**Year of publication 2021**

Florian Constanty, Alena Shkumatava (2021 Jan 14)

**lncRNAs in development and differentiation: from sequence motifs to functional characterization.**

*Development (Cambridge, England)*: [DOI: dev182741](https://doi.org/dev182741)

**Summary**

The number of long noncoding RNAs (lncRNAs) with characterized developmental and cellular functions continues to increase, but our understanding of the molecular mechanisms underlying lncRNA functions, and how they are dictated by RNA sequences, remains limited. Relatively short, conserved sequence motifs embedded in lncRNA transcripts are often important determinants of lncRNA localization, stability and interactions. Identifying such RNA motifs remains challenging due to the substantial length of lncRNA transcripts and the rapid evolutionary turnover of lncRNA sequences. Nevertheless, the recent discovery of specific RNA elements, together with their experimental interrogation, has enabled the first step in classifying heterogeneous lncRNAs into sub-groups with similar molecular mechanisms and functions. In this Review, we focus on lncRNAs with roles in development, cell differentiation and normal physiology in vertebrates, and we discuss the sequence elements defining their functions. We also summarize progress on the discovery of regulatory RNA sequence elements, as well as their molecular functions and interaction partners.

**Year of publication 2020**

Ansgar Zoch, Tania Auchynnikava, Rebecca V Berrens, Yuka Kabayama, Theresa Schöpp, Madeleine Heep, Lina Vasiliauskaitė, Yuvia A Pérez-Rico, Atlanta G Cook, Alena Shkumatava, Juri Rappasilber, Robin C Allshire, Dónal O’Carroll (2020 Jul 17)

**SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation.**

*Nature*: [DOI: 10.1038/s41586-020-2557-5](https://doi.org/10.1038/s41586-020-2557-5)

**Summary**

In mammals, the acquisition of the germline from the soma provides the germline with an essential challenge: the need to erase and reset genomic methylation. In the male germline, RNA-directed DNA methylation silences young, active transposable elements. The PIWI protein MIWI2 (PIWIL4) and its associated PIWI-interacting RNAs (piRNAs) instruct DNA methylation of transposable elements. piRNAs are proposed to tether MIWI2 to nascent transposable element transcripts; however, the mechanism by which MIWI2 directs the de novo methylation of transposable elements is poorly understood, although central to the immortality of the germline. Here we define the interactome of MIWI2 in mouse fetal gonocytes undergoing de novo genome methylation and identify a previously unknown MIWI2-associated factor, SPOCD1, that is essential for the methylation and silencing of young transposable elements. The loss of Spocd1 in mice results in male-specific infertility but does
not affect either piRNA biogenesis or the localization of MIWI2 to the nucleus. SPOCD1 is a nuclear protein whose expression is restricted to the period of de novo genome methylation. It co-purifies in vivo with DNMT3L and DNMT3A, components of the de novo methylation machinery, as well as with constituents of the NURD and BAF chromatin remodelling complexes. We propose a model whereby tethering of MIWI2 to a nascent transposable element transcript recruits repressive chromatin remodelling activities and the de novo methylation apparatus through SPOCD1. In summary, we have identified a previously unrecognized and essential executor of mammalian piRNA-directed DNA methylation.

Yuvia A Pérez-Rico, Emmanuel Barillot, Alena Shkumatava (2020 Apr 26) Demarcation of Topologically Associating Domains Is Uncoupled from Enriched CTCF Binding in Developing Zebrafish. iScience: 101046: DOI: S2589-0042(20)30231-5

Summary

CCCTC-binding factor (CTCF) is a conserved architectural protein that plays crucial roles in gene regulation and three-dimensional (3D) chromatin organization. To better understand mechanisms and evolution of vertebrate genome organization, we analyzed genome occupancy of CTCF in zebrafish utilizing an endogenously epitope-tagged CTCF knock-in allele. Zebrafish CTCF shares similar facets with its mammalian counterparts, including binding to enhancers, active promoters and repeat elements, and bipartite sequence motifs of its binding sites. However, we found that in vivo CTCF binding is not enriched at boundaries of topologically associating domains (TADs) in developing zebrafish, whereas TAD demarcation by chromatin marks did not differ from mammals. Our data suggest that general mechanisms underlying 3D chromatin organization, and in particular the involvement of CTCF in this process, differ between distant vertebrate species.


Summary

Year of publication 2019


Summary
Strategies for Genetic Inactivation of Long Noncoding RNAs in Zebrafish.

RNA (New York, N.Y.) : DOI : rna.069484.118

Summary

The number of annotated long noncoding RNAs (lncRNAs) continues to grow, however their functional characterization in model organisms has been hampered by the lack of reliable genetic inactivation strategies. While partial or full deletions of lncRNA loci disrupt lncRNA expression, they do not permit the formal association of a phenotype with the encoded transcript. Here, we examined several alternative strategies for generating lncRNA null alleles in zebrafish and found that they often resulted in unpredicted changes to lncRNA expression. Removal of the transcriptional start sites (TSSs) of lncRNA genes resulted in hypomorphic mutants due to the usage of either constitutive or tissue-specific alternative TSSs. Deletions of short, deeply conserved lncRNA regions can also lead to overexpression of truncated transcripts. By contrast, a knock-in of a polyadenylation signal enabled complete inactivation of malat1, the most abundant vertebrate lncRNA. In summary, lncRNA null alleles require extensive in vivo validation and we propose insertion of transcription termination sequences as the most reliable approach to generate lncRNA-deficient zebrafish.

Year of publication 2018


MicroRNA degradation by a conserved target RNA regulates animal behavior

Nat Struct .Mol Biol

Summary

Year of publication 2017


Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution.

EMBO reports : 1231-1247 : DOI : 10.15252/embr.201744059

Summary

Spermatogenesis is associated with major and unique changes to chromosomes and
chromatin. Here, we sought to understand the impact of these changes on spermatogenic transcriptomes. We show that long terminal repeats (LTRs) of specific mouse endogenous retroviruses (ERVs) drive the expression of many long non-coding transcripts (lncRNA). This process occurs post-mitotically predominantly in spermatocytes and round spermatids. We demonstrate that this transposon-driven lncRNA expression is a conserved feature of vertebrate spermatogenesis. We propose that transposon promoters are a mechanism by which the genome can explore novel transcriptional substrates, increasing evolutionary plasticity and allowing for the genesis of novel coding and non-coding genes. Accordingly, we show that a small fraction of these novel ERV-driven transcripts encode short open reading frames that produce detectable peptides. Finally, we find that distinct ERV elements from the same subfamilies act as differentially activated promoters in a tissue-specific context. In summary, we demonstrate that LTRs can act as tissue-specific promoters and contribute to post-mitotic spermatogenic transcriptome diversity.

Year of publication 2016

Yuvia A Pérez Rico, Valentina Boeva, Allison C Mallory, Angelo Bitetti, Sara Majello, Emmanuel Barillot, Alena Shkumatava (2016 Dec 15)

Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes.

Genome research: DOI: gr.203679.115

Summary

Super-enhancers (SEs) are key transcriptional drivers of cellular, developmental and disease states in mammals, yet the conservational and regulatory features of these enhancer elements in non-mammalian vertebrates are unknown. To define SEs in zebrafish and enable sequence and functional comparisons to mouse and human SEs, we used genome-wide histone H3 lysine 27 acetylation (H3K27ac) occupancy as a primary SE delineator. Our study determined the set of SEs in pluripotent state cells and adult zebrafish tissues and revealed both similarities and differences between zebrafish and mammalian SEs. Although the total number of SEs was proportional to the genome size, the genomic distribution of zebrafish SEs differed from that of the mammalian SEs. Despite the evolutionary distance separating zebrafish and mammals and the low overall SE sequence conservation, ~42% of zebrafish SEs were located in close proximity to orthologs that also were associated with SEs in mouse and human. Compared to their non-associated counterparts, higher sequence conservation was revealed for those SEs that have maintained orthologous gene associations. Functional dissection of two of these SEs identified conserved sequence elements and tissue-specific expression patterns, while chromatin accessibility analyses predicted transcription factors governing the function of pluripotent state zebrafish SEs. Our zebrafish annotations and comparative studies show the extent of SE usage and their conservation across vertebrates, permitting future gene regulatory studies in several tissues.
Baptiste Renaud, Sylvie Schneider-Maunoury, Alena Shkumatava, Lydia Teboul, Jim Kent, Jean-Stephane Joly, Jean-Paul Concordet (2016 Jul 7)

**Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR.**


**Summary**

The success of the CRISPR/Cas9 genome editing technique depends on the choice of the guide RNA sequence, which is facilitated by various websites. Despite the importance and popularity of these algorithms, it is unclear to which extent their predictions are in agreement with actual measurements.

---

Allison C Mallory, Alena Shkumatava (2015 Mar 28)

**LncRNAs in vertebrates: advances and challenges.**

*Biochimie* : 3-14 : [DOI : 10.1016/j.biochi.2015.03.014]

**Summary**

Beyond the handful of classic and well-characterized long noncoding RNAs (lncRNAs), more recently, hundreds of thousands of lncRNAs have been identified in multiple species including bacteria, plants and vertebrates, and the number of newly annotated lncRNAs continues to increase as more transcriptomes are analyzed. In vertebrates, the expression of many lncRNAs is highly regulated, displaying discrete temporal and spatial expression patterns, suggesting roles in a wide range of developmental processes and setting them apart from classic housekeeping ncRNAs. In addition, the deregulation of a subset of these lncRNAs has been linked to the development of several diseases, including cancers, as well as developmental anomalies. However, the majority of vertebrate lncRNA functions remain enigmatic. As such, a major task at hand is to decipher the biological roles of lncRNAs and uncover the regulatory networks upon which they impinge. This review focuses on our emerging understanding of lncRNAs in vertebrate animals, highlighting some recent advances in their functional analyses across several species and emphasizing the current challenges researchers face to characterize lncRNAs and identify their in vivo functions.

---

Igor Ulitsky, Alena Shkumatava, Calvin H Jan, Alexander O Subtelny, David Koppstein, George W Bell, Hazel Sive, David P Bartel (2012 Jun 23)

**Extensive alternative polyadenylation during zebrafish development.**

*Genome research* : 2054-66 : [DOI : 10.1101/gr.139733.112]
Summary

The post-transcriptional fate of messenger RNAs (mRNAs) is largely dictated by their 3′ untranslated regions (3′ UTRs), which are defined by cleavage and polyadenylation (CPA) of pre-mRNAs. We used poly(A)-position profiling by sequencing (3P-seq) to map poly(A) sites at eight developmental stages and tissues in the zebrafish. Analysis of over 60 million 3P-seq reads substantially increased and improved existing 3′ UTR annotations, resulting in confidently identified 3′ UTRs for >79% of the annotated protein-coding genes in zebrafish. mRNAs from most zebrafish genes undergo alternative CPA, with those from more than a thousand genes using different dominant 3′ UTRs at different stages. These included one of the poly(A) polymerase genes, for which alternative CPA reinforces its repression in the ovary. 3′ UTRs tend to be shortest in the ovaries and longest in the brain. Isoforms with some of the shortest 3′ UTRs are highly expressed in the ovary, yet absent in the maternally contributed RNAs of the embryo, perhaps because their 3′ UTRs are too short to accommodate a uridine-rich motif required for stability of the maternal mRNA. At 2 h post-fertilization, thousands of unique poly(A) sites appear at locations lacking a typical polyadenylation signal, which suggests a wave of widespread cytoplasmic polyadenylation of mRNA degradation intermediates. Our insights into the identities, formation, and evolution of zebrafish 3′ UTRs provide a resource for studying gene regulation during vertebrate development.

Year of publication 2011

Igor Ulitsky, Alena Shkumatava, Calvin H Jan, Hazel Sive, David P Bartel (2011 Jul 28)
Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution.

Summary

Thousands of long intervening noncoding RNAs (lincRNAs) have been identified in mammals. To better understand the evolution and functions of these enigmatic RNAs, we used chromatin marks, poly(A)-site mapping and RNA-Seq data to identify more than 550 distinct lincRNAs in zebrafish. Although these shared many characteristics with mammalian lincRNAs, only 29 had detectable sequence similarity with putative mammalian orthologs, typically restricted to a single short region of high conservation. Other lincRNAs had conserved genomic locations without detectable sequence conservation. Antisense reagents targeting conserved regions of two zebrafish lincRNAs caused developmental defects. Reagents targeting splice sites caused the same defects and were rescued by adding either the mature lincRNA or its human or mouse ortholog. Our study provides a roadmap for identification and analysis of lincRNAs in model organisms and shows that lincRNAs play crucial biological roles during embryonic development with functionality conserved despite limited sequence conservation.
Year of publication 2010


Expanding the microRNA targeting code: functional sites with centered pairing.
Molecular cell : 789-802 : DOI : 10.1016/j.molcel.2010.06.005

Summary

Most metazoan microRNA (miRNA) target sites have perfect pairing to the seed region, located near the miRNA 5’ end. Although pairing to the 3’ region sometimes supplements seed matches or compensates for mismatches, pairing to the central region has been known to function only at rare sites that impart Argonaute-catalyzed mRNA cleavage. Here, we present “centered sites,” a class of miRNA target sites that lack both perfect seed pairing and 3’-compensatory pairing and instead have 11-12 contiguous Watson-Crick pairs to the center of the miRNA. Although centered sites can impart mRNA cleavage in vitro (in elevated Mg(2+)), in cells they repress protein output without consequential Argonaute-catalyzed cleavage. Our study also identified extensively paired sites that are cleavage substrates in cultured cells and human brain. This expanded repertoire of cleavage targets and the identification of the centered site type help explain why central regions of many miRNAs are evolutionarily conserved.

Year of publication 2009

Alena Shkumatava, Alexander Stark, Hazel Sive, David P Bartel (2009 Feb 26)

Coherent but overlapping expression of microRNAs and their targets during vertebrate development.
Genes & development : 466-81 : DOI : 10.1101/gad.1745709

Summary

MicroRNAs (miRNAs) are small noncoding RNAs that direct post-transcriptional repression of protein-coding genes. In vertebrates, each highly conserved miRNA typically regulates hundreds of target mRNAs. However, the precise relationship between expression of the miRNAs and that of their targets has remained unclear, in part because of the scarcity of quantitative expression data at cellular resolution. Here we report quantitative analyses of mRNA levels in miRNA-expressing cells of the zebrafish embryo, capturing entire miRNA expression domains, purified to cellular resolution using fluorescent-activated cell sorting (FACS). Focus was on regulation by miR-206 and miR-133 in the developing somites and miR-124 in the developing central nervous system. Comparison of wild-type embryos and those lacking miRNAs revealed predicted targets that responded to the miRNAs and distinguished miRNA-mediated mRNA destabilization from other regulatory effects. For all three miRNAs examined, expression of the miRNAs and that of their predicted targets usually overlapped. A few targets were expressed at higher levels in miRNA-expressing cells than in the rest of the embryo, demonstrating that miRNA-mediated repression can act in opposition
to other regulatory processes. However, for most targets expression was lower in miRNA-expressing cells than in the rest of the embryo, indicating that miRNAs usually operate in concert with the other regulatory machinery of the cell.