shaking. Depending on the size of fish, we increased the volume of the lysis buffer up to 2 ml. Embryos were digested in the same buffer for 1 hour without homogenization. When the solution came clear, 210 μl of 5 M NaCl was added, and genomic DNA was then extracted once with chloroform, twice with phenol–chloroform–isoamyl alcohol (50:48:2), and once with chloroform. To the aqueous phase, we added 24 μg of glycogen and 700 μl of isopropanol, then cooled it at −20°C for 20 minutes, followed by centrifugation at 16,000*g* for 20 minutes. The DNA pellet was rinsed twice with 1 ml of 70% ethanol, briefly dried, and suspended in water.

Extraction of Zebrafish RNA. Total RNA was isolated from single zebrafish embryos at 50%-epiboly stage as follows. An embryo whose chorion had been removed by forceps was extensively homogenized with a pestle in a microcentrifuge tube (Kontes) containing 100 μl of Ultraspec RNA solution (Biotecx). The sample was cooled on ice for 5 minutes, followed by the addition of 20 μl of chloroform. It was vigorously vortexed for 15 minutes and centrifuged at 13,800*g* for 15 minutes at 4°C. Then 140 μl of isopropanol was added to the top (aqueous) phase, and it was centrifuged at 13,800 *g* for 10 minutes at 4°C.

Polymerase Chain Reaction. Except where noted, polymerase chain reaction (PCR) was done in a volume of 10 μl containing 200 nM of each primer, 200 μM of dNTP, 1 U of thermostable DNA polymerase, and 1× buffer supplied with the polymerase on a GeneAmp 9700 (PerkinElmer).

Synthesis of* ntl *Complementary DNA. The total RNA isolated as described above was reverse transcribed by SuperScript II (Gibco) using oligo-dT as a primer to synthesize single-strand DNA according to the manufacturer's instruction. We amplified the *ntl* coding region using the single-strand DNA as a template with the primers GCGCTGTCAAAGCAACAGTAT (forward) and ATCAACCCGTTTTCTGATTGTCA (reverse). Cycling conditions were 3 minutes at 94°C, 30 × (15 seconds at 94°C, 30 seconds at 62°C, 2 minutes at 68°C), and 7 minutes at 68°C. To minimize PCR error, we used KOD DNA polymerase that exhibits high accuracy (Nishioka et al., 2001). The 1.5-kb PCR products were kinased, digested with *Xba*I, and subcloned into the pUC18 vector.

Detection of Mutations in* ntl *Complementary DNA. Twelve 120 *ntl* cDNA clones per embryo were sequenced on a LiCor 4200 sequencer. We also

analyzed the size of 566 clones of *ntl* cDNA by PCR. For each clone we amplified 6 different regions of *ntl* by using 6 primer sets such that we could scan the entire coding region without losing sensitivity to small deletions or insertions (> ±10 bp). For PCR we used *EX Taq* polymerase (TaKaRa) with cycling conditions of 25× (30 seconds at 94°C, 30 seconds at 60°C, 30 seconds at 72°C). The 6 primer sets used were as follows primer set 1, forward (GCGCTGTCAA AGCAACAGTAT); reverse (AAGTTGGG TGAGTCCGGG TGGATGTAG); primer set 2, forward (CGGTCTCGACC CTAATGCAAT); reverse (AATCCCACCGACTTTCACGA); primer set 3, forward (GCACAAATACGAACCCAGGA); reverse (CAGCCACCGAGTTGTGAA TA); primer set 4, forward (GTCCCAGACCACAGCACTG ACAACCAGG); reverse (GAGCAGCTCTGTGGTTCCTC); primer set 5, forward (CTTGAGGAACCACAGAGCTG CTC); reverse (AAC GCCAACCTCGCTTAGG); primer set 6, forward (ACATCT CAGTTCCTACGCGG); reverse (TCCT CCTGAAGCCAAG ATCAAGTCCA).

Sequences of* ntl *Introns. To determine the structures of 7 introns of *ntl*, we amplified 4 segments of *ntl* in the zebrafish genome with 4 primer sets. The primer sets 1, 2, and 5 described above were used to amplify intron1, intron2, and intron6-7, respectively. For intron3-5, we used the forward primer of primer set 3 and the reverse primer GAGCAGCTCTGTGGTTCCTC. For this experiment we used *EX Taq* polymerase with cycling conditions of 25 × (30 seconds at 94°C, 30 seconds at 60°C, 30 seconds at 72°C). The sequences of *ntl* introns are deposited in GenBank (accession number AB088068).

Identification of* ntl *Mutants at Bud Stage. Since the phenotypes of *ntl* homozygous mutants were not evident at bud stage, we identified *ntl* mutations from their DNA. We first extracted genomic DNA from single embryos and then PCR-amplified DNA segments carrying each mutation. For ntl^{b195} we amplified an *ntl* region containing a transposon-like sequence (Schulute-Merker et al., 1994) by 2 step PCR with *Ex Tag* polymerase: 28× (30 seconds at 94°C and 2 minutes at 68°C) with a forward primer, CGGTCTCGACCCTAATGCAAT, and a reverse primer, GTGAGTCCGGGTGGATGTAG. The PCR products were run on 1% agarose gels. The ntl^{b195} homozygous mutants could be easily identified by the lengths of the amplified DNA fragments; a fragment from mutation allele is approximately 1.5 kb longer than that from wild-type allele. For ntl^{tc41} an *ntl* region containing a base substitution (Odenthal et al., 1996) was amplified by PCR: 30× (30 seconds at 94°C, 30 seconds at 60°C, 30 seconds at 72°C) with a forward primer, CGGTCTCGACCC-TAATGCAAT, and a reverse primer, AATCCCACC-GACTTTCACGA. To identify ntl^{tc41} homozygous mutants, the amplified DNA fragment was digested with *Hpy* CH4 V (New England Biolabs), which cleaves wild-type allele, but not ntl^{tc41} allele. The digested samples were directly run on 2% agarose gels.

Ligation-Mediated PCR. Ligation-Mediated PCR (LMPCR) assays for the detection of DSEs in AT repeats were performed essentially as described (Schlissel et al., 1993). To make a linker, a long oligonucleotide, NB-1 (5′-CGTGGCGATGGAGATCAGACT TCCG-3′), and a short nucleotide, NB-2 (5′-CGGAAGT CTGA-3′), were annealed by mixing 2 nmol of each in a volume of 50 μl of 250 mM Tris (pH 8.0), heating the mixture to 90°C for 5 minutes, incubating it at 60°C for 5 minutes, and allowing the mixture to cool to room temperature. The sequences of the 2 oligonucleotides were slightly different from those used in the original protocol (Schlissel et al., 1993). Purified zebrafish genomic DNA (1–20 μg for adult fish and 1 μg for embryos) was ligated to the linker in a reaction volume of 50 μl containing 66 mM Tris (pH 7.5), 5 mM $MgCl_2$, 1 mM dithiothreitol, 1 mM ATP, 20 pM linker, and 400 U of T4 DNA ligase (NEB). Ligations were incubated at 16°C for 12 to 16 hours. The reactions were then inactivated at 65°C for 10 minutes.

A nested PCR strategy was used to identify the sites of linker ligation. Ligated DNA (40–800 ng) was used in a 25-μl PCR assay containing 25 ng of the linker primer, either NB-ATA (5′-ATGGAGATCA GACTTCCGATA-3′) or NB-TAT (5′-ATGGAGATCAGAC TTCCGTAT-3′), and 25 ng of each locus-specific primer, 10 mM Tris at pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, and 0.5 U of *Taq* polymerase (Roche). The sequences of locus-specific primers were as follows: upstream of *ntl*, 5′-CTTGAGGAACCACAGAGCTGCTC-3′; downstream of *nog2*, 5′-CCCGCAGTGTTCTTCAGTCT-3′. Samples were amplified for 25 cycles of 1 minutes at 94°C, 15 seconds at 60°C, and 1 minute at 72°C, followed by a final 10-minute extension step at 72°C. The reaction product (1 μl) was then used to program an additional 35 cycles of PCR in an identical buffer containing a second, nested, locus-specific primer and NB-ATA or NB-TAT. The sequences of nested, locus-specific primers were as follows: upstream of *ntl*, 5′-TATCCCAGCCATTACTCCCA-3′; downstream of *nog2*; 5′-ACCACGAGCTGAAGTGCTC-3′. NB-ATA and NB-TAT were identical to NB-1 except for 3 additional nucleotides, ATA and TAT, respectively, added at their 3′ ends for the enhancement of amplification specificity.

Table 1. Mutations in *no tail* cDNA Clones

Strain	*Embryo ID*	*Deletions*	*Insertions*	*N*[a]
AB	#1	2	1	69
	#2	0	0	68
AB/WIK hybrid	#3	5	0	66
	#4	2	0	67
WIK	#5	3	0	80
	#6	3	0	65
	#7	0	0	70
TL	#8	7	0	65
	#9	0	0	68
	#10	0	0	68
Total	10	22	1	686

[a]Total number of offspring checked for insertions and deletions by PCR or by sequencing (10 clones per embryo).

Results

Presence of Aberrant* ntl *Transcripts in Wild-type Embryos. We speculated that somatic mutation might spontaneously occur in the *ntl* gene in wild-type zebrafish. To examine the possibility we sequenced 120 *ntl* cDNA clones derived from 4 strains of the zebrafish. We identified 4 mutations in 4 clones: 2 clones had a single-base substitution, and the other 2 carried the same deletion of 41 bp (unpublished results). Because the length of cDNA was 1.4 kb, the frequency of point mutations was only 1.2×10^{-5} per base pair. This frequency of point mutations approximated the error frequency of the thermostable DNA polymerase used for the amplification of reverse-transcribed *ntl* mRNA (Nishioka et al., 2001). We thus concluded that point mutations were not introduced in *ntl* transcripts at a detectable level.

The frequency of deletions that occurred in 2 (1.7%) of 120 clones seemed to be significantly higher because frameshift mutations happen much less frequently than point mutations during PCR (Eckert and Kunkel, 1991). To investigate if deletions could be a type of alternation that occurs in *ntl* transcripts, we analyzed the sizes of an additional 566 *ntl* cDNAs by PCR. We found 20 new deletions and an insertion in the *ntl* cDNAs: in total, 23 mutations (22 deletions and one insertion) were found in 686 *ntl* cDNAs (3.4%) (Table 1). Schematic representation of these mutations is shown in Figure 2 (A).

We observed 4 characteristics in the mutations. First, deletions were found in cDNA clones from all 4 strains of zebrafish, with a frequency that was not significantly different between the strains: the lowest were observed in clones from AB strain (2.2%) and the highest in AB/WIK hybrid strain (5.3%). Second, the mutation rate varied from embryo to embryo: in the case of one embryo (#8 in Figure 2, A), 7 in 65 *ntl* cDNA clones (11%) carried the same deletion, whereas no deletions or insertions were found in 4 of 10 embryos (Table 1). Among the 22 deletions found, at least 12 could be regarded as independent because they had different deletion sites. Among the 12 deletions, 2 had the same deletion site (mutations "e" and "f" in Figure 2, A). They were regarded as independent deletions because they were derived from different embryos: the lowest frequency of deletion was estimated to be 1.7% (12 of 686 clones). Of the 12 independent deletion sites found, 6 were located between nucleotide positions 740 and 788, suggesting a possible deletion hotspot (Figure 2, A). The other deletions seemed to be located at random.

Third, most of the deletions (9 of 12) changed the reading frame of *ntl*. Fourth, all the deleted sequences in those clones had a short direct repeat composed of between 2 and 6 bp at the borders of the rearrangement (Figure 2, A and C). This was also the case for an EST clone with a deletion, deposited in GenBank (Figure 2, A) (Kudoh et al., 2001). The one insertion identified had the same repeat feature as the deletions: a 3 bp-direct repeat at the border of the rearrangement (Figure 2, A).

These results indicate that the aberrant transcripts were generated by a common mechanism. With regard to the mechanism, it is interesting to note that the insertion turned out to be the entire region of intron7, and 3 deletions in Figure 2 (A) were apparently caused by missplicing (Figure 2, D). Moreover, the deletion found in a TL embryo (#8) had GT and AG dinucleotides at the left and right borders, respectively, as was the case for most introns. These results suggested that missplicing was prone to happen in pre-mRNA of *ntl*.

Assessment of Artificial Generation of Aberrant* ntl *cDNA. During cDNA synthesis artificial generation of deletions has been reported between 2 direct repeats adjacent to a strong stem-loop secondary structure (Mader et al., 2001). Although no obvious sequences that potentially adopt stem-loop structures at deletion points in *ntl* cDNAs were predicted by a program that detects RNA secondary structures (Zuker, 2003), we assessed the possibility of artificial generation of aberrant *ntl* cDNA clones as described (Mader et al., 2001): in vitro–synthesized *ntl* mRNA was reverse transcribed, amplified, and cloned in *Escherichia coli*. No deletions or insertions were, however, detected in 480 cDNA clones analyzed (data not shown). This suggested that *ntl* mRNA did not adopt a secondary structure that elicited the artificial generation of aberrant cDNA. However, it could still be argued that mutation was prone to occur during reverse transcription of mRNA extracted by our procedure. We thus divided total RNA

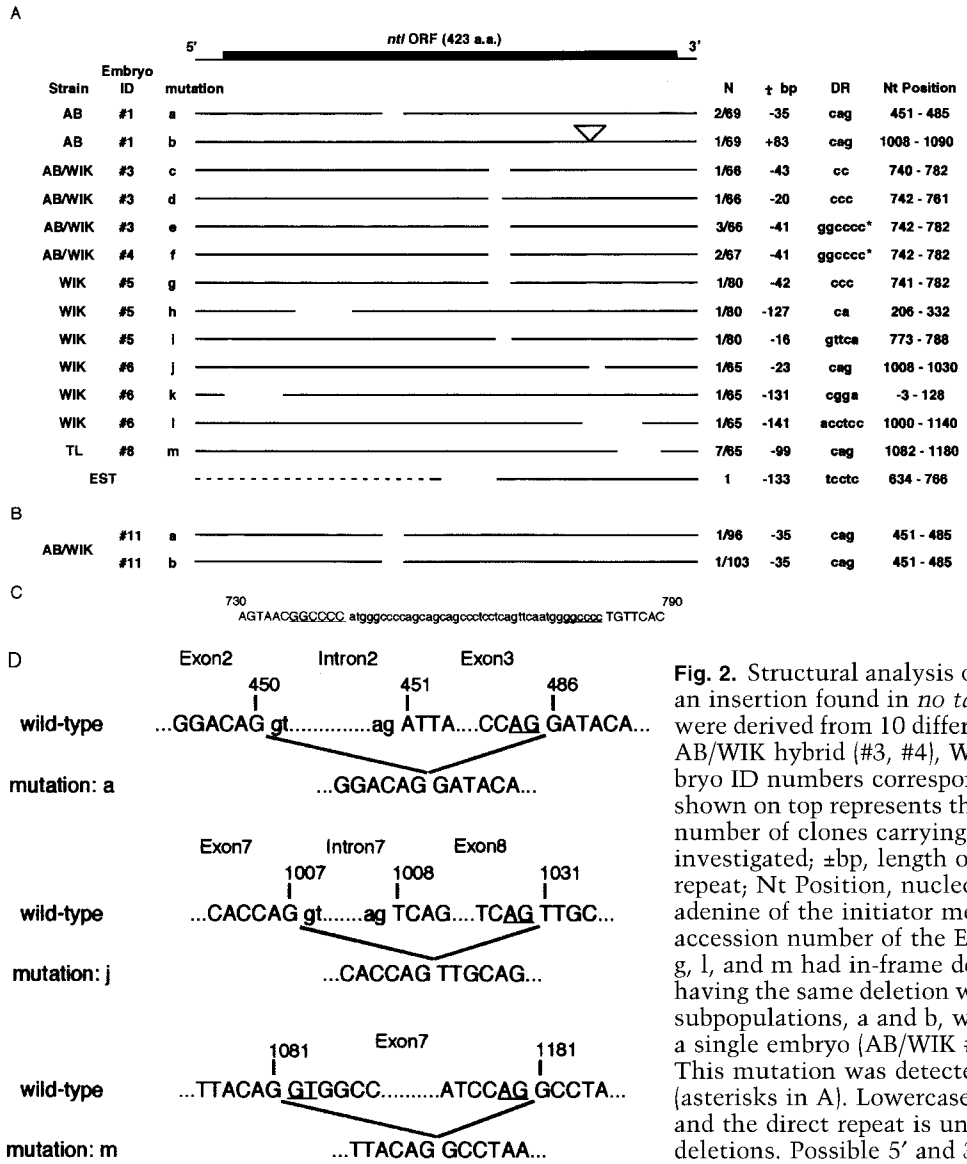

Fig. 2. Structural analysis of *ntl* transcripts. **A**: Deletions and an insertion found in *no tail* cDNAs. The cDNAs examined were derived from 10 different embryos of strains AB (#1, #2), AB/WIK hybrid (#3, #4), WIK (#5–#7), and TL (#8–#10). Embryo ID numbers corresponds to those in Table 1. The bar shown on top represents the coding region of *ntl*. N indicates number of clones carrying mutation/total number of clones investigated; ±bp, length of deletion or insertion; DR, direct repeat; Nt Position, nucleotide position of mutation. The adenine of the initiator methionine is numbered as 1. The accession number of the EST clone is BG985630. Mutations g, l, and m had in-frame deletions. **B**: Two *ntl* cDNA clones having the same deletion were isolated from 2 of the 3 cDNA subpopulations, a and b, which were originally derived from a single embryo (AB/WIK #11). **C**: An example of a deletion. This mutation was detected in 2 independent embryos (asterisks in A). Lowercase letters show deleted nucleotides, and the direct repeat is underlined. **D**: Missplicing causes deletions. Possible 5′ and 3′ splice sites are underlined.

isolated from a single embryo (#11 in Figure 2, B) into 3 subpopulations (a, b, and c) and generated cDNA from these subpopulations.

We examined the sizes of approximately 100 cDNA clones from each cDNA subpopulation and found 7 cDNAs with deletions. Among the 7 clones, 5 had independent deletions. However, the other 2 had an identical deletion, and each was derived from 2 different cDNA subpopulations (Figure 2, B). Considering that the mutation rate fluctuated between embryos, and that the deletions seemed to occur at random positions in *ntl* cDNAs, it was unlikely that the identical deletion found in the 2 cDNA subpopulations was independently generated in vitro. Furthermore, as described below, we were unable to find any deletions or insertions in cDNA clones of another gene, *noggin2* (*nog2*) (Fürthauer et al., 1999), which were amplified from the same reverse-transcribed RNA as the one used for *ntl* cDNA. On the basis of these results, we concluded that aberrant *ntl* mRNA was already present in the total RNA used for reverse transcription.

To assess if translation of altered *ntl* mRNA occurs, we then examined if altered *ntl* mRNA could be recovered from polysomes. RNA that was associated with polysomes was isolated as described (Krichevsky et al., 1999) and used for the generation of cDNA. The presence of insertions or deletions was examined for 672 clones, but no mutations could be found. This result suggests the presence of a cellular system that eliminates aberrant mRNA before it is translated into defective protein.

Absence of Mutations in Genomic* ntl *DNA. The presence of defective *ntl* transcripts in zebrafish embryos suggested that they originated from somatic mutations that occurred in genomic *ntl* DNA. We thus extracted genomic DNA from the same embryo (AB/WIK#4) from which *ntl* transcripts with a deletion had been isolated and amplified an *ntl* region, which flanked the deletion site, by PCR (Figure 3, B). However, no mutations could be found in the 355 clones we examined (data not shown).

We then extracted DNA from the most caudal part of a tailfin-less fish and checked the size of the *ntl* DNA by PCR. We expected that there would be deletions or insertions in the DNA if somatic mutation in *ntl* caused the phenotype as we suspected. Although we checked 500 *ntl* clones, no mutation was found. These results suggests that somatic mutation did not occur in *ntl* DNA.

***Formation of DSEs in* ntl.** While searching for mutations in *ntl* DNA, we also tried to find intrinsic factors that may destabilize the structure of the gene. Candidates were repetitive DNA sequences such as transposon and inverted/direct repeats, which are known to cause DNA rearrangements (Lupski and Stankiewicz, 2002). Because no distinct repeats were found in the coding region of *ntl*, we amplified all 7 introns of *ntl* (Schulte-Merker et al., 1994) by PCR and determined their sequences (GenBank accession number AB088068). We found one long microsatellite repeat, $(AT)_{41}$, in intron6 (Figure 3, A).

An AT repeat has been shown to form a cruciform structure in vivo (Haniford and Pulleyblank, 1985), a possible barrier to replication (Hyrien, 2000). Furthermore, it has been proposed that one or two DNA DSEs are formed where a replication fork encounters such a barrier (Figure 4) (Fujimura and Horiuchi, 1995; Connelly and Leach, 1996; Seigneur et al., 1998). Since DSE formation is likely required for somatic hypermutation of immunoglobulin genes (Papavasiliou and Schatz, 2000; Jacobs et al., 2001), by using LMPCR (Schlissel et al., 1993), we examined if DSEs were formed at the AT repeat in the *ntl* gene as a consequence of replication fork arrest. If genomic DNA contains a DSE in the AT repeat, a linker can be ligated to the cleaved end, and subsequent PCR with a primer pair, one on the linker and the other on *ntl*, can amplify *ntl* DNA fragments. With the *ntl*-specific primer that hybridizes upstream of the AT repeat (Figure 3, A), we could amplify fragments that hybridized an *ntl*-specific oligonucleotide probe from genomic DNA of zebrafish embryos (Figure 3, B and C). No specific PCR product was, however, obtained with the *ntl* primer hybridizing downstream of the AT repeat. Likewise, we tried to recover DSE sites from adult fish, and the amplified fragments were cloned with 10 clones being sequenced (Figure 3, D). We confirmed that, in all 10 clones, the linker was joined just inside the 5′ border of the AT repeat (Figure 3, D). This result indicates that DSEs were actually formed at the AT repeat in *ntl*.

DSE Formation in Another AT Repeat. To assess if DSE formation at the AT repeat was restricted to the *ntl* AT repeat, we further surveyed another 13 AT repeats in zebrafish (Table 2) and found another DSE in an AT repeat in *noggin2* (*nog2*) (Figure 3, C). The AT repeat was located in the 3′ untranslated region of *nog2*, and only a *nog2*-specific primer hybridizing downstream of the repeat could amplify specific PCR products. As was the case for DSEs in *ntl*, the other end of the AT repeat in *nog2* could not be recovered. The DSEs in *ntl* and *nog2* were considered to be blunt and phosphorylated, because we used a blunt, unphosphorylated linker for LMPCR. Treatment of genomic DNA of adult fish with T4 DNA polymerase (to blunt ends with 5′ or 3′ overhangs) or mung bean nuclease (to open hairpin ends) before linker ligation did not enhance the frequency of detection of DSEs in *ntl* or *nog2* (data not shown). This may imply that most DSEs at these AT repeats have blunt ends.

To examine if DSE formation in *nog2* led to somatic mutations in the transcripts, we analyzed the size of *nog2* cDNA clones by PCR. Although we checked 480 clones, no deletions or insertions could be found.

Possible Link Between DSE and Replication. For the LMPCR described above, we used genomic DNA extracted from tailbud-stage embryos and 1-year-old wild-type adult fish. The experiments revealed that DSB formation in either *ntl* or *nog2* occurred at a much higher frequency in embryos than in adult fish. Based on the quantity of the genomic DNA used for LMPCR, the frequency of the DSE formation in embryos was 1 or 2 orders of magnitude higher than that in adults (Table 3). These results were consistent with the idea that the DSE formation was coupled to replication.

We were unable to detect *ntl* DSE in genomic DNA from a zebrafish cell line, AB9 (data not shown). Because the cell cycle of the cells we used was not synchronized, the number of replicating cells, and thus *ntl* DSEs, might not have been high enough to be detected in the sample. Alternatively, the cells might not express a gene or genes required for DSE formation in *ntl*.

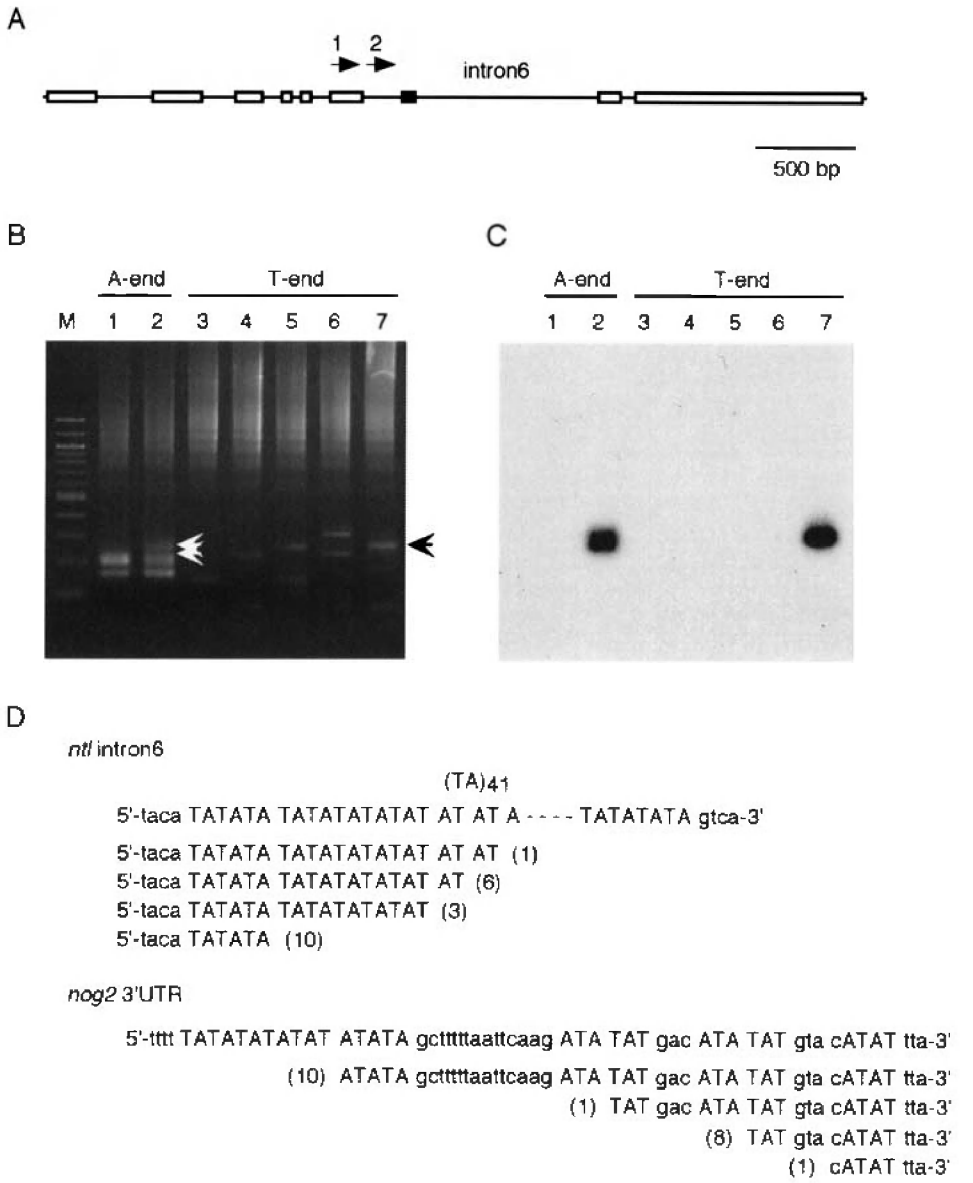

Fig. 3. Detection of DNA double-strand breaks in *ntl* using LMPCR. **A**: Genomic structure of *ntl*. White boxes and a black box show exons and a $(AT)_{41}$ dinucleotide repeat, respectively. Arrows show relative positions of the 2 locus-specific primers used for LMPCR. Primers 1 and 2 represent primers used for the first and second (nested) rounds of PCR, respectively. **B**: LMPCR products of upstream break sites from zebrafish embryos were separated on a 2% (w/v) agarose gel and visualized with ethidium bromide. We slightly modified the primer sequence used in the original LMPCR method so that DSEs ending in adenine and thymine could be distinguished. LMPCR products, which were expected to have DSEs ending in adenine and thymine, were placed on lanes A-end and T-end, respectively. M indicates 100-bp ladder; lanes 1–7, each lane contains a separate aliquot of the same DNA mixture from 5 zebrafish embryos. Two white arrowheads and a black arrow indicate the PCR products that gave the positive signals in C. **C**: Southern blotting with an *ntl*-specific probe reveals the existence of DSEs in the *ntl*. The signal in lane 2 turned out to be derived from the 2 distinct PCR products shown with white arrowheads in B. **D**: Location of DSEs in the *ntl* and *nog2* genes. DSE sites that end in either adenine or thymine were separately amplified as in B. Ten positive clones for each end were sequenced. The number of clones identified in the 10 clones is shown in parentheses. A clone obtained from *nog2* with a "T-end" primer ended in cytosine.

We then examined if transcription would influence the DSE formation, as the generation of DSEs in immunoglobulin genes has been shown to depend on their transcriptional activity (Papavasiliou and Schatz, 2000; Jacobs et al., 2001). For this purpose we employed 2 zebrafish mutations, ntl^{b195} and ntl^{tc41}, in both of which *ntl* expression is severely reduced at tailbud stage because the *ntl* product is an activator of its own gene (Schulte-Merker et al., 1994; Odenthal et al., 1996). We extracted genomic DNA from these homozygous embryos at tailbud stage and performed LMPCR. We found that the frequency of

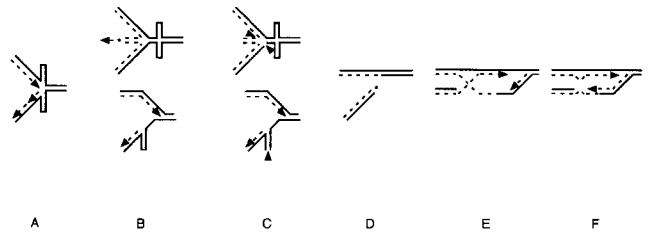

Fig. 4. DSE formation at an AT repeat, based on models presented by Connelly and Leach (1996) and Seigneur et al. (1998). **A**: A replication fork stall at an AT-cruciform. **B**: A Holliday structure with a DSE arises by replication fork reversal (top) (Seigneur et al., 1998), or a hairpin structure is formed on a lagging strand, which inhibits the extension or synthesis of Okazaki fragments (bottom) (Connelly and Leach, 1996). **C**: Endonucleolytic cleavage of these structures (arrowheads). **D**: Generation of a sister chromatid with a DSE. **E**: The broken replication fork is likely restored by homologous recombination. **F**: The resultant Holliday structure will be cleaved by resolvases. Dotted lines show nascent DNA strands. Asterisks indicate DSEs to which a LMPCR linker can be ligated.

DSE formation was not significantly different between these mutants and wild-type (Table 3). Thus transcription may not influence DSE formation at the *ntl* AT repeat. The same experiment could not be performed for the *nog2* AT repeat because no mutants affecting the expression of *nog2* have been isolated.

Discussion

Aberrant Transcription of* ntl *in Wild-type Zebrafish. We formed a hypothesis, based on the phenotypic similarity between *Brachyury* mutant mouse and tailfin-less zebrafish, that *ntl* is genetically instable. As shown above, we actually found aberrant *ntl* transcripts in wild-type zebrafish embryos. Although we did not have conclusive data showing the relation of the generation of aberrant *ntl* transcripts to the tailfin phenotype, two characteristics of the aberrant transcripts implied that the activity of *ntl* could be spontaneously reduced, to various degrees, in a subpopulation of zebrafish embryos. First, the chance of the appearance of aberrant *ntl* mRNA was different between embryos (Table 1). In an embryo aberrant *ntl* transcripts occupied more than 10% of total *ntl* transcripts (Table 1). Second, there seemed to be no useful biological function for the products of the aberrant *ntl* transcripts, since most of the alternations changed the reading frame of *ntl* (Figure 2).

Unlike *Brachyury* heterozygous mice, no phenotypes are observed in *ntl* heterozygotes. However, it might be possible that the tailfin phenotype appears when aberrant *ntl* transcripts occupy more than half of total *ntl* transcripts in the cells that participate in the formation of the tail. To assess if the tail-less phenotype would be caused by a significant increase in aberrant *ntl* mRNA, it will be necessary to establish a system that allows monitoring of the level of *ntl* activity in vivo.

All of the alternations in *ntl* cDNAs we found give rise to termination codons upstream of the normal stop codon; Therefore, translation could yield C-terminal truncated *ntl* proteins. Such polypeptides can often act in a dominant-negative manner, leading to deleterious effects on the cell or organism. In fact, *Brachyury* mutations that yield C-terminal truncated forms of the protein cause more severe phenotypes than those of null mutations (Herrmann et al., 1990). However, it may be less likely that aberrant *ntl* transcripts act in a dominant-negative fashion as they were not recovered from polysome fraction. Presumably these transcripts became targets of nonsense-mediated mRNA decay (NMD), a surveillance mechanism that selectively degrades nonsense mRNA (Singh and Lykke-Andersen, 2003). This system is conserved from yeast to

Table 2. DSE Formation at Zebrafish AT Repeats

Locus	*Position*	$(AT)_n$	*DSE*	*Accession no.*
stat3	Intron	84	–	AF322857
hoxalla	Intron	56	–	AF071240
no tail	Intron	41	+	S57147
myf-5	Intergenic region	41	–	AF184166
MHC Brre-DBB	Intron	33	–	BRU08869
Odorant receptor gene cluster	Intergenic region	32	–	AF112374
hagoromo	Intergenic region	31	–	AB037997
RP71-3C13 (BAC clone)		30	–	AL590150
fibronectin	Intron	21	–	AF342953
fibroblast growth factor 3	3′ UTR	11	–	AF342953
noggin 2	3′ UTR	9	+	AF159148
jak2a	3′ UTR	8	–	AJ005690
activin receptor-like kinase 8	3′ UTR	5	–	AF038425
bone morphogenetic protein 4	Intron	5	–	AF056336

Table 3. Frequencies of DSE Formation at AT Repeats

Cross	*Genotype*	*Stage*	*Haploid genomes/DSB*[a]	*Position of AT repeats*
+/+ × +/+				
	+/+	Adult	1.1×10^6	*ntl*
	+/+	Tailbud	1.1×10^4	*ntl*
	+/+	Adult	0.9×10^6	*nog2*
	+/+	Tailbud	4.7×10^4	*nog2*
ntl^{b195}/+ × ntl^{b195}/+				
	+/+	Tailbud	3.5×10^4	*ntl*
	ntl^{b195}/ntl^{b195}	Tailbud	6.6×10^4	*ntl*
ntl^{tc41}/+ × ntl^{tc41}/+				
	+/+	Tailbud	1.2×10^5	*ntl*
	ntl^{tc41}/ntl^{tc41}	Tailbud	5.9×10^4	*ntl*

[a]The average numbers of zebrafish haploid genomes required for the detection of a DSE by LMPCR. In most cases we used 100 ng of genomic DNA per LMPCR, which corresponds to 5.9×10^4 haploid genomes given that the molecular weight of the haploid genome is 1.7 pg.

human, and we found that the zebrafish genome contains homologues of the genes required for NMD, such as up-frameshift (Upf) proteins (data not shown). Furthermore, a possible occurrence of NMD in the zebrafish has been reported for a mutant in which the expression of the responsible gene was barely detectable (Parsons et al., 2002).

How aberrant *ntl* transcripts were generated remains to be determined. Since no mutation could be recovered from *ntl* DNA, from all our examinations, they must have been generated by a mechanism that does not accompany the alternation of the *ntl* DNA structure. On the basis of the data shown in Figure 2 (D), we speculated that missplicing might be involved in the mechanism, although how missplicing of *ntl* transcript happens remains unknown.

DSE Formation at AT Repeats. In addition to the generation of aberrant *ntl* transcripts, we found that another genetic event happened in *ntl*: DSE formation at an AT repeat, which was also observed in another gene, *nog2*. Two characteristics of the DSE formation in *ntl* and *nog2* suggested that it was related to replication. First, we detected more DSEs in rapidly growing zebrafish embryos than in adults. Second, only either side of the breakpoints was recovered by LMPCR. This result can be expected if the DSEs were formed as proposed by Connelly and Leach (1996), or by Seigneur et al. (1998) (Figure 4). In their models, one (Connelly and Leach, 1996) or two (Seigneur et al., 1998) free DNA ends to which a linker for LMPCR can ligate will be generated at a replication fork upon encountering a barrier by the action of endonucleases of SbcCD and RuvC, respectively (Figure 4). One site is the free end of a broken sister chromatid. The other site is the end of a protruding DNA formed by annealing of two nascent strands (a branch of a Holliday structure). Although it remains to be determined which DSE site we detected by LMPCR, in either case, PCR products will be derived from only one side of template DNA, which is consistent with our results. We thus consider that certain AT repeats in the zebrafish genome halt the progression of a replication fork, perhaps by adopting hairpin or cruciform structures, and thereby forming DSEs (Figure 4).

The frequencies of DSE formation in *ntl* and in *nog2* were considerably low: at most, one DSE was found in 5.9×10^4 and 4.7×10^4 haploid genomes, respectively. This may reflect infrequent, sporadic occurrence of DNA breaks at the AT repeats. In this case there should not be any consequences for the breaks. Alternatively, it may be that DSEs are formed at much higher frequencies during replication than the frequencies above, but are repaired so rapidly that only a fraction of them can be detected by LMPCR. In this case it might be possible that DSEs are formed to exert unknown functions. We need methods other than LMPCR to determine how frequently a DSE is formed at the AT repeats per replication.

At present it is unclear if DSE formation is directly related to the generation of aberrant *ntl* transcripts; unlike somatic mutation in immunoglobulin genes, no correlation was observed between the positions of the DSEs and the mutations in *ntl* transcripts. Furthermore, we did not observe any mutations in *nog2* transcripts, regardless of the occurrence of DSE formation in the gene. Isolation of zebrafish mutations that change the frequency of these 2 incidents will help to clarify their relation and biological significance, if any, and involvement in the tailfin phenotype.

Acknowledgments

We thank Dr. H. Siomi for providing facilities throughout the course of this study; F. Unemi, T. Nagami and R. Makimoto for fish maintenance; H.

Takeda and M. Seiki-Furutani for the *ntl* mutants; and the members of H. Siomi lab for technical assistance and advice.

References

1. Connelly JC, Leach DRF (1996) The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. Genes Cells 1, 285–291
2. Eckert KA, Kunkel TA (1991) *PCR: A Practical Approach.* (New York, NY: Oxford University Press) pp 225–244
3. Fujimura Y, Horiuchi T (1995) Recombinational rescue of the stalled DNA replication fork: a model based on analysis of an *Escherichia coli* strain with a chromosome region difficult to replicate. J Bacterial 177, 783–791
4. Fürthauer M, Thisse B, Thisse C (1999) Three different *noggin* genes antagonize the activity of bone morphogenetic proteins in the zbrafish embryo. Dev Biol 214, 181–196
5. Halpern ME, Ho RK, Walker C, Kimmel CB (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish *no tail* mutation. Cell 75, 99–111
6. Haniford DB, Pulleyblank DE (1985) Transition of a cloned d(AT)n-d(AT)n tract to a cruciform in vivo. Nucleic Acids Res 13, 4343–4363
7. Haworth K, Putt W, Cattanach B, Breen M, Binns M, Lingaas F, Edwards YH (2001) Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. Mamm Genome 12, 212–218
8. Herrmann BG, Kispert A (1993) The *Brachyury* gene encodes a novel DNA binding protein. EMBO J 12, 3211–3220
9. Herrmann BG, Kispert A (1994) Immunohistochemical analysis of the *Brachyury* protein in wild-type and mutant mouse embryos. Dev Biol 161, 179–193
10. Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H (1990) Cloning of the *T* gene required in mesoderm formation in the mouse. Nature 343, 617–622
11. Hyrien O (2000) Mechanisms and consequences of replication fork arrest. Biochimie 82, 5–17
12. Jacobs H, Rajewsky K, Fukita Y, Bross L (2001) Indirect and direct evidence for DNA double-strand breaks in hypermutating immunoglobulin genes. Philos Trans R Soc Land B Biol Sci 356, 119–125
13. Krichevsky AM, Metzer E, Rosen H (1999) Translational control of specific genes during differentiation of HL-60 cells. J Biol Chem 274, 14295–14305
14. Kudoh T, Tsang M, Hukriede NA, Chen X, Dedekian M, Clarke CJ, Kiang A, Schultz S, Epstein JA, Toyama R, Dawid IB (2001) A gene expression screen in zebrafish embryogenesis. Genome Res 11, 1979–1987
15. Lupski JR, Stankiewicz P (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18, 74–82
16. MacMurray A, Shin HS (1988) The antimorphic nature of the T^c allele at the mouse *T* locus. Genetics 120, 545–550
17. Mader RM, Schmidt WM, Sedivy R, Rizovski B, Braun J, Kalipciyan M, Exner M, Steger GG, Mueller MW (2001) Reverse transcriptase template switching during reverse transcriptase—polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. J Lab Clin Med 137, 422–428
18. Nishioka M, Mizuguchi H, Fujiwara S, Komatsubara S, Kitabayashi M, Uemura H, Takagi M, Imanaka T (2001) Long and accurate PCR with a mixture of KOD DNA polymerase and its exonuclease deficient mutant enzyme. J Biotechnol 88, 141–149
19. Odenthal J, Haffter P, Vogelsang E, Brand M, van Eeden FJM, Furutani-Seiki M, Granato M, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Warga RM, Allende ML, Weinberg ES, Nüsslein-Volhard C (1996) Mutations affecting the formation of the notochord in the zebrafish, *Danio rerio.* Development 123, 103–115
20. Papavasiliou FN, Schatz DG (2000) Cell-cycle-regulated DNA double-strand breaks in somatic hypermutation of immunoglobulin genes. Nature 408, 216–221
21. Parsons MJ, Polland SM, Saúde L, Feldman B, Coutinho P, Hirst EMA, Stemple DL (2002) Zebrafish mutants identify an essential role for laminins in notochord formation. Development 129, 3137–3146
22. Schlissel M, Constantinescu A, Morrow T, Baxter M, Peng A (1993) Double-strand signal sequence breaks in *V(D)J* recombination are blunt, 5′-phosphorylated, RAG-dependent, and cell cycle regulated. Genes Dev 7, 2520–2532
23. Schulte-Merker S, Ho RK, Herrmann EG, Nüsslein-Volhard C (1992) The protein product of the zebrafish homologue of the mouse *T* gene is expressed in nuclei of the germ ring and the notochord of the early embryo. Development 116, 1021–1032
24. Schulte-Merker S, van Eeden FJM, Halpern ME, Kimmel CB, Nüsslein-Volhard C (1994) *no tail* (*ntl*) is the zebrafish homologue of the mouse *T* (*Brachyury*) gene. Development 120, 1009–1015
25. Seigneur M, Bidnenko V, Ehrlich SD, Michel B (1998) RuvAB acts at arrested replication forks. Cell 95, 419–430
26. Singh G, Lykke-Anderson J (2003) "New insights into the formation of active nonsense-mediated decay complexes". Trends Biochem Sci 28: 464—466
27. Stemple DL, Solnica-Krezel L, Zwartkruis F, Neuhauss SCF, Schier AF, Malicki J, Stainier DYR, Abdelilah S, Rangini Z, Mountcastle-Shah E, Driever W (1996) Mutations affecting development of the notochord in zebrafish. Development 123, 117–128
28. Yanagisawa KO (1990) Does the *T* gene determine the anteroposterior axis of a mouse embryo? Jpn J Genet 65, 287–297
29. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31, 3406–3415