Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility

Mehdi Goudarzi¹, Kathryn Berg¹, Lindsey M. Pieper¹, and Alexander F. Schier¹,²,³,⁴,⁵

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA.
²Center for Brain Science, Harvard University, Cambridge, MA 02138, USA., ³FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA., ⁴Allen Discovery Center for Cell Lineage Tracing, University of Washington, Seattle, WA 98195, USA., ⁵Biozentrum, University of Basel, Switzerland.

Correspondence should be addressed to M.G. (mgoudarzi@fas.harvard.edu) or A.F.S. (schier@fas.harvard.edu).

Hundreds of long non-coding RNAs (IncRNAs) have been identified as potential regulators of gene expression, but their functions remain largely unknown. To study the role of IncRNAs during vertebrate development, we selected 25 zebrafish IncRNAs based on their conservation, expression profile or proximity to developmental regulators, and used CRISPR-Cas9 to generate 32 deletion alleles. We observed altered transcription of neighboring genes in some mutants, but none of the IncRNAs were required for embryogenesis, viability or fertility. Even RNAs with previously proposed non-coding functions (cyrano and squint) and other conserved IncRNAs (gas5 and Inc-setd1ba) were dispensable. In one case (Inc-phox2bb), absence of putative DNA regulatory-elements, but not of the IncRNA transcript itself, resulted in abnormal development. LncRNAs might have redundant, subtle, or context-dependent roles, but extrapolation from our results suggests that the majority of individual zebrafish IncRNAs have no overt roles in embryogenesis, viability and fertility.
Introduction

Long non-coding RNAs (lncRNAs) comprise a heterogeneous group of transcripts longer than 200 nucleotides that do not encode proteins. LncRNAs have been proposed to affect the expression of neighboring or distant genes by acting as signaling, guiding, sequestering or scaffolding molecules\textsuperscript{1-5}. The functions of specific lncRNAs in dosage compensation (\textit{xist}\textsuperscript{6,7}, \textit{tsix}\textsuperscript{8}, \textit{jpx}\textsuperscript{9}) and imprinting (\textit{Airn}\textsuperscript{10,11}, \textit{MEG3}\textsuperscript{12,13}, \textit{H19}\textsuperscript{14,15}) are well established, and mutant studies in mouse have suggested that \textit{fendrr}, \textit{peril}, \textit{mdget}, \textit{linc-brn1b}, \textit{linc-pint}\textsuperscript{16}, and \textit{upperhand}\textsuperscript{17} are essential for normal development. However, other studies have questioned the developmental relevance of several mouse lncRNAs, including \textit{Hotair}\textsuperscript{18}, \textit{MIAT/Gumafu}\textsuperscript{19}, \textit{Evx1-as}\textsuperscript{20}, \textit{upperhand}, \textit{braveheart} and \textit{haunt}\textsuperscript{21}. In zebrafish, morpholinos targeting the evolutionarily-conserved lncRNAs \textit{megamind} (TUNA\textsuperscript{22}) and \textit{cyrano} resulted in embryonic defects\textsuperscript{23}. However, a mutant study found no function for \textit{megamind} and revealed that a \textit{megamind} morpholino induced non-specific defects\textsuperscript{24}. These conflicting results have led to a controversy about the importance of lncRNAs for vertebrate development\textsuperscript{16,21}. We therefore decided to mutate a group of selected zebrafish lncRNAs using CRISPR-Cas9, and assay their roles in embryogenesis, viability and fertility.

Transcriptomic studies of early embryonic development\textsuperscript{23,25} and five adult tissues\textsuperscript{26} have identified over 2,000 lncRNAs in zebrafish\textsuperscript{27}, of which 727 have been confirmed as non-coding based on ribosome occupancy patterns\textsuperscript{28}. For our mutant analysis we selected 24 bona fide lncRNAs based on synteny (conserved relative position on at least one other vertebrate genome), sequence conservation, expression dynamics (expression
levels, onset and pattern) and proximity to developmental regulatory genes (see Table 1). These criteria were chosen to increase the likelihood of potential functional requirements of the selected lncRNAs. In addition, we selected a protein-coding RNA with a proposed non-coding function (*squint*).

**Results and discussion**

The genomic location of selected lncRNAs are depicted in Fig 1. The neighbor-relationship, and expression levels of the selected lncRNAs and their neighboring genes are shown in Fig 1- S1, Fig 1- S2, respectively.
Figure 1: Genomic location of selected lncRNAs

The chromosomal positions of selected lncRNAs are depicted. lncRNAs discussed in the text are underlined. The corresponding genomic coordinates for all lncRNAs are provided in the supplementary file 2.

Using CRISPR-Cas9 (Fig 1-S3) we generated 32 knockout-alleles. 24 alleles removed regions containing transcription start sites (TSS-deletion; 244bp to 736bp), and 8 alleles fully or partially removed the gene (1kb to 203kb) (Table 1). qRT-PCR analysis demonstrated effective reduction in the levels of the targeted lncRNA transcripts (average reduction of 94 ± 6%; Table 1), which was further tested and confirmed for a subset of lncRNAs by in situ RNA hybridization (Fig 2B, 3B, 3C, 4D, 5B and 6D).

Previous observations in mammalian cell culture systems suggested that lncRNA promoters can affect the expression of nearby genes. To test if these results hold true in vivo, we measured the changes in the expression of neighboring genes (a 200 kb window centered on each lncRNA) in lncRNA mutants. Several mutants displayed changes in the expression of neighboring genes (Fig 1-S4). In particular, 10 out of 40 neighboring genes showed more than two-fold changes in expression, lending in vivo support to observations in cell culture systems.

To determine the developmental roles of our selected lncRNAs, we generated maternal-zygotic mutant embryos (lacking both maternal and zygotic lncRNA activity) and analyzed morphology from gastrulation to larval stages, when all major organs have formed. Previous large-scale screens have shown that the visual assessment of live
embryos and larvae is a powerful and efficient approach to identify mutant phenotypes, ranging from gastrulation movements and axis formation to the formation of brain, spinal cord, floor plate, notochord, somites, eyes, ears, heart, blood, pigmentation, vessels, kidney, pharyngeal arches, head skeleton, liver, and gut. No notable abnormalities were detected in 31/32 mutants. Moreover, these 31 mutants survived to adulthood, indicating functional organ physiology, and were fertile (Table 1). In the following section we describe the results for five specific lncRNAs and put them in the context of previous studies.

<table>
<thead>
<tr>
<th>lncRNA mutant, deletion type</th>
<th>lncRNA Transcript ID</th>
<th>Deletion size</th>
<th>Percent reduction</th>
<th>Embryonic phenotype</th>
<th>Viability and fertility</th>
<th>Neighboring genes Up 100Kb</th>
<th>Neighboring genes Down 100Kb</th>
<th>Selection Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>cyrano\textsuperscript{a173}, TSS-del.</td>
<td>ENSDART00000139872</td>
<td>326 bp</td>
<td>98%</td>
<td>No</td>
<td>Yes</td>
<td>trim39b</td>
<td>op3</td>
<td>Syntenic and sequence conservation, Reported phenotype</td>
</tr>
<tr>
<td>cyrano\textsuperscript{a173}, gene del.</td>
<td>ENSDART00000139872</td>
<td>4374 bp</td>
<td>94%</td>
<td>No</td>
<td>Yes</td>
<td>trim39b</td>
<td>op3</td>
<td>Syntenic and sequence conservation, Reported phenotype</td>
</tr>
<tr>
<td>gas5\textsuperscript{e173}, TSS-del.</td>
<td>ENSDART00000156268</td>
<td>296 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
<td>oshp79</td>
<td>tor3a</td>
<td>Syntenic and sequence conservation, Well studied lncRNA, host of several snoRNA</td>
</tr>
<tr>
<td>lnc-setd1ba\textsuperscript{a274}, gene del.</td>
<td>ENSDART00000141500</td>
<td>3137 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
<td>setd1ba</td>
<td>rhoF</td>
<td>Syntenic and sequence conservation, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>squint\textsuperscript{a173}, gene del.</td>
<td>ENSDART00000079692</td>
<td>1032 bp</td>
<td>95%</td>
<td>No</td>
<td>Yes</td>
<td>hr1ab</td>
<td>eif4ebp1</td>
<td>Evolutionary conservation, Reported phenotype, Putative ncRNA</td>
</tr>
<tr>
<td>lnc-phox2bb\textsuperscript{a173}, TSS-del.</td>
<td>ENSDART00000158002</td>
<td>652 bp</td>
<td>99%</td>
<td>No</td>
<td>Yes</td>
<td>smnt1</td>
<td>phox2bb</td>
<td>Syntenic conservation</td>
</tr>
<tr>
<td>lnc-phox2bb\textsuperscript{a173}, gene del.</td>
<td>ENSDART00000158002</td>
<td>9361 bp</td>
<td>87%</td>
<td>Yes</td>
<td>No</td>
<td>smnt1</td>
<td>phox2bb</td>
<td>Syntenic conservation</td>
</tr>
<tr>
<td>lnc-3852\textsuperscript{a197}, TSS-del.</td>
<td>ENSDART00000153852</td>
<td>447 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
<td>lim1a</td>
<td>hoxc1a</td>
<td>Maternal expression, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-1562\textsuperscript{a173}, TSS-del.</td>
<td>ENSDART00000131562</td>
<td>409 bp</td>
<td>90%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>fgf10a</td>
<td>Maternal expression, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-3962\textsuperscript{a196}, TSS-del.</td>
<td>ENSDART00000153982</td>
<td>352 bp</td>
<td>97%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>bmp2b</td>
<td>Maternal expression, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-6269\textsuperscript{a130}, TSS-del.</td>
<td>ENSDART00000156269</td>
<td>535 bp</td>
<td>99%</td>
<td>No</td>
<td>Yes</td>
<td>tbx1</td>
<td>*</td>
<td>Maternal expression, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-2154\textsuperscript{a130}, TSS-del.</td>
<td>ENSDART00000132154</td>
<td>546 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
<td>ryc</td>
<td>nrf2f5</td>
<td>Maternal expression, Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-1200\textsuperscript{a125}, TSS-del.</td>
<td>Chr12:1708389-1925779:1</td>
<td>590 bp</td>
<td>95%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>zp11</td>
<td>Maternal expression, Longest selected lncRNA</td>
</tr>
<tr>
<td>lnc-1200\textsuperscript{a125}, gene del.</td>
<td>Chr12:1708389-1925779:1</td>
<td>203.8 kb</td>
<td>84%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>zp11</td>
<td>Maternal expression, Longest selected lncRNA</td>
</tr>
<tr>
<td>lnc-2646\textsuperscript{a130}, TSS-del.</td>
<td>ENSDART00000152646</td>
<td>240 bp</td>
<td>97%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>dkk1b</td>
<td>Proximity to developmental regulatory genes</td>
</tr>
<tr>
<td>lnc-4468\textsuperscript{a196}, TSS-del.</td>
<td>ENSDART00000154468</td>
<td>306 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
<td>fam169ab</td>
<td>l0a5</td>
<td>Proximity to developmental regulatory genes, Low expression level</td>
</tr>
<tr>
<td>lnc-0600\textsuperscript{a196}, TSS-del.</td>
<td>Chr6:59414652-59443141:1</td>
<td>244 bp</td>
<td>95%</td>
<td>No</td>
<td>Yes</td>
<td>*</td>
<td>pl3</td>
<td>Proximity to developmental regulatory genes, Low expression level</td>
</tr>
<tr>
<td>lnc-0600\textsuperscript{a196}, TSS-del.</td>
<td>Chr9:6684669-6691350:1</td>
<td>377 bp</td>
<td>83%</td>
<td>No</td>
<td>Yes</td>
<td>pou3f3a</td>
<td>*</td>
<td>Syntenic conservation, Low expression level</td>
</tr>
</tbody>
</table>
Table 1: Summary of lncRNA features and mutant phenotypes

<table>
<thead>
<tr>
<th>lncRNA</th>
<th>Feature</th>
<th>Transcript ID</th>
<th>Deletion Size</th>
<th>Percentage Decrease</th>
<th>Embryonic Phenotypes</th>
<th>Selection Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnc-8507a79, mTSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000158507</td>
<td>323 bp</td>
<td>81%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-8507a79, mTSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000158507</td>
<td>9773 bp</td>
<td>95%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-1300a79, TSS-del.</td>
<td>Leaderlike</td>
<td>Chr13:4535992-4538275</td>
<td>367 bp</td>
<td>92%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-7118a79, TSS-del.</td>
<td>Trailerlike</td>
<td>ENSDART00000157118</td>
<td>438 bp</td>
<td>82%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-6912a79, TSS-del.</td>
<td>Trailerlike</td>
<td>ENSDART00000156913</td>
<td>333 bp</td>
<td>72%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-5888a79, TSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000155888</td>
<td>606 bp</td>
<td>96%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-6490a79, TSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000146490</td>
<td>607 bp</td>
<td>99%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-5888a79, gene del.</td>
<td>Leaderlike</td>
<td>ENSDART00000155888</td>
<td>5568 bp</td>
<td>93%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-3266a79, TSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000141666</td>
<td>544 bp</td>
<td>96%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-6490a79, gene del.</td>
<td>Leaderlike</td>
<td>ENSDART00000146490</td>
<td>8378 bp</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-6666a79, TSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000141666</td>
<td>597 bp</td>
<td>96%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-4145a79, TSS-del.</td>
<td>Leaderlike</td>
<td>ENSDART00000154149</td>
<td>491 bp</td>
<td>98%</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>lnc-4145a79, gene del.</td>
<td>Leaderlike</td>
<td>ENSDART00000154149</td>
<td>35.11 kb</td>
<td>100%</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Cyrano
cyrano is evolutionarily conserved lncRNA and based on morpholino studies, has been suggested to have essential functions during zebrafish embryogenesis and brain.
morphogenesis. *cyrano* has also been suggested to act as a sponge (decoy-factor) for HuR during neuronal proliferation, regulate *miR-7* mediated embryonic stem cell differentiation, and control the level of *miR-7* in the adult mouse brain. We generated two mutant alleles that removed the TSS (*cyrano*a171) or the gene (*cyrano*a172), including the highly conserved *miR-7* binding-site (Fig 2A, B). The expression level of the nearby gene (*oip5*) was not affected in either of these mutants (Fig 1- S4). In contrast to previous morpholino studies in zebrafish but in support of recent findings in mouse, *cyrano* mutants developed normally and were viable and fertile.

The difference between morphant and mutant phenotypes might be caused by compensation in the mutants. To test this possibility, we injected the previously used morpholinos targeting the first exon-intron boundary (e1i1) or the conserved *miR-7* binding site (CMiBS) into wild type and homozygous deletion mutants. The TSS-mutant allele lacked the e1i1 morpholino binding site and the gene deletion allele lacked the CMiBS morpholino binding site (Fig 2A). The previously reported phenotypes, including small heads and eyes, heart edema, and kinked tails were found in both wild type and mutants (Fig 2C), demonstrating that the morpholino-induced phenotypes were non-specific. These results reveal that *cyrano* transcripts or their evolutionarily conserved *miR-7* binding site, are not required for embryogenesis, viability or fertility.
Figure 2: Normal embryogenesis of cyrano mutants

A) The positions of TSS-deletion allele and gene deletion allele are marked by dashed red lines. Green box represents the conserved element in cyrano which is complementary to miR-7. Solid red lines indicate the position of the first exon-intron boundary (e1i1) morpholino and conserved microRNA binding site (CMiBS) morpholinos. Arrows flanking black dotted line mark the primer binding sites for qRT-PCR product. B) Representative images of in situ hybridization for cyrano in wild type (15/15) and both homozygous TSS-deletion (21/22) and gene deletion (18/18) 1-dpf. C) At 2-dpf gene deletion mutants (lower-left), (and TSS-deletion mutants, not shown) were not different from the wild type embryos (upper-left). Morpholino injected wild-type embryos (upper-middle and upper-left) reproduced observed phenotype in Ulitsky et. al. Morpholino injected deletion-mutants, lacking the corresponding binding sites for morpholinos, (lower-middle and lower-left) were comparable to morpholino injected wild types.
gas5

gas5 is an evolutionarily conserved lncRNA (growth-arrest specific 5)\(^{38}\) that is highly expressed in early development (Fig 3B) and hosts several snoRNAs implicated in zebrafish development\(^{39}\). Knockdown and knockout studies in cell culture\(^{40}\) have indicated that gas5 might act as a tumor suppressor\(^{41}\) and exert effects at distant genomic sites\(^{42}\). However, the role of this lncRNA in development has not been studied in any vertebrate. Our gas5\(^{a173}\) mutant allele removed the sequences containing the TSS (-169 to +127) (Fig 3A) and resulted in complete elimination of its expression (Fig 3B, 3D). Expression of the neighboring gene osbpl9, encoding a lipid binding protein, was increased by 50% (Fig 3D). Previous studies have shown that gas5 lncRNA can act in trans to affect pten expression (ptena and ptenb in zebrafish) by sequestering specific microRNAs\(^{43-45}\). Additionally, gas5 transcript can mimic Glucocorticoid Response Element and act as a decoy factor (riborepressor) for the Glucocorticoid Receptor (nr3c1) mediated transcription\(^{46}\). We analyzed the expression level changes of these genes in MZgas5\(^{a173}\) embryos (at 1-dpf) and found significant upregulation for ptena in MZgas5\(^{a173}\) mutants (Fig 3E). Despite these changes in gene expression, gas5\(^{a173}\) mutants were indistinguishable from wild type (Fig 3C), reached adulthood and were fertile.
Figure 3: Normal embryogenesis of gas5 mutants

A) Position of the TSS-deletion allele in gas5 is marked by dashed red line. Arrows flanking black dotted lines mark the primer binding sites for 5'-qPCR and 3'-qPCR products. B) Representative in situ hybridization images for gas5 in wild type (11/11) and homozygous TSS-deletion mutants (11/11). C) Maternal and Zygotic gas5 (MZgas5) mutant embryos at 1-dpf were indistinguishable from the wild-type embryos at the same developmental stage (not shown). D) Expression level of gas5 and osbpl9 measured by qRT-PCR. Tor3A, the other neighboring gene, was not expressed at the investigated time-point. E) Expression level of gas5, its trans targets ptena, ptenb and nr3c1 measured by qRT-PCR. The statistical significance of the observed changes was determined using t-test analysis and represented by star marks (*, **, *** and **** respectively mark p-values < 0.05, <0.01, <0.001 and <0.0001).
*Lnc-setd1ba*

*Lnc-setd1ba* is the zebrafish orthologue of human LIMT\(^47\) (LncRNA Inhibiting Metastasis), which has been implicated in basal-like breast cancers. It is expressed from a shared promoter region that also drives the expression of the histone methyltransferase *setd1ba* in opposite direction (Fig 4A). Evolutionary conservation in vertebrates and proximity to *setd1ba*, whose mouse homolog is essential for embryonic development\(^48,49\) prompted us to investigate the function of this lncRNA in zebrafish. We removed the gene of *lnc-setd1ba* downstream of its TSS (3137bp deletion) (*lnc-setd1ba\(^{a174}\)*). In situ hybridization and qRT-PCR revealed absence of lncRNA expression (Fig 4C and 4E) and strong upregulation of *setd1ba* (Fig 4D and 4E) during cleavage stages and slight upregulation of *setd1ba* and the other neighboring gene *rhoF* at one-day post fertilization (1-dpf) (Fig 4E). Despite these changes, maternal-zygotic *lnc-setd1ba\(^{a174}\)* mutants were indistinguishable from wild type (Fig 4B), reached adulthood and produced normal progeny.
Figure 4: Normal embryogenesis of \textit{inc-setd1ba} mutants

A) The relative position of \textit{inc-setd1ba} and the protein-coding gene \textit{setd1ba}. The gene deletion region is marked by dashed red line. Arrows flanking black dotted line mark the primer binding sites for qRT-PCR product. B) Maternal and zygotic \textit{inc-setd1ba} mutants were not different from wild-type embryos at 1-dpf. C) Representative images of in situ hybridization for \textit{inc-setd1ba} at 4-8 cell stage mutant (18/18) and wild-type (25/25) embryos. D) In situ hybridization for the protein-coding mRNA, \textit{setd1ba} (9/11) in \textit{inc-setd1ba} mutants compared to the wild-type embryos (15/15). E) qRT-PCR at 1-cell stage and 1-dpf for the lncRNA and its neighboring genes \textit{rhoF} and \textit{setd1ba}. The statistical significance of the observed changes was determined using t-test analysis and represented by star marks (ns, *, **, ***, and **** respectively mark p-values \( \geq 0.05, < 0.05, <0.01, <0.001 \) and <0.0001).
Squint encodes a Nodal ligand involved in mesendoderm specification\textsuperscript{50,51}. The previously studied *squint* insertion mutant alleles (*squint\textsuperscript{Hi975Tg} \textsuperscript{50} and *squint\textsuperscript{cz35} \textsuperscript{51*}) lead to delayed mesendoderm specification and partially penetrant cyclopia\textsuperscript{52}. Morpholino and misexpression studies have suggested an additional, non-coding role for maternally provided *squint*, wherein the *squint* 3'UTR mediates dorsal localization of *squint* mRNA, induces the expression of dorsal mesoderm genes, and is required for the development of dorsal structures\textsuperscript{53,54}. This mode of activity assigns *squint* to the cncRNA family - RNAs with both protein-coding and non-coding roles\textsuperscript{55}. To investigate the non-coding roles of *squint* mRNA we generated a deletion allele (*squint\textsuperscript{a175}*) that lacked most of the protein coding region and the 3'UTR, including the Dorsal Localization Element (DLE) implicated in maternal *squint* RNA localization\textsuperscript{56} (Fig 5A). In this allele 525bp (178bp 5'UTR, 280bp first exon and 67bp of second exon) out of the 1592bp-long mature transcript remain in the genome (Fig 5A). In situ hybridization (Fig 5B) and qRT-PCR (Fig 5C) showed that the level of remaining *squint* transcript was greatly reduced (~90%). MZ*squint*\textsuperscript{a175} embryos displayed partially penetrant cyclopia, similar to existing protein-disrupting *squint* alleles (Fig 5D)\textsuperscript{50,51,57}, but the defects proposed to be caused by interference with *squint* non-coding activity\textsuperscript{53} were not detected.

To further test whether *squint* mRNA might have non-coding roles, we injected wild-type and MZ*squint*\textsuperscript{a175} embryos with either control RNA, full-length *squint* mRNA, a non-coding version of *squint* mRNA, or the putative transcript produced in *squint*\textsuperscript{a175} (Fig 5-S1). We found that in contrast to wild-type *squint* mRNA, control RNA, non-protein
coding *squint* RNA or *squint*\(^{a175}\) RNA did not cause any phenotypes and did not rescue MZ*squint*\(^{a175}\) mutants. These results indicate that *squint* 3'UTR does not have the previously proposed non-coding functions and that the *squint* transcript may not be a member of the cncRNA family.

![Diagram](image)

**Figure 5: No non-coding function for *squint* 3'UTR**

A) The position of untranslated regions (brown), coding region (green), putative Dorsal Localization Element- DLE (blue) and the gene deletion (red dashed line) in the *squint* genomic locus. Arrows flanking black dotted line mark the primer binding sites for qRT-PCR product. B) In situ hybridization for *squint* at 8-cell stage on wild-type (18/20) and MZ*squint*\(^{a175}\)(17/17) embryos. C) qRT-PCR for *squint* and *eif4ebp1*...
on wild-type and MZsquintEmbryos at 1 cell stage. D) Two representative MZsquintEmbryos. E) MZsquintEmbryonic phenotype (N=4 independent crosses, n=360 embryos). The statistical significance of the observed changes was determined using t-test analysis and represented by star marks (ns, *, **, ***, and **** respectively mark p-values ≥ 0.05, < 0.05, <0.01, <0.001 and <0.0001).

Transcript-independent phenotype at Inc-phox2bb locus

Lnc-phox2bb neighbors phox2bb and smtnl1. Phox2bb is a transcription factor implicated in the development of the sympathetic nervous system, while smtnl1 has been implicated in smooth muscle contraction. Whole-gene deletion of Inc-phox2bb (Inc-phox2bbEmb) (Fig 6A) led to jaw deformation and failure to inflate the swim-bladder (Fig 6B), and no homozygous mutant fish survived to adulthood. Like the whole-gene deletion allele, the TSS-deletion allele (Inc-phox2bbEmb) lacked Inc-phox2bb RNA (Fig 6E), but in contrast to the whole-gene deletion mutants, TSS-deletion mutants developed normally and gave rise to fertile adults. To determine the cause of this difference, we analyzed the expression level and pattern of neighboring genes. We found that the anterior expression domain of phox2bb in the hindbrain was absent in the whole-gene deletion allele (Fig 6D). This finding is consistent with the observation that the deleted region contains enhancer elements for phox2bb, conserved non-coding elements (CNEs), and histone marks related to enhancer regions (H3K4me1 and H3K27Ac). We also found that the expression level of smtnl1 increased in gene deletion mutants relative to the TSS-deletion mutant and wild type (Fig 6E). These results indicate that Inc-phox2bb RNA is not required for normal development but that the Inc-phox2bb overlaps with regulatory elements required for proper expression of phox2bb and smtnl1 (Fig 6E).
Figure 6: Requirement for *Inc-phox2bb* genomic elements but not RNA

A) The red dashed lines depict the respective positions of the *Inc-phox2bb* TSS and gene deletion. Arrows flanking black dotted line mark the primer binding sites for qRT-PCR product. 

B) Homozygous gene deletion mutants but not the TSS-deletion mutants show embryonic defects in jaw formation (arrow head) and swim bladder inflation (asterisk) by 4-dpf. 

C) Histone marks (H3K4me1 and H3K27ac) associated with enhancer activity\(^{64}\) and conserved noncoding elements (CNEs)\(^{63}\) overlap with gene deletion. 

D) *phox2bb* expression pattern in the TSS and gene deletions. 

E) qRT-PCR analysis on MZ TSS-deletion and gene deletion mutants. The statistical significance of the observed changes was determined using t-test analysis and represented by star marks (*, **, ***, and **** respectively mark p-values < 0.05, <0.01, <0.001 and <0.0001).
In summary, our systematic mutant studies indicate that none of the 25 lncRNAs analyzed here are essential for embryogenesis, viability or fertility, including the prominent lncRNAs cyrano, gas5, and lnc-setd1ba. Additionally, they refute the proposed non-coding function of squint RNA. Our phenotypic screen does not exclude more subtle phenotypes; e.g. in behavior or brain activity\textsuperscript{65-67}. This mutant collection can now be analyzed for subtle, context specific or redundant functions, but extrapolation suggests that most individual zebrafish lncRNAs are not required for embryogenesis, viability or fertility.
Materials and Methods

Animal care

TL/AB zebrafish (Danio rerio) were used as wild-type fish in this study. Fish were maintained on daily 14hr (light): 10hr (dark) cycle at 28°C. All animal work was performed at the facilities of Harvard University, Faculty of Arts & Sciences (HU/FAS). This study was approved by the Harvard University/Faculty of Arts & Sciences Standing Committee on the Use of Animals in Research & Teaching (IACUC; Protocol #25-08).

Cas9 mediated mutagenesis

Guide RNAs (gRNAs) were designed using CHOPCHOP and synthesized in pool for each candidate as previously described. (See supplementary file 1 for the gRNA sequences). gRNAs were combined with Cas9 protein (50 μM) and co-injected (~1 nL) into the one-cell stage TL/AB wild-type embryos. Genomic DNA from 10 injected and 10 un-injected siblings was extracted and screened for the difference in amplified band pattern from the targeted region (See supplementary file 1 for the genotyping primer sequences). The rest of injected embryos were raised to adulthood, crossed to wild-type fish and screened for passing the mutant allele to the next generation. Founder fish with desirable mutations were selected and confirmed by Sanger sequencing of the amplified mutant allele. Heterozygous mutants were crossed together to generate homozygous mutants. At least 15 adult homozygous mutant pairs per allele were crossed to test fertility of mutants and to generate maternal and zygotic mutants (MZ) devoid of maternally and zygotic IncRNA activity.
Phenotype scoring procedure

Visual assessment of live embryos and larvae performed\textsuperscript{30,31} to identify mutant phenotypes, ranging from gastrulation movements and axis formation to the formation of brain, spinal cord, floor plate, notochord, somites, eyes, ears, heart, blood, pigmentation, vessels, kidney, pharyngeal arches, head skeleton, liver, and gut.

At day five formation of swim bladder and overall appearance of the embryos were checked again (at any stage 60-100 embryos were scored). Sixty to hundred fish from heterozygous mutant crosses were grown to adulthood and genotyped to identify the viability of adult homozygous fish. Validated homozygous mutant fish were further crossed together to test for potential fertility phenotypes or putative maternal functions of candidate IncRNAs.

Antisense RNA synthesis and in situ hybridization

Antisense probes for in situ hybridization were transcribed using the DIG RNA labeling kit (Roche). All RNAs were purified using EZNA Total RNA kits (Omega Biotek). Embryos were fixed in 4% formaldehyde overnight at 4°C (embryos younger than 50% epiboly fixed for 2 days). In situ hybridizations were performed according to standard protocols\textsuperscript{71}. NBT/BCIP/Alkaline phosphatase-stained embryos were dehydrated in methanol and imaged in benzyl benzoate:benzyl alcohol (BBBA) using a Zeiss Axio Imager.Z1 microscope.
qRT-PCR

Total RNA was isolated from 3 individuals or sets of 10-20 embryos per condition using EZNA Total RNA kits (Omega Biotek). cDNA was generated using iScript cDNA Synthesis kit (Bio-Rad). qPCR was conducted using iTaq Universal SYBR Green Supermix (Bio-Rad) on a CFX96 (Bio-Rad). Gene expression levels were calculated relative to a reference gene, *ef1a*. Three technical replicates were used per condition. The qPCR primer sequences are listed in supplementary file 1.

Bright field Imaging

Embryos were anesthetized in Tricaine (Sigma) and mounted in 1% low melting temperature agarose (Sigma) with Tricaine, then imaged using a Zeiss SteREO Discovery.V12 microscope or Zeiss Axio Imager.Z1 microscope. Images were processed in FIJI/ImageJ72. Brightness, contrast and color balance was applied uniformly to images.

Sense RNA synthesis and injections

The sequences for the wild-type *squint* mRNA, non-protein coding *squint* transcript (One Adenine base was added after 8 in-frame ATG codons, and the 3'UTR sequence kept unchanged) and the *squinta175* transcript were synthesized as gBlocks (IDT) containing 5’ Xhol cut site and 3’ NotI site. Fragments were digested and inserted the pCS2 plasmid. Positive colonies were selected, and sanger sequenced to assure the accuracy of the gene synthesis process. Sequences of the constructs are provided in supplementary file 1. mRNA was in vitro transcribed by mMessage mMACHINE (Ambion)
and purified by EZNA Total RNA kits (Omega Biotek). *h2b-gfp* was used as control mRNA. Each injection mix contained 30ng/ul of *squint* or control mRNA. 1nl of mRNA mix was injected into the yolk of one-cell stage embryos. Morpholinos were ordered from Gene Tools and injected based on Ulitsky *et al.*\textsuperscript{23}.
Acknowledgments

We thank current and former members of the Schier laboratory, particularly Andrea Pauli and Guo-Liang Chew for their helpful suggestions and support during the early phases of this project, Jeffrey Farrell, Nathan Lord and Maxwell Shafer for their critical comments on the manuscript, and the Harvard zebrafish facility staff for technical support. This work was supported by Leopoldina postdoctoral fellowship LPDS2014-01 to M.G. and NIH grant R01HD076708 to A.F.S.
References:


Johnston, C. M., Newall, A. E., Brockdorff, N. & Nesterova, T. B. Enox, a novel
gene that maps 10 kb upstream of Xist and partially escapes X inactivation.  

Wutz, A. *et al.* Imprinted expression of the Igf2r gene depends on an intronic 

Latos, P. A. *et al.* Airn transcriptional overlap, but not its IncRNA products, 
induces imprinted Igf2r silencing. *Science (New York, N.Y.)* **338**, 1469-1472, 

Miyoshi, N. *et al.* Identification of an imprinted gene, Meg3/Gtl2 and its human 
homologue MEG3, first mapped on mouse distal chromosome 12 and human 
chromosome 14q. *Genes to cells : devoted to molecular & cellular mechanisms* 

Kobayashi, S. *et al.* Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally 
expressed imprinted genes closely located to the maternally expressed imprinted 
genes: mouse Meg3/Gtl2 and human MEG3. *Genes to cells : devoted to 

Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse 

Feil, R., Walter, J., Allen, N. D. & Reik, W. Developmental control of allelic 
methylation in the imprinted mouse Igf2 and H19 genes. *Development* 


Borman, M. A., Freed, T. A., Haystead, T. A. & Macdonald, J. A. The role of the calponin homology domain of smoothelin-like 1 (SMTNL1) in myosin


Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. 


**Supplementary Figures**

**Figure 1 supplementary Figure 1**: Size, relative distance and orientation of selected lncRNAs and their neighboring genes

A) lncRNA names and sizes are shown in the middle section (blue columns). The distance, size and transcriptional orientation of the neighboring genes, in a 200kb window centered on lncRNA’s TSS are shown on the left (upstream neighbor) and on the right (downstream neighbor). The transcription orientation is represented by green (in the same direction as lncRNA) and magenta (in the opposite direction of lncRNA). B) Visual representation of data in A. All sizes and distances are in Kb.
Figure 1 supplementary Figure 2: Expression levels of selected lncRNAs and their neighboring protein-coding genes

LncRNAs are color coded as blue (Intergenic), brown (Overlapping) and green (Divergent/Promoter associated) (see Fig S1B). For each lncRNA and its upstream (top) and downstream (bottom) neighbor, the expression levels at 10 early-developmental stages are shown\textsuperscript{25}. The scale is log2 (FPKM+1) value, represented as gradient between 0 (white) and 8 (magenta).
Figure 1 supplementary Figure 3: Cas9-mediated deletion approach for generating lncRNA knockouts

6 gRNAs (three at either side of the TSS) were used to remove TSS. 9 guide RNAs (the first 6 plus three additional gRNAs around the Transcriptional Termination Site, TTS) were used to generate the gene deletions. Relative positions of genotyping primers are indicated by numbered circles.
Figure 1 supplementary Figure 4: Summary of qRT-PCR analysis for IncRNA and their neighboring genes

Visual representation of the expression level changes for each IncRNA and its neighboring genes in homozygous deletion mutants. Three biological replicates for homozygous mutant and wild-type samples.

Log2 of fold change between -4 (magenta) and 4 (green) is shown.
Figure 5 supplementary Figure 1: Dorsalization induced by Overexpression of *squint* mRNA but not its non-protein coding version

A) Schematic representation of injected mRNAs. Cap-analog is indicated by in blue circles at the beginning of each mRNA. *squint* non-protein coding mRNA was generated by adding 8 Adenine-nucleotides (red circles) after in-frame ATG codons. B) Table shows scoring outcome of observed phenotypes in embryos injected with 30pg of each indicated mRNA. C) Representative embryos showing typical wild-type, *squint* mutant or dorsalized morphology. Ambiguous phenotypes were scored as “Affected”.

<table>
<thead>
<tr>
<th>Genotype, Injected RNA</th>
<th>% at Bud stage</th>
<th>% at 1 day</th>
<th>% at 1 day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dead</td>
<td>Dead</td>
<td>Dorsalized</td>
</tr>
<tr>
<td>WT, Ctrl. RNA</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MZsquint&lt;sup&gt;ATG&lt;/sup&gt;, Ctrl. RNA</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WT, <em>squint</em>&lt;sup&gt;WT&lt;/sup&gt;, Ctrl. RNA</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WT, <em>squint</em>&lt;sup&gt;WT&lt;/sup&gt;</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>MZsquint&lt;sup&gt;ATG&lt;/sup&gt;, <em>squint</em>&lt;sup&gt;WT&lt;/sup&gt;</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WT, <em>squint</em> non-protein coding</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MZsquint&lt;sup&gt;ATG&lt;/sup&gt;, <em>squint</em> non-protein coding</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>WT, <em>squint</em>&lt;sup&gt;WT&lt;/sup&gt;</td>
<td>13</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>MZsquint&lt;sup&gt;ATG&lt;/sup&gt;, <em>squint</em>&lt;sup&gt;WT&lt;/sup&gt;</td>
<td>15</td>
<td>3</td>
<td>19</td>
</tr>
</tbody>
</table>
Supplementary File 1:
This compressed folder contains three Excel files for the sequences of gRNAs, genotyping and qRT-PCR primers (for lncRNAs and their neighboring genes) and also the annotated sequence files (.ape) for each lncRNA and their deleted segments.

Supplementary File 2:
This genome-browser-compatible fie is in the bed formant, containing the coordinates for all the lncRNAs investigated in this manuscript based on the GRCz11 (GCA_000002035.4).