Dynamic Expression of Long Non-Coding RNAs (IncRNAs) in Adult Zebrafish

Kriti Kaushik1,3,*, Vincent Elvin Leonard1,*, Shamsudheen KV1, Mukesh Kumar Lalwani1, Saakshi Jalali2,3, Ashok Patowary1, Adita Joshi1, Vinod Scaria2,3, Sridhar Sivasubbu1,3*

1 Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology, Delhi, India, 2 G.N. Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology, Delhi, India, 3 Academy of Scientific and Innovative Research (ACSIR), Anusandhan Bhavan, New Delhi, India

Abstract

Long non-coding RNAs (IncRNA) represent an assorted class of transcripts having little or no protein coding capacity and have recently gained importance for their function as regulators of gene expression. Molecular studies on IncRNAs have uncovered multifaceted interactions with protein coding genes. It has been suggested that IncRNAs are an additional layer of regulatory switches involved in gene regulation during development and disease. LncRNAs expressing in specific tissues or cell types during adult stages can have potential roles in function, formation, maintenance and repair of tissues and organs. We used RNA sequencing followed by computational analysis to identify tissue restricted IncRNA transcript signatures from five different tissues of adult zebrafish. The present study reports 442 predicted IncRNA transcripts from adult zebrafish tissues out of which 419 were novel IncRNA transcripts. Of these, 77 IncRNAs show predominant tissue restricted expression across the five major tissues investigated. Adult zebrafish brain expressed the largest number of tissue restricted IncRNA transcripts followed by cardiovascular tissue. We also validated the tissue restricted expression of a subset of IncRNAs using independent methods. Our data constitute a useful genomic resource towards understanding the expression of IncRNAs in various tissues in adult zebrafish. Our study is thus a starting point and opens a way towards discovering new molecular interactions of gene expression within the specific adult tissues in the context of maintenance of organ form and function.

Introduction

The enormous success of major genome sequencing projects in this century was soon consumed with greater challenge of discovering and functionally annotating transcripts encoded by the genome. Thousands of novel RNA transcripts were uncovered by systematic sequencing of full-length cDNA libraries in eukaryotes [1,2]. These studies estimated that over 70–75% of the eukaryotic genome encoded for transcripts of diverse nature [3]. Many of these transcripts did not have an obvious potential to encode for protein and were popularly called non-coding RNAs (ncRNAs). Genome-wide large-scale projects such as FANTOM 3 uncovered incomprehensible nature of the non-coding RNA transcription by detecting ~35,000 non-coding RNA transcripts from ~10,000 distinct loci in the mouse genome [1]. In human cells, genome-wide transcriptome mapping as part of the ENCODE project annotated about 18,400 non-coding RNAs including tRNA, rRNA, HYPERLINK “http://en.wikipedia.org/wiki/MicroRNA”\microrn\^{A}A\text{A}\text{A} and other non-coding RNA genes [3,4].

The non-coding RNAs (ncRNA) are broadly classified into long and small ncRNAs depending upon length of the transcript and have been implicated in regulating expression of key genes involved in the maintenance of biological processes [5–7]. At least four classes of regulatory small ncRNAs have been described including short interfering RNAs (siRNA), small nucleolar RNA ( snoRNA), piwi-interacting RNAs (piRNAs) and microRNAs (miRNAs) [3]. Among the small ncRNAs, miRNAs are the most well studied, phylogenetically conserved and are found to be indispensable for the development and functioning of an organism [6].

Long non-coding RNAs (LncRNAs) have emerged as a major class of novel regulating transcripts, which are ≥200 nucleotides and display spatio-temporal expression suggesting precise function [8]. In contrast to small ncRNAs, LncRNAs form an enigmatic class of transcripts, which regardless of having characteristic mRNA signatures such as 5' capping, splicing, and poly-adenylation are not functionally well annotated [9–11]. Xist and H19 were amongst the earliest discovered LncRNAs using conventional gene discovery methods [12–14]. Subsequently, several other LncRNAs have been discovered [7]. The Allen Brain Atlas has documented ~49 LncRNAs within the mouse brain, similarly ~1,600 long intervening non-coding RNA (lincRNAs) have been identified in mouse cell types using epigenetic marks and ~3,300 lincRNAs have been discovered in human cell types [15–17]. The importance of long non-coding RNA transcription is underscored...
by the fact that the human genome has four times more lncRNA sequences represented than the protein coding transcripts [10]. Projects like GENCODE (http://www.gencodegenes.org) and NONCODE (http://www.noncode.org) have focused on identification and annotation of lncRNAs. At least 9,640 human lncRNA loci, representing ~15,512 transcripts have been reported by GENCODE 7 and over 11,000 lncRNAs were identified in the mouse genome by the FANTOM consortium [19,20].

Several model organisms including zebrafish have been explicitly used for deciphering the functional role of lncRNAs [6,20–25]. Zebrafish has emerged as an excellent vertebrate model organism for studies focusing on discovery and biology of non-coding RNA transcription in developing embryos as well as adult tissues [26–28]. The functional roles and interactions of small and long ncRNA transcriptome have been well studied in developing zebrafish embryos, worms and flies [29–33]. A recent study identified 550 lncRNAs in three developmental stages of zebrafish by using chromatin marks, RNA sequencing and Poly (A) site mapping. Conserved lncRNAs such as Cyrano (linc-oip5) and megamind (linc-bire6) have been documented to have specific function during zebrafish brain morphogenesis and eye development respectively [25]. An independent study also identified 1,133 long non-coding transcripts originating from diverse genomic loci through transcriptome sequencing of eight developmental stages of zebrafish. Furthermore, the study also documented tissue-specific expression and sub-cellular localization patterns of long non-coding RNA transcripts [24]. Collectively, these studies suggest that lncRNAs may have spatial and temporal expression with potentially important roles during embryogenesis in zebrafish. However, relatively less is known about lncRNAs and their biological functions in adult tissues of zebrafish. Deciphering the repertoire and expression profiles of lncRNAs in adult tissues of zebrafish would enable better understanding of gene regulation within individual tissues types.

In this study, we report a compendium of lncRNAs expressed in five major tissue types of adult zebrafish. In complement to the recent studies in zebrafish that focused on identification of lncRNAs across narrow windows of early developmental time points [24,25], we have analyzed and compiled the lncRNA transcriptome within functional tissues in adult zebrafish. Using RNA sequencing of five tissue types of adult zebrafish viz heart, liver, muscle, brain and blood of adult zebrafish and RNA sequence reads were generated using sequencing-by-synthesis method [34]. Approximately, 193 million raw paired-end sequence reads of 51 base pairs (bp) were obtained from five tissue libraries. Sequence reads were aligned to the zebrafish reference genome (Ensembl Zv9 build; hereafter called as Zv9). Approximately, 171 million sequencing reads (88.66%) were successfully mapped back to the reference genome (Table 1). These mapped reads were processed further for analysis.

### Results

**Sequence data generation and mapping**

Poly-A RNA was obtained from total RNA for five tissues viz, heart, liver, muscle, brain and blood of adult zebrafish and RNA sequence reads were generated using sequencing-by-synthesis method [34]. Approximately, 193 million raw paired-end sequence reads of 51 base pairs (bp) were obtained from five tissue libraries. Sequence reads were aligned to the zebrafish reference genome (Ensembl Zv9 build; hereafter called as Zv9). Approximately, 171 million sequencing reads (88.66%) were successfully mapped back to the reference genome (Table 1).

| Tissue       | Heart       | Liver       | Muscle      | Brain       | Blood       | Total       |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Raw Reads    | 43,928,174  | 50,627,322  | 34,505,562  | 29,073,480  | 34,161,882  | 193,196,420 |
| Mapped Reads | 38,076,909  | 43,416,137  | 32,421,815  | 27,347,501  | 30,024,662  | 171,287,024 |

| Tissue       | Heart       | Liver       | Muscle      | Brain       | Blood       | Total       |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| (85.6%)      | (85.7%)     | (94%)       |             |             |             |             |

The total number of sequence reads obtained from the five zebrafish tissues using RNA sequencing is described. Mapped reads represent all transcripts that aligned back to the zebrafish reference genome (Zv9).

doi:10.1371/journal.pone.0083616.t001

Further, the study also documented tissue-specific expression and sub-cellular localization patterns of long non-coding RNA transcripts [24]. Collectively, these studies suggest that lncRNAs may have spatial and temporal expression with potentially important roles during embryogenesis in zebrafish. However, relatively less is known about lncRNAs and their biological functions in adult tissues of zebrafish. Deciphering the repertoire and expression profiles of lncRNAs in adult tissues of zebrafish would enable better understanding of gene regulation within individual tissues types.

In this study, we report a compendium of lncRNAs expressed in five major tissue types of adult zebrafish. In complement to the recent studies in zebrafish that focused on identification of lncRNAs across narrow windows of early developmental time points [24,25], we have analyzed and compiled the lncRNA transcriptome within functional tissues in adult zebrafish. Using RNA sequencing of five tissue types of adult zebrafish viz heart, liver, muscle and blood followed by multi-filter computational analysis pipeline, we predicted 442 putative lncRNA transcript loci that could represent putative non-coding transcripts. These remaining 6,942 putative non-coding transcript loci were subjected to an independent open reading frame (ORF) prediction in all six frames [36]. Based on the ORF prediction, 6,038 transcript loci that could potentially code for thirty or more amino acids were removed from the analysis, as these would represent potential small RNA loci in the genome. The remaining 23,624 predicted transcript loci were evaluated for their coding potential [35]. Of the 23,624 predicted transcripts, 17,132 transcripts had a positive coding potential score, thus representing potential protein coding transcript loci and were removed from further analysis. The 6,492 transcripts with negative coding potential score were retained, as these would represent putative non-coding transcripts. These remaining 6,492 putative non-coding transcript loci were subjected to an independent open reading frame (ORF) prediction in all six frames [36]. Based on the ORF prediction, 6,038 transcript loci that could potentially code for thirty or more amino acids were removed from the analysis, as these would represent potential small peptides [20]. This resulted in a total of 454 non-coding transcript loci. Of the set, 12 transcripts that showed partial overlap with predicted protein coding gene isoforms were removed from further analysis. The remaining 442 predicted transcript loci represent potential lncRNAs identified from the zebrafish tissues. The 442 predicted lncRNAs were analyzed for overlaps with previously known
lncRNA loci in zebrafish [24,25]. We found that 23 lncRNA loci derived from our analysis overlapped with the previous studies. Thus from this study, we identified 419 potential novel lncRNAs (Table S1).

Of the 419 potential novel lncRNAs, we found that 342 lncRNAs were expressed in more than one tissue investigated in this study (Figure 2A, 2B). The remaining 77 lncRNA displayed putative restricted expression to a single tissue and were labeled as “tissue restricted lncRNAs” (Figure 2C, Table S2). Among the five tissues, brain tissue expressed the maximum number of lncRNAs (47) followed by heart tissue (12) and blood tissue (12). Muscle tissue (4) and liver tissue (2) had relatively low number of lncRNAs. Brain as a tissue accounted for 61%, followed by cardiovascular tissues such as heart and blood, which together accounted for 31% of the putative novel lncRNAs. Liver and muscle represented 3–5% of the total collection (Figure 1).
As a part of this study we identified 419 putative novel lncRNAs from five zebrafish tissues, of which 77 putative lncRNAs show tissue restricted differential expression (Figure 1). We have performed detailed expression analysis of 419 potential novel lncRNAs using Fragments Per Kilo base of exons per Million fragments generated (FPKM) scores derived from the RNA sequencing data in order to examine distribution of these lncRNAs across five tissues of zebrafish. Approximately, 50% of the transcripts were expressed in 2–3 tissues and 15% were expressed in all the five tissues (Figure 2A). A Venn diagram representing the overlapping expression of all 419 transcripts in five tissues is shown (Figure 2B), suggesting their dynamic expression across five tissues.

We have also observed that amongst the 77 tissue restricted...

---

Expression profile of tissue specific lncRNome

As a part of this study we identified 419 putative novel lncRNAs from five zebrafish tissues, of which 77 putative lncRNA show tissue restricted differential expression (Figure 1). We have performed detailed expression analysis of 419 potential novel lncRNAs using Fragments Per Kilo base of exons per Million fragments generated (FPKM) scores derived from the RNA sequencing data in order to examine distribution of these lncRNAs across five tissues of zebrafish. Approximately, 50% of the transcripts were expressed in 2–3 tissues and 15% were expressed in all the five tissues (Figure 2A). A Venn diagram representing the overlapping expression of all 419 transcripts in five tissues is shown (Figure 2B), suggesting their dynamic expression across five tissues. We have also observed that amongst the 77 tissue restricted...
Expression of embryonic lncRNA transcripts in adult tissues of zebrafish

Non-coding RNAs have been documented to display a high degree of specificity in their domain of expression. A number of studies have shown tissue-restricted expression for short non-coding RNA such as microRNAs [40–43]. Recently, we reported that expression of miR-142a-3p was restricted to the vasculature endothelium and has a role in developmental angiogenesis in zebrafish [32]. In contrast to rich literature on the tissue-specific expression domain and function of miRNAs, evidence for tissue restricted expression for long non-coding RNA is still forthcoming. Studies have described tissue and cell type specific, spatiotemporal regulated expression of the lncRNA transcripts, suggesting putative functional roles [15, 44, 45]. Studies on the lncRNA expression indicate that brain as a tissue expresses the largest repertoire of lncRNA transcripts and displays conserved expression in the brain and heart tissues also (Figure 4C). "Milkine a (mikla), a protein coding gene that uniquely expresses in brain tissue [39], was chosen to evaluate relative expression of putative brain specific lncRNA transcripts. LncBrM_002 and LncBrM_028 show predominant expression in the brain with trace expression in other tissue types (Figure 4D). T cell acute lymphocytic leukemia protein 1 (tal1) was used as protein coding marker and displayed predominant expression in blood tissue with minimal expression in the brain (Figure 4E). The transcript lncHBl_017 was found to express specifically in the blood tissue and its expression was absent in the other tissues investigated.

We further compared the RNA sequencing derived FPKM values of predicted lncRNAs transcripts with the fold change values of RT-PCR assay in order to evaluate the reproducibility of the tissue restricted lncRNA expression (Figure 5). Analysis showed good concordance between RT-PCR data and FPKM score (Figure 5). This suggests that the trends of tissue restricted lncRNA expression were similar in RNA sequencing and RT-PCR assays. In summary our RT-PCR assay reproduced the relative transcript abundance of predicted tissue restricted lncRNAs similar to that observed by RNA sequencing.

To further verify whether the predicted lncRNA transcripts were predominantly expressed and localized in the specific tissues, we performed whole mount RNA in situ hybridization (WISH) for both brain restricted lncRNAs, LncBrHM_035 and LncBrM_002 in adult brain organ as well as developing embryos (Figure 6). Prior to examining the expression of lncRNA using WISH, we performed 3’ RACE (Rapid Amplification of cDNA Ends) of lncRNA transcripts LncBrHM_035 and LncBrM_002, in order to confirm the directionality of the lncRNA transcript in the genome (data not shown).

LncBrHM_035 transcript displayed distinct localization in the eye, mid and hind brain of 24hpf zebrafish embryos (Figure 6C) and was found to be expressing explicitly in cerebellum of adult zebrafish brain (Figure 6D). Another brain restricted lncRNA transcript, LncBrM_002 could be detected in mid and hind-brain of 24hpf zebrafish embryos (Figure 6E) and showed restricted expression in cerebellum and EG (eminentia granularis) of adult zebrafish brain (Figure 6F). The WISH data revealed that the predicted tissue restricted lncRNA expressed in adult organs and displayed slightly overlapping expression profiles in developing organs during early embryogenesis. In summary, we have used three independent approaches, namely RNA sequencing, RT-PCR and WISH for determining the expression of putative lncRNAs across five tissues. Collectively, the results of the assays suggest that the predicted lncRNAs display defined tissue restricted boundaries of expression.

Discussion

Non-coding RNAs have been documented to display a high degree of specificity in their domain of expression. A number of studies have shown tissue-restricted expression for short non-coding RNA such as microRNAs [40–43]. Recently, we reported that expression of miR-142a-3p was restricted to the vasculature endothelium and has a role in developmental angiogenesis in zebrafish [32]. In contrast to rich literature on the tissue-specific expression domain and function of microRNAs, evidence for tissue restricted expression for long non-coding RNA is still forthcoming. Studies have described tissue and cell type specific, spatiotemporal regulated expression of the lncRNA transcripts, suggesting putative functional roles [15, 44, 45]. Studies on the lncRNA expression indicate that brain as a tissue expresses the largest repertoire of lncRNA transcripts and displays conserved expres-
Evf2, a long non-coding RNA, transcribed from an ultra-conserved genomic region, displays explicit expression in mouse brain and regulates activity of Dlx homeodomain genes across vertebrates [48]. LncRNAs such as Gomafu show distinct localization within sub-cellular compartments (nuclear) in neurons [49]. Another study found that a neural specific lncRNA, CASK regulatory gene (CRG) in Drosophila participates in locomotor and climbing activity [50]. LncRNAs are also known to express as pairs with protein coding genes and co-localize at genomic level in developing brain [51]. LncRNA such as tie-1AS are known to express specifically in vascular endothelium and regulate the tie-1 coding transcript [52].

The roles of lncRNAs such as braveheart, Fendrr and LINCRNA-EPS have been documented in early cardiovascular lineage commitment, heart development and erythroid differentiation respectively [23,53,54]. Apart from directly interacting with protein coding genes, lncRNAs also act as a decoy of miRNA as in the case of linc-MD1, a muscle specific lncRNA [55].

Majority of the literature pertaining to lncRNA in zebrafish is primarily focused on describing functional roles during early developmental stages. However, information regarding their expression profile and biological role in adult organ function and maintenance is limited. This study describes the lncRNA expression landscape from tissues of diverse function in an adult zebrafish. Next generation high throughput sequencing technology was used to capture the polyadenylated transcripts, which were then subjected to a computational analysis pipeline leading to the identification of putative novel lncRNAs from five tissues derived from adult zebrafish. A total of 52,008 transcripts were reconstructed from our RNA sequencing data. A similar number of transcripts 56,535 were reported by Pauli and co-workers in their description of zebrafish embryonic transcriptome. Of 52,008 transcripts identified in our study, 27,691 transcripts corresponded to the RefSeq transcripts and were removed from analysis. The remaining 24,317 transcripts were subjected to the computational analysis for identification of putative lncRNAs (Figure 1).

In this study we identified 442 putative lncRNAs with high confidence from five major tissues of adult zebrafish. Of these, 14 lncRNA transcripts overlapped with those identified from zebrafish developing embryos [24]. We also noticed that only 9 transcripts in our dataset overlapped with the lincRNA dataset of developing zebrafish embryos reported by Ulitsky and co-workers [25]. Reasons for the minimal overlap in lncRNA transcripts between the previous studies and the present work could be attributed to the stringent computational analysis used in this study, which filtered out a large portion of embryonic lncRNAs that are otherwise present as RNA transcripts in the transcriptome dataset obtained from the five tissues of adult zebrafish. We have also examined the overlap of lncRNA transcripts after modifying the ORF cut off from 30 amino acid to 100 amino acid as used by Pauli and co-workers. When the ORF cut off was set to 100 amino acid, the total number of lncRNA transcripts increased from 442 to 6,214. In addition, the overlap of the lncRNA transcripts with the previous studies also increased from 9 to 176 in case of Ulitsky et al.,2011 and 14 to 197 in case of Pauli et al.,2012 (Table S4). However, it is well known that the higher ORF length could potentially add to the false positive predictions of lncRNA transcripts [56]. Therefore, to avoid false predictions, we have followed stringent criteria of 30 amino acid cut off in our study. Furthermore, we have used a non-stranded RNA sequencing approach in our study and this limits the number of lncRNA transcripts that could be predicted. Lastly, we have investigated...
transcriptome from adult tissues of zebrafish, which is known to harbor distinctly different transcriptome repertoire from embryonic stages [57–59].

Out of the 442 lncRNA transcripts predicted from this study, 419 lncRNAs were classified as putative novel as these have not been reported before. Of the 419 putative novel lncRNAs, 342 lncRNAs were found to be expressed in more than one tissue investigated, suggesting that these lncRNAs transcripts may be important regulators of protein coding genes that may be required for maintenance of the corresponding organs/tissues. The remaining 77-lncRNA transcripts were predicted to have predominant expression restricted to one single zebrafish tissue investigated. The expression of individual lncRNA transcripts varies widely in the tissues investigated. All the five tissues have different subsets of uniquely restricted lncRNA transcripts with almost no expression elsewhere. The expression profiles of lncRNA transcripts derived from the RNA sequencing and RT-PCR for the five tissues indicate a good concordance. In addition, the WISH assay showed the unique and non-overlapping expression domains of the two brain restricted lncRNA transcripts lncBrHM_035 and lncBrM_002 in adult brain, which clearly suggests that lncRNA transcripts within a single organ (brain) may have discrete localization patterns that might signify restricted functional activity.

The present study is not without caveats; firstly, we have applied a non-stranded RNA sequencing approach, which limits the

Figure 4. Real time assay for putative tissue restricted lncRNAs. Expression of candidate lncRNA transcripts was analyzed by semi quantitative RT-PCR in A) heart; B) liver; C) muscle; D) brain and E) blood tissues. A tissue specific protein coding marker gene viz cmlc2 (heart); tfr (liver); mdka (brain); mrcub (muscle) and tal1 (blood) was used as standard control. See text for details on selection of protein coding marker genes. LncRNA transcripts investigated for a particular tissue type showed relatively predominant expression in the specific tissue when compared with other tissues.
doi:10.1371/journal.pone.0083616.g004
Figure 5. FPKM values are consistent with lncRNA expression. Expression of lncRNAs validated via RT-PCR for each tissue is compared with their corresponding FPKM values obtained from RNA sequencing. By and large, tissue specificity of the lncRNA transcripts as reflected by FPKM values shows reasonable overlap with their relative expression profiles across tissues obtained from RT-PCR assay. A(i), A(ii) Heart; B(i), B(ii) Liver; C(i), C(ii) Muscle; D(i), D(ii) Brain; and E(i), E(ii) Blood tissues.

doi:10.1371/journal.pone.0083616.g005
number of lncRNA transcripts that could be predicted. Secondly, we have not investigated the chromatin marks flanking the predicted lncRNA transcript loci, which could have revealed additional information on transcript loci. Thirdly, we sequenced only the poly (A) containing RNA transcripts in our study, which prevented the identification of lncRNA transcripts that are devoid of a poly (A) tail [60]. Nevertheless, this catalogue of tissue restricted lncRNA transcripts will be useful for exploring the role of non-protein coding transcriptome in maintenance and repair of tissues. The predominant tissue restricted expression of the lncRNA transcripts may suggest specific functional roles in each tissue type. We speculate that the lncRNA transcripts identified in this study may also help to better understand the recently identified functional interactions amongst mRNA, miRNA and lncRNA [22] in a broader context of processes such as tissue maintenance, repair and regeneration. The strategy outlined here for identifying putative novel lncRNA transcripts can be employed as a methodology for prioritizing and understanding biologically significant of non-coding RNA transcripts. Further, this methodology could be readily applied to a large number of tissue specific fluorescent zebrafish lines for identification of functionally

Figure 6. LncRNAs show tissue restricted expression patterns. Whole mount in situ hybridization of lncRNA transcripts. Shown are images with probes specific to the two indicated brain restricted lncRNAs. Arrow heads indicate the expression domains. A and B Anatomical cartoons of 24 hpf developing zebrafish embryo and adult zebrafish brain. C and D Expression of lncRNA transcript IncBrHM_035. (C) Dorsal view (anterior up) and lateral view (anterior to the left) showing expression in mid-hind brain boundary and hind brain of 24hpf zebrafish embryos. (D) Dorsal view (anterior up) of the adult zebrafish brain showing expression in regions of cerebellar crest (CC). E and F Expression of lncRNA transcript IncBrM_002. (E) Dorsal view (anterior up) and lateral view (anterior to the left) showing expression in fore-brain (FB), mid-hind brain boundary (MHB) and hind brain (HB) of 24hpf zebrafish embryos. (F) Dorsal view (anterior up) of the adult zebrafish brain showing expression in the regions of CC and a localized signal in eminentia granularis (EG). MB, mid brain; OB, olfactory bulb; Tel, telencephalon; Ha, habenula; Teo, optic tectum; MO, medulla oblongata. doi:10.1371/journal.pone.0083616.g006
significant non-coding RNA transcripts in specific biological pathways.

**Materials and Methods**

**Ethics Statement**

Fish experiments were performed in strict accordance with the recommendations and guidelines laid down by the CSIR Institute of Genomics and Integrative Biology, India. The protocol was approved by the Institutional Animal Ethics Committee (IAEC) of the CSIR Institute of Genomics and Integrative Biology, India. All efforts were made to minimize animal suffering.

**RNA isolation**

Adult wild type zebrafish were maintained at CSIR-Institute of Genomics and Integrative Biology as per standard practices described [61]. Tissue isolation was performed by anaesthetizing an adult zebrafish with Tricaine (Sigma, USA). Individual tissues viz. heart, liver, muscle, brain and blood were dissected out and utmost care was taken to ward off contamination to obtain pure homogenous samples for each tissue type. The tissues were washed in PBS several times to clean up any debris. The tissue samples were homogenized in Trizol (Invitrogen, USA). RNA isolation from the homogenized tissue samples was carried out using RNeasy kit (Qiagen, USA) as previously described [32].

**Next generation sequencing and data generation**

Approximately, 5–10 μg of RNA isolated from the individual tissues was used to capture poly-(A) RNA using Sera-Mag oligo (dT) magnetic beads. The captured poly-(A) RNA was fragmented into small pieces of size ranging from 200–500 bp. This size selected RNA was used for cDNA synthesis followed by second strand synthesis using reverse transcriptase and DNA polymerase I respectively. The overhangs at cdNA ends were repaired to blunt ends with the 3’ to 5’ exo-nuclease activity of Klenow enzyme and synthesis activity of T4 DNA Polymerase. To the blunt ends, single ‘A’ base overhang was added by Klenow (3’ to 5’ Exo minus) activity to facilitate specific pairing with manufacturer specified paired end adaptor with a single ‘T’ base overhang. This was followed by the adaptor ligation to the generated cDNA. These ligated A-tail products were run on a 2% agarose gel and fragments corresponding to 300 bp size were purified and selectively enriched by PCR using adaptor specific primers. Quality of the purified library was verified by agarose gel electrophoresis and the concentrations were measured using Qubit (Life Technologies, USA). The RNA libraries were amplified on the Genome Analyzer IIx (GAIIx) flow cell to generate clusters using Illumina’s cBot cluster generation system as per manufacturer specified protocols. Genome Analyzer IIx (GA IIx) sequencing platform from Illumina, USA, was used for sequencing of the RNA libraries. The clusters were sequenced in the GAIIx using sequencing-by-synthesis methodology [34]. High resolution images were captured after every cycle and processed for base calling using Illumina Pipeline software (v1.9). Reads that passed the initial threshold values for quality filter were only used for further analysis. The study accession number (SRA) is PRJNA207719 (SRR891495, SRR891504, SRR891510, SRR891511, SRR891512).

**Assembly of the tissue restricted lncRNome**

The RNA sequencing reads were aligned independently to the zebrafish genome (Zv9) using Bowtie and TopHat (v2.0.3) software (http://tophat.cbcb.umd.edu/). Short read aligner Bowtie was used to align the reads to the exons. These aligned reads were processed by TopHat for demarcating splice junctions between the exons. Further, the mapped reads were assembled into transcripts using Cufflinks software (http://cufflinks.cbcb.umd.edu/), which calculates a transcript’s relative abundance based on the number of reads supporting the transcript, using a reference annotation file. The Cufflink assembler generates the output in the form of FPKM (Fragments Per Kilo base of exons per Million fragments generated) values. The value of FPKM score is directly proportional to the relative abundance of a transcript in a given sample. Transcriptome assembly corresponding to each of the five tissue types was generated. Following this Cuffmerge script (http://cufflinks.cbcb.umd.edu/manual.html#cuffmerge) was used to merge transcriptome data from all the five tissue samples and to filter out reads representing sequencing artifacts owing to the use of random hexamer primers. Next, all the Refseq genes were eliminated and the remaining transcripts formed the corpus of data that was used for downstream analysis. Any lncRNAs that overlapped with Refseq genes were also removed from further analysis. In the next step, transcripts with a length of more than 200 bp were selected and were checked for their coding potential using Coding Potential Calculator software (http://cpc.cbi.pku.edu.cn/), which distinguishes coding and non-coding transcripts with high accuracy [35]. Coding Potential Calculator applies sequence based features to predict the protein-coding potential of transcripts, and has been widely used to discover long non-coding RNAs [62]. Transcripts with a negative score correspond to a non-coding transcript. Transcripts with a score of <−1 score were selected for further analysis. Further, the selected transcripts were checked for open reading frame prediction (ORF) by Getorf software (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html). The transcripts with an ORF length of less than thirty amino acids (as defined for lncRNA) were chosen. The final predicted long non-coding RNAs were aligned back to previously known datasets for developmental stages in zebrafish [24,25]. At this point we also checked for matches to any protein coding isoforms. Those transcripts that had any match with protein coding isoforms were removed from further analysis. The remaining corpus of transcriptome was screened for tissue specific expression using Cuffdiff software (http://cufflinks.cbcb.umd.edu/manual.html#cuffdiff). Cuffdiff software determines the differential expression of transcripts in various tissues. The resulting transcripts were classified as putative tissue restricted lncRNA. The genomic co-ordinates of the identified lncRNA transcripts (BED file) are given in Table S5.

**Quantitative Real-Time PCR (QRT-PCR) assay**

RNA was isolated from the tissues dissected from the adult zebrafish, using RNeasy kit (Qiagen) according to manufacturer’s instructions. cDNA was prepared from 1μg of RNA using Superscript II (Invitrogen, USA). Quantitative Real Time Polymerase Chain Reaction (63) [qRT-PCR] was carried out using Sybr Green mix (Roche, Germany) for detection in Light cycler LC 480 (Roche). The lncRNAs for each tissue were selected based on their FPKM values. Protein coding genes that expressed predominantly in specific tissue types were analyzed in parallel for ensuring purity of the isolated tissues. These protein coding genes were selected on the basis of the in-situ data and publicly available gene expression profiles. Regulatory nervous light chain (cmi2), muscle-related coiled-coil protein b (mucb), midkine a (midk) and T-cell acute lymphocytic leukemia protein 1 (tal 1) were chosen as protein coding gene markers for heart, muscle, brain, liver and blood respectively. The sequences of primers for the protein coding genes and predicted lncRNAs are given in the Table S6.
Whole mount In Situ hybridization (WISH)
Paraformaldehyde-fixed embryos were processed for in situ hybridization according to standard zebrafish protocols (http://zfin.org/ZFIN/Methods/ThissieProtocol.html) [64]. The brain specific lncRNA sequences were amplified from cDNA by PCR using primers (Table S6) and cloned into Topo TA vector (Invitrogen, USA). The lncRNA clones were linearized with NotI and digoxygenin (DIG) labeled at sit probes were generated by in vitro transcription with SP6 or T7 polymerases using DIG RNA Labeling kit (Roche, Germany).

Supporting Information
Figure S1 Differential expression of lncRNA transcripts identified in situ hybridization. Heat maps of 442 lncRNA transcripts across the five tissues viz heart (H), liver (L), muscle (M), brain (Br) and blood (B) are represented. Each individual heat map represents the number of lncRNA transcripts predicted for the corresponding tissue type and its expression levels in the parent tissue vs other tissues based on the FPKM values. The colour key represents the FPKM values in the range of 0 for transcripts with the least expression to 196 for those with the highest expression.

Table S1 of 419 putative lncRNAs that are predicted to express in five tissues of adult zebrafish.

Table S2 of 77 putative lncRNAs that are predicted to have predominant expression restricted to particular tissue type investigated.

References
1. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563. 10.1126/science.1112014 [doi].
2. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiura T, et al. (2004) Complete sequencing and characterization of 2,423 full-length human cDNAs. Nat Genet 36: 40–45. 10.1038/ng1285 [doi]; ng1285 [pii].
3. Pennisi E. (2012) Genomics, ENCODE project writes eulogy for junk DNA. Science 337: 1159, 1161. 337/6099/1159 [pii];10.1126/science.337.6099.1159 [doi].
4. Birney E, Stamatoyannopoulos JA, Datta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799–816. 10.1038/nature05874 [doi].
5. Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs regulators of disease. J Pathol 220: 126–139. 10.1002/path.2630 [doi].
6. Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell 43: 904–914. S1097-2765(11)00636-8 [pii];10.1016/j.molcel.2011.08.018 [doi].
7. Wapinski O, Chang HY (2011) Long non-coding RNAs and human disease. Trends Cell Biol 21: 354–361. S0962-8924(11)00614-4 [pii];10.1016/j.tcb.2011.04.001 [doi].
8. Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, et al. (2006) Functionality of intergenic transcription: an evolutionary comparison. PLoS Genet 2: e171. 06-PGEC-RA-0268R2 [pii];10.1371/journal.pgen.0020171 [doi].
9. Bhartiya D, Kapoor S, Jalali S, Sati S, Kaushik K, et al. (2012) Conceptual approaches for lncRNA drug discovery and future strategies. Expert Opin Drug Discov 7: 503–513. 10.1517/17460441.2012.682055 [doi].
10. Liao Q, Liu C, Yuan X, Kang S, Miao R, et al. (2011) Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 39: 3864–3876. gki3348 [pii];10.1093/nar/gki3348 [doi].
11. Lipovich L, Johnson R, Liu CY (2010) MacroRNA underlogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. Biochim Biophys Acta 1799: 597–615. S1874-gkq1348 [pii].
12. Branman CL, Dees EC, Ingram RS, Tilghman SM (1990) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71: 515–526. 0092-8647(92)90519-4 [pii].
13. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, et al. (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71: 527–542. 0092-8647(92)90520-M [pii].
14. Guttmann M, Garber M, Levin JP, Donovan J, Robinson J, et al. (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28: 510–514. MCB31874 [pii];10.1038/nbt.1633 [doi].
15. Khalil AM, Guttmann M, Haurte M, Garber M, Raj A, et al. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106: 11667–11672. 0904715106 [pii];10.1073/pnas.0904715106 [doi].
16. Lein ES, Hawrylycz MJ, Ao N, Ayres A, Bensinger A, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445: 168–176. nature05453 [pii];10.1038/nature05453 [doi].
17. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, et al. (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316: 1408–1408. 1136341 [pii];10.1126/science.1136341 [doi].
18. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, et al. (2007) GENCODE: producing a reference annotation for ENCODE. Genome Biol 7 Suppl 1: S4-S9. gb-2006-7-s1-s4 [pii];10.1186/gb-2006-7-s1-s4 [doi].
19. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, et al. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 418: 542–547. 0904715106 [pii];10.1038/nature05453 [doi].
20. Guttman M, A夢t I, Garber M, French C, Lin MF, et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458: 223–227. nature07672 [pii];10.1038/nature07672 [doi].
21. Harrow J, Denny F, Frankish A, Reynolds A, Chien CK, et al. (2006) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71: 527–542. 0092-8647(92)90520-M [pii].
22. Jalali S, Bhartiya D, Lalwani MK, Sivasubbu S, Scaria V (2013) Systematic transcriptome wide analysis of lncRNA-miRNA interactions. PLoS One 8: e53823. 10.1371/journal.pone.0053823 [doi];PONE-D-12-06014 [ pii].
23. Kastenholz CA, Scheuermann JC, Vastenhouw NL, Fields PA, et al. (2013) In silico reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28: 510–514. MCB31874 [pii];10.1913/journal.pone.0053823 [doi].
24. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, et al. (2012) Systematic identification of long non-coding RNAs expressed during zebrafish embryogenesis.
40. Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC (2005) Drosophila microRNAs regulate brain morphogenesis in zebrafish. Science 308: 633-638. [pii:10.1126/science.1109820 [doi]]

41. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, et al. (2002) Identification of tissue-specific microRNAs from mouse. Curr Biol 12: 735–739. [pii]:10.1126/curr-biol.12-8.735 [doi]

42. Choppard RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, et al. (2010) Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol 11: R72. gb-2010-11-7-r72 [pii];10.1186/gb-2010-11-7-r72 [doi]

43. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22: 1775–1789. 22/9/1775 [pii];10.1101/gr.135199.113 [doi]

45. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long non-coding RNAs in the mouse brain. Proc Natl Acad Sci U S A 105: 1269–1274. 10.1073/pnas.0706729105 [doi];100563620721210 [pii]

46. Thisse C, Thisse B (2008) High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc 3: 59–69. nprot.2007.514 [pii];10.1038/nprot.2007.514 [doi]

47. Feng J, Bi C, Clark BS, Mady R, Shah P, et al. (2006) The E(spl)-C2 non-coding RNA is transcribed from the Ddx-5/6 ultracorvered region and functions as a Ddx-2 transcriptional coactivator. Genes Dev 20: 1470–1484. 1471-2164-11-548 [pii];10.1101/gad.17446611 [doi]

48. Bagger MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25: 1915–1927. 17446611 [pii];10.1101/gad.17446611 [doi]

49. Li M, Wen S, Guo X, Bai B, Gong Z, et al. (2012) The novel long non-coding RNA CRG regulates Drosophila locomotor behavior. Nucleic Acids Res 40: 11714–11727. gkq493 [pii];10.1093/nar/gkq493 [doi]

50. Li K, Blum Y, Verma A, Liu Z, Pramuk K, et al. (2010) A noncoding antisense RNA in tie-1 locus regulates tie-1 function in vivo. Blood 115: 133–139. blood-2009-04-241810 [pii];10.1182/blood-2009-04-241810 [doi] [pii:10.1093/blood/c102573 [doi];100563620721210 [pii]

51. Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of Caenorhabditis elegans piRNAs. Science 317: 574–578. science.1120932 [pii];10.1126/science.1120932 [doi]

52. Khurana JS, Theurkauf W (2010) piRNAs, transposon silencing, and Drosophila germ line development. J Cell Biol 191: 905–913. jcb.201006034 [pii];10.1083/jcb.201006034 [doi]

53. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Depletion of zebrafish essential and regulatory myosin light chains reduces myotube maturation and heart contractility. PLoS One 7: e52588. 10.1371/journal.pone.0052588 [pii];10.1093/plosone/pone.0052588 [doi]

54. Olson SA (2002) EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite. Brief Bioinform 3: 87–91.

55. Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, et al. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neuronal nuclei. J Cell Sci 120: 2908–2920. jcs.095537 [pii];10.1242/jcs.095537 [doi]

56. Soni K, Choudhary A, Patowary A, Singh AR, Bhatia S, et al. (2013) miR-34 is maternally inherited in Drosophila melanogaster and Danio rerio. Nucleic Acids Res 41: 4470–4480. gkq139 [pii];10.1093/nar/gkq139 [doi]

57. Luo Q, Zhao Y, Zhao YQ, Liu QX, Zhao SQ, et al. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35: W543–W549. 35/suppl_2/W345 [pii];10.1093/nar/gkm091 [doi]

58. Westerfield M (2000) The Zebrafish Book. A guide for the laboratory use of Danio rerio. Univ. of Oregon Press, Eugene.

59. Sleep E, Boue S, Jopling C, Raya M, Raya A, et al. (2010) Transcriptomics reveals global properties and specific subclasses. Genes Dev 25: 1915–1927. 17446611 [pii];10.1101/gad.17446611 [doi]

60. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Depletion of zebrafish essential and regulatory myosin light chains reduces cardiac function through distinct mechanisms. Cardiovasc Res 79: 97–101. cvr073 [pii];10.1093/cvr/cvr073 [doi]

61. Westerfield M (2000) The Zebrafish Book. A guide for the laboratory use of Danio rerio. Univ. of Oregon Press, Eugene.

62. Boerner S, McGinnis KM. (2012) Computational identification and functional validation of putative zebrafish (Danio rerio) microRNAs. BMC Genomics 13: 1–12. 1471-2164-11-548 [pii];10.1186/1471-2164-11-548 [doi]

63. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.