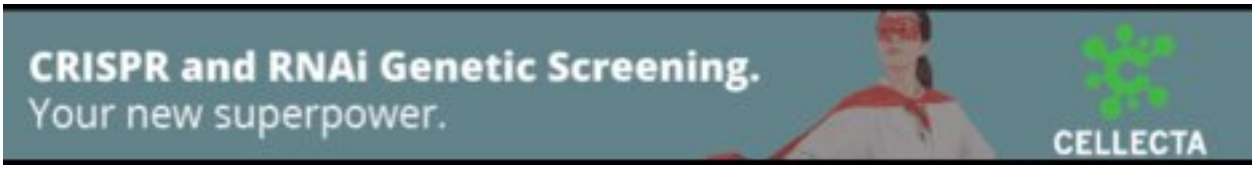
Antisense RNAs during early vertebrate development are divided in groups with distinct features

Sanjana Pillay, Hazuki Takahashi, Piero Carninci, et al.

Genome Res. published online April 1, 2021
Access the most recent version at doi:10.1101/gr.262964.120



Antisense RNAs during early vertebrate development are divided in groups with distinct features

Sanjana Pillay[1], Hazuki Takahashi[2], Piero Carninci[2,3], Aditi Kanhere[4*]

[1] Department of Cell, Developmental and Regenerative Biology, Mount Sinai School of Medicine, New York, NY, United States

[2] Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, 230-0045, Japan

[3] Fondazione Human Technopole - Via Cristina Belgioioso 171 - 20157 Milano Italy

[4] Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, L69 3GE, United Kingdom*

* To whom correspondence should be addressed.

Tel: +44 (0)151 795 1292

Email: a.kanhere@liverpool.ac.uk

Present Address: Dr Aditi Kanhere, Department of Molecular Physiology and Cell Signaling, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 3BX, United Kingdom.

Running Title: Antisense RNAs in Zebrafish development

Key Words: Natural Antisense Transcripts, Development

ABSTRACT

Long non-coding RNAs or lncRNAs are a class of non-protein-coding RNAs that are >200 nucleotides in length. Almost 50% of lncRNAs during zebrafish development are transcribed in an antisense direction to a protein-coding gene. However, the role of these Natural Antisense Transcripts or NATs during development remains enigmatic. To understand NATs in early vertebrate development, we took a computational biology approach and analyzed existing as well as novel datasets. Our analysis indicates that zebrafish NATs can be divided into two major classes based on their co-expression patterns with respect to the overlapping protein-coding genes. Group-1 NATs have characteristics similar to maternally deposited RNAs in that their levels decrease as development progresses. Group-1 NAT levels are negatively correlated with that of overlapping sense-strand protein-coding genes. Group-2 NATs, on the other hand, are co-expressed with overlapping protein-coding genes. In contrast to group-1, which is enriched in genes involved in developmental pathways, group-2 protein-coding genes are enriched in house-keeping functions. Group-1 NATs also show larger overlap and higher complementarity with the sense-strand mRNAs as compared to other NATs. In addition, our transcriptomics data, quantifying RNA levels from cytoplasmic and nuclear compartments, indicates that group-1 NATs are more populated in the cytosol. Based on their expression pattern, cytosolic nature and their higher complementarity to the overlapping developmental mRNAs, we speculate that group-1 NATs function post-transcriptionally to silence spurious expression of developmental genes.

Introduction:

Natural Antisense Transcripts or NATs are non-coding RNAs that are transcribed in an antisense direction to the overlapping genes. NATs are prevalent in eukaryotic genomes, ranging from yeast to humans. Estimates suggest that, in the human genome, as many as ~38% of protein-coding genes show evidence of antisense transcription (Katayama et al. 2005; He et al. 2008; Balbin et al. 2015). Although the importance of the majority of NATs still remains enigmatic, expression of sense-antisense transcript pairs is often linked, either positively or negatively, much more than expected by chance (Chen et al. 2005). While the possibility that the antisense transcription is a mere consequence of transcriptional status of the overlapping protein-coding gene cannot be ruled out, a number of studies also suggest that there is a regulatory relationship between the expression of overlapping sense-antisense pairs. Individual examples suggest that NATs can regulate the overlapping protein-coding partner using variety of transcriptional and post-transcriptional mechanisms, such as influencing chromatin landscape (Nagano et al. 2008; Wang et al. 2011; Magistri et al. 2012; Fatica and Bozzoni 2014) , genomic imprinting (Berteaux et al. 2008; Zhao et al. 2008; Fatica and Bozzoni 2014), RNA processing (Mandal et al. 2013), alternative splicing (Morrissy et al. 2011), RNA stability and the rate of translation (Beltran et al. 2008; Ebralidze et al. 2008). NATs such as *Kcnq1ot1* (Pandey et al. 2008; Fatica and Bozzoni 2014), *Airn* (Nagano et al. 2008) and *HOTTIP* (Wang et al. 2011) regulate transcription of the overlapping genes by influencing chromatin environment at the genomic locus. On the other hand, NATs like *HTTAS* and *qrf* use transcriptional interference to downregulate the expression of overlapping genes (Chung et al. 2011; Xue et al. 2015). Some NATs can also function post-transcriptionally e.g. by increasing the stability of overlapping sense strand mRNAs by protecting them from RNA-degrading machineries such as micro-RNAs and ribonucleases. Antisense RNAs overlapping *Sirt1* and *BACE1* genes are good examples of this mechanism (Faghihi et al. 2008; Li et al. 2018). On the other hand, some NATs are shown to decrease the stability of overlapping sense-strand mRNA. In these cases, by virtue of sequence complementarity, NAT and sense strand mRNA can hybridize resulting in inhibition of mRNA translation and leading to its degradation (Villegas and Zaphiropoulos, 2015, Faghihi and Wahlestedt, 2009). Finally, NATs like *TRalpha2 NAT* participate in regulating alternative splicing of overlapping *TRalpha2* mRNA (Hastings et al. 2000).

Recent studies suggest that NATs play an important role during development. For example, NATs, such as *Kcnq1ot1*(Pandey et al. 2008) and *Airn* (Nagano et al. 2008; Fatica and Bozzoni 2014) are needed for imprinting which is crucial for early vertebrate development (Wang and Chang 2011). NATs like *HOTTIP* and *HOTAIRM1* are involved in regulation of spatiotemporal expression of developmentally important *HOXA* genes (Wang and Chang 2011; Wang et al. 2011). In addition, studies in animal models indicate that NATs play a role in embryonic development in vertebrates. Previous transcriptomics studies have identified large number of lncRNAs during embryonic development in vertebrates such as zebrafish (Ulitsky et al. 2011; Pauli et al. 2012; Haque et al. 2014). Up to 50% of maternally deposited RNAs in zebrafish embryo are NATs (Pauli et al. 2011; Pauli et al. 2012). A transcriptomics study by Pauli *et al.,* e.g., identified 1133 lncRNAs across eight stages of zebrafish development, out of these 397 were intergenic lncRNAs, 184 intronic overlapping lncRNAs and the rest 566 were classified as exonic overlapping NATs (Pauli et al. 2012). The abundance of NATs during early development combined with functional studies on selected NATs further points to their importance in early embryonic development (Li et al. 2010; Pauli et al. 2012; Wei et al. 2014). However, for the majority of NATs, functional mechanism and relationship to the overlapping protein-coding transcript during development remains largely unclear (de Hoon et al. 2015).

NATs and their regulatory mechanisms during development are likely to be conserved across different vertebrates. *spi1b* NAT and *tie1* AS are perfect examples of this (Li et al. 2010; Wei et al. 2014). *Tie1* AS is transcribed in zebrafish, mouse as well as humans in the antisense direction to the *tie1* gene (Li et al. 2010). Tie1 protein is a tyrosine kinase receptor for angioproteins and is essential for vascular development in vertebrates. The NAT, *tie1* AS, binds to *tie1* mRNA through RNA:RNA hybridization and downregulates it, thus, resulting in a loss of the protein. This mechanism is evolutionary conserved as imbalance in regulation of Tie1 protein by *tie1AS* results in vascular defects in zebrafish as well as human (Li et al. 2010). Another example is that of *spi1b* NAT which regulates the expression of transcription factor Spib that regulates myeloid and lymphoid cell development in both, zebrafish as well as human (Wei et al. 2014). Similar to *tie1* AS, *spi1b* NAT also downregulates *spi1b* mRNA by forming RNA-RNA duplex and preventing its translation.

Zebrafish is one of the popular animal models that is routinely used to understand early vertebrate development (Mork and Crump 2015; Drummond and Davidson 2016; Jung et al. 2017). In zebrafish,

embryonic development starts by fertilization of externally laid eggs and spans across a period of 3 days post-fertilization (dpf). Initially, the embryo undergoes 10 rapid and asynchronous cell divisions followed by more lengthened cell cycles. In all vertebrates, the embryo is in a transcriptionally inactive state during initial period of cell divisions. As a result, during this period when the zygotic genome is inactive, early development of the embryo is completely dependent on maternally provided products (Tadros and Lipshitz 2009). As development progresses, the transcription of the zygotic genome is activated and simultaneous clearance of maternal RNAs and proteins leads to their replacement with newly synthesized zygotic RNAs. This process is called maternal-to-zygotic transition (MZT). In zebrafish, MZT coincides with mid-blastula transition (MBT) and occurs at 3 hours post-fertilization (hpf) at the 1000-cell stage (Lee et al. 2014). The availability of large amount of transcriptomics data during developmental stages spanning MZT makes zebrafish a good model to further interrogate relationship between NATs and their sense protein-coding partners during early development.

Unlike many other classes of ncRNAs, NATs show varied relationship with overlapping protein-coding genes. They also differ in size, their genomic organization, conservation, expression pattern and cellular localization. The lack of any unifying characteristics or relationship between NATs and their overlapping protein-coding genes, makes it difficult to assess their role in regulating protein-coding genes. This is also one of the reasons why the role of NATs during early vertebrate development remains enigmatic. Here, we present results of transcriptomics analyses of different groups of NATs during early zebrafish developmental stages spanning MZT. The aim of this study is to identify distinct features of each group of NATs which will be useful in speculating their role in development and in predicting their function in regulating sense-strand gene expression.

Results

Comparison of NATs with long intergenic RNAs during early development.

First, we sought to understand if NATs have features that are distinct from other long ncRNAs. We compared them to the other major category of lncRNAs, long intergenic ncRNAs (lincRNAs). LincRNAs, unlike NATs, are expressed from genomic loci that are away from the protein-coding regions of the genome. We first analyzed expression levels of NATs and lincRNAs in the 8 stages of early development (2-4 cell, 1000 cell, dome, shield, bud, 24hpf, 48hpf and 120hpf). For this purpose, RNA-seq reads from previously published study (Pauli et al. 2012) were remapped to the zebrafish

genome and normalized abundance of annotated non-coding RNAs and protein-coding genes was calculated.

We first compared the percentage of annotated NATs and lincRNAs that were expressed (>1 fpkm) during different stages of development (Figure 1A). The percentage of NATs detected during maternal stages (2-4 cell and 1000 cell) was higher than percentage lincRNAs (Figure 1A). However, after zygotic genome activation, the percentage of NATs was lower than lincRNAs. Before MZT, NAT percentages (~13% in 2-4 cell stage and 15% in 1000 cell stage) were significantly higher than that of lincRNAs (0.2%-3.5%, p-value < 10^{-4}). On the other hand, after MZT, the percentage of NATs was significantly lower compared to lincRNAs (Figure 1A). This indicates that a higher percentage of NATs is deposited among maternal RNAs than lincRNAs suggesting possible relevance of NATs in the pre-MZT stages.

In addition, we also examined expression levels of NATs and lincRNAs. In all the eight stages that were considered in this study, the average RNA levels of NATs were lower (p-value < 0.01) as compared to lincRNAs (Figure 1B). This suggests that, although types of NAT species present in the pre-MZT was more than lincRNAs (Figure 1A), they were on an average less abundant than lincRNAs (Figure 1B).

We also inspected the stability and stage specificity of these two classes of lncRNAs using UpSet diagrams (Lex et al. 2014). This allowed us to visualize the frequency of NATs and lincRNAs that were present in consecutive stages of development (Figure 1C, Supplemental Table S1). This analysis indicated that the percentage of NATs stable during first four stages spanning MZT was higher than lincRNAs (7.5% vs 3.9%; Chi-square p-value < 0.01). On the other hand, the percentage of NATs present in the six post-MZT stages was almost half that of the lincRNAs (11.6% vs 21%; Chi-square p-value < 0.0001). On the other hand, the combined frequency of NATs (1.5%) present or expressed in stage specific manner, i.e. occurring only in one particular stage, was very similar to the lincRNAs (1.1%) (Figure 1C). This suggests that maternally deposited NATs are probably more stable as compared to maternally deposited lincRNAs. In order to assess the significance of these observations, we compared these results to the stability patterns of mRNAs during development (Supplemental Fig S1A). In contrast to NATs and lincRNAs, lesser percentages of mRNAs were stable specifically in either pre-MZT or post-MZT (only 2% in first four stages and only 4% were stable

in the six stages post-MZT). On the other hand, higher percentage of mRNAs were stable all throughout the eight stages (~25%) as compared to NATs (~9%) and lincRNAs (~12%).

Can these differences between NATs and lincRNAs be related to the functions of protein-coding genes they are associated with? In order to answer this, we compared functions of the protein-coding genes overlapping all NATs annotated in zebrafish to protein-coding genes adjacent to all lincRNAs (Figure 1D). A Gene Ontology analysis showed that NATs significantly overlapped genes coding for DNA-binding proteins such as transcription factors (p-value <= 5.40×10^{-9}). On the other hand, the top functional category for protein-coding genes adjacent to lincRNAs (within ±2kb) was prostaglandin transmembrane transporter activity (Figure 1D, below). We also saw some enrichment of transcription factor related genes, in case of lincRNAs, however, this was much lesser than that seen in the case of NATs (transcription factor activity, p-value <= 7.60×10^{-3}). The observation, that transcription factors (TFs) were significantly overrepresented among the genes overlapping NATs, might suggest that antisense transcription is a general feature of genes coding TFs. However, only 4% of all TF genes in zebrafish overlap NATs, making it unlikely that it is a general feature of TF genes (Supplemental Fig S1B). To examine if this association of NATs and lincRNAs to the functions of associated genes is evolutionarily conserved, we carried out a similar analysis in human and mouse (Figure 1D). In mouse, NATs were also associated with genes with DNA-binding functions, however, in humans, we did not see any significant enrichment (Figure 1D), presumably indicating evolutionary changes in association of NATs and protein-coding genes.

Lastly, to understand if they are distinct in any other way, we also compared other properties NATs and lincRNAs in zebrafish, mouse and human. In zebrafish and mouse, lincRNAs showed a significantly higher exon count and transcript length as compared to NATs (Supplemental Fig S1C). However, in human, NATs and lincRNAs were similar in exon count and transcript length (Supplemental Fig S1C). In all three species, NATs and lincRNAs, showed significant differences in the conservation level. To rule out the possibility that this is because the of the bias introduced by differences in the genomic position of intragenic NATs as compared to intergenic lincRNAs, we corrected conservation levels by taking into consideration only those NAT exons which overlap <10% of any sense strand protein-coding exon and NAT promoters which are away from promoters of protein-coding genes and do not overlap protein-coding exons. After this correction, the conservation of NATs and lincRNAs did not show any significant difference (Supplemental Fig S1D).

NATs can be grouped into classes with distinct relation to the overlapping protein-coding genes

Although a large of number of NATs are present before MZT in zebrafish, the role of NATs in gene expression regulation during early zebrafish development remains unclear. In addition, the relevance of expression changes in NATs to early embryonic development has not been fully understood. Therefore, as a first step, we sought to analyse the expression patterns of NATs during early development. Although not a proof of functionality, their co-expression patterns vis-à-vis the overlapping protein-coding genes and other genomic features can provide insights into the early events during development.

For this analysis, we selected NATs which had minimum 10% overlap with the overlapping sense-strand coding transcript (Figure 2A). Our analysis based on Ensembl annotations showed that there are 1482 such protein-coding/NAT pairs. To understand the relation between protein-coding/NAT pairs, we carried out a correlation calculation between normalized RNA levels of each protein-coding/NAT pair across the 8 developmental stages. Out of the 1482 protein-coding and NAT pairs, 60 pairs did not express at all in any of the developmental stages we considered and, therefore, were excluded (Figure 2A). Among remaining 1422 protein-coding/NAT pairs, 696 pairs were negatively correlated, 580 pairs were positively correlated and the rest 146 showed no correlation. Pairs with significant correlations (p-value < 0.05 and $r > 0.71$) were retained for further analysis. As a result, 127 anti-correlated (group-1) and 326 positively correlated protein-coding/NAT pairs (group-2) were shortlisted (Figure 2A). The 146 pairs which did not show any correlation were used as a control set (group-3).

First, we examined if the NATs in these three groups are predominantly present in either the pre- or the post-MZT stages. In order to understand this, we plotted average abundance of NATs as well as mRNAs in the three groups. We found that group-1 NATs were more abundant in the pre-MZT stages and their levels decreased in the post-MZT stages. On the other hand, the protein-coding transcript levels in group-1 increased post-MZT (Figure 2B). Thus, suggesting that the NATs in group-1 category were predominantly maternally deposited and, like other maternal transcripts, probably degraded upon zygotic genome activation at 10hpf. In contrast, group-2 and group-3 NATs did not show any particular tendency to be abundant either pre-MZT or post-MZT (Figure 2B). On an

average, NATs and mRNAs belonging to the positively correlated group-2 and uncorrelated group-3 showed both maternal (0.75hpf to 10hpf) and zygotic (24hpf onwards) presence. Interrogation of individual RNA levels in group-1 and group-2 confirmed that the average expression patterns reflect correlations between individual protein-coding/NAT pairs (Figure 2B, Supplemental Fig S2, S3).

These results were comparable to that obtained by another frequently used RNA-seq analysis software, STAR (Dobin and Gingeras 2015). A comparison showed that, independent of the alignment protocols, ~80% of NAT/protein-coding pairs showed similar correlation and are classified in same groups as our original pipeline (Supplemental Fig S4A). In addition, the relationship between expression patterns of NATs and their overlapping protein-coding partners in the three groups was very similar using the two alignment protocols (Figure 2B and Supplemental Fig S4B). To rule out the possibility that co-expression relation between NAT/protein-coding pairs is a mere consequence of expression changes that occur during MZT and is not affected if the NATs were to be randomly paired with non-overlapping mRNAs, we paired NATs with neighbouring non-overlapping protein-coding transcripts as well as with protein-coding transcripts expressed elsewhere in the zebrafish genome. In both the scenarios, ~64-80% of NATs showed no correlation and around 8-18% showed opposite relation with randomly selected mRNAs as compared to the overlapping protein-coding transcript partner (Supplemental Fig S4C, D). This suggests that the correlations in the RNA levels between NATs and their overlapping protein-coding partners is not random and is independent of genome-wide gene expression changes linked to maternal to zygotic transition.

If group-1 NATs were maternally deposited, then like other maternal RNAs, this group of NATs should show distinct stability pre-MZT and post-MZT. Therefore, using UpSet plots, we also analyzed if NATs in the three groups showed differences in the stability during MZT (Figure 2C, Supplemental Table S2). This analysis showed that large number of the NATs in group-1 category were present in the early stages (2-4 Cell, 1000 Cell and Dome; 13%; Chi-square p-value < 0.005). The percentage of group-1 NATs present only in a single stage was very low suggesting that they are not highly stage-specific. In contrast, the highest percentage for group-2 NATs was observed in single stages (14% for 120hpf and 11% for 1000 cell). Moreover, stability pattern of group-2 was very comparable to the control group-3 NATs. Thus, suggesting that the group-1 NATs are more stable and are present through multiple stages spanning MZT as compared to the group-2 NATs and the group-3 NATs.

These results show that during early vertebrate development two very different classes of NATs are present and they show distinct relationship to their overlapping protein-coding genes.

NATs in different groups are distinct in their organisation vis-à-vis the overlapping genes

Examples supporting positive as well as negative regulation of protein-coding genes by NATs are described in literature (Wang and Chang 2011; Wight and Werner 2013). In some cases, NATs are shown to have no effect on the function of the overlapping protein-coding partner or might instead have an effect on a gene at distant loci (Wight and Werner 2013). This is reflected in our analysis which shows that antisense RNA-protein pairs can be divided into either negative (group-1), positive (group-2) or no correlation groups (group-3). Besides displaying different co-expression relations, NATs are also known to be organized in different manner vis-à-vis the overlapping gene. Well-studied NATs are shown to be transcribed near 5'-end (HOTTIP), near 3'-end (HOTAIRM1) or from the gene body of the sense strand gene (BACE1-AS, Kcnq1ot1). Understanding how different categories of NATs are organized in relation to overlapping protein-coding genes would help in postulating reasons behind positive, negative and no-correlation patterns between NATs/protein-coding pairs as well as gaining insights into functional mechanisms of NATs. However, the link between co-expression patterns between NAT/protein-coding gene pairs and their genomic positioning relative to each other has not been clearly understood.

We first calculated the extent of the overlap between NAT and overlapping protein-coding gene pairs in the three categories. We defined NAT/protein-coding overlap as the percent length of the protein-coding gene that is completely covered by the NAT in the antisense direction (Figure 3A, Supplemental Fig S5A). NAT/protein-coding transcript pairs in group-1 showed significantly greater overlap (average = 13003bp, p-value < 0.01) as compared to the pairs in group-2 (average = 4548bp) and no correlation group (average = 6722bp). To further understand the nature of NAT/protein-coding transcript overlap, we also calculated TSS-TSS distances i.e. genomic distances between the transcription start sites (TSS) of NAT and protein-coding genes in a pair (Figure 3B). TSS-TSS distance distribution in the three groups of NATs was distinct from one another (Figure 3B, Supplemental Fig S5B). The average distance between TSS-TSS distances in group-1 NAT and overlapping protein-coding TSS (log(distance) = 4.6) was much greater than that in group-2 (log(distance) = 3.6) and group-3 (log(distance) = 4.0). However, when we calculated TES-TSS

distances i.e genomic distances between transcription end sites (TES) of NATs to the protein-coding genes in the pair, we found that TES-TSS distances in group-1 (average log(distance) = 4.1) and group-2 (average log(distance) = 3.7) were significantly different (p-value < 0.01; Figure 3C, Supplemental Fig S5C). However, group-1 was similar to the control group-3 (log(distance) = 4.0) in this aspect.

We further interrogated whether NATs in different groups also start and end distinctly vis-à-vis genomic features *viz.* exons, introns, intergenic regions, 5' UTRs and 3' UTRs of protein-coding genes on the opposite strand (Figure 3D, E). The majority of NATs (~40%) in group-1 started in intergenic regions in contrast to the group-2 and group-3 NATs which mainly started within the introns (50% and 48.7% respectively) of the protein-coding gene on opposite strand (Figure 3D). On the contrary, the TESs of a large majority of NATs in all the three categories were in the introns of protein-coding genes on the sense strand (Figure 3E).

This shows that group-1 NATs start much further away from protein-coding TSS as compared to group-2 NATs. However, the group-2 NATs end nearer to the protein-coding TSS as compared to group-1 and group-3 NATs. This might suggest that the average length of NATs in group-1 might be more than other groups. However, transcript lengths NATs in group-1 was only marginally greater than group-2 and group-3 (Supplemental Fig S5D). The number of exons in group-1 NATs was significantly higher than other two groups (Supplemental Fig S5E).

These observations can be summarized in a model showing that NATs in group-1 start in intergenic regions and span large percentage of the length of overlapping protein-coding gene (Figure 3F). Group-2 and Group-3 NATs, on the other hand, start in introns of the overlapping protein-partner. Group-2 NATs end in introns nearer to the TSS of overlapping protein-coding gene (Figure 3F).

Group-1 NATs overlap developmental genes while group-2 NATs overlap housekeeping genes

Given the distinct expression patterns of NATs and protein-coding transcripts in the three groups (Figure 2B), a pertinent question would be if the protein-coding genes in these groups are distinct in their biological and cellular functions. To address this question, a Gene Ontology (GO) analysis was carried out on the protein-coding transcripts in the three groups. We found that the genes in the three groups were enriched in distinct molecular functions (Figure 4A). The transcripts belonging to group-1 were highly enriched in sequence-specific DNA-binding proteins (p-value < 3.70×10^{-9}),

developmental genes (p-value 1.40×10^{-7}) and proteins involved in transcription regulation (p-value < 8.90×10^{-6}). The genes involved in transcription factor activity were also enriched in group-2 (p-value < 1.60×10^{-4}), however, less significantly than the group-1 protein-coding genes (p-value < 3.70×10^{-9}). In addition, group-2 genes also showed enrichment of housekeeping functions related to metabolism and signalling processes. In contrast, the transcripts in the non-correlated group-3 did not show significant enrichment of any particular category.

To confirm the association of group-1 NATs with developmental genes, we sought to investigate if these genes also display other characteristics of developmental genes. It is well-documented that, during early stages of development, genes involved in developmental pathways are repressed by polycomb group of proteins. Polycomb Repressive Complex-2 catalyzes trimethylation of the lysine 27 on the histone H3 (H3K27me3) at the promoters of developmental genes (Aloia et al. 2013; Deevy and Bracken 2019). Given the enrichment of developmental genes in the group-1 (Figure 4A), we expected that the promoters of protein-coding genes in this group would show histone mark patterns, such as enrichment of H3K27me3, that are normally seen at developmental genes. We mined genome-wide Chromatin Immunoprecipitation data (ChIP-seq) for three different histone modification marks in the dome stage of zebrafish development. In addition to H3K27me3 modification, we also analysed H3K4me3 (Histone 3 lysine 4 trimethylation) and H3K27ac (Histone 3 lysine27 acetylation) marks. H3K27me3 mark is largely enriched at transcriptionally repressed genes while H3K4me3 and H3K27ac marks are usually enriched at transcriptionally active genes (Vastenhouw et al. 2010; Black et al. 2012; Zhang et al. 2014). We checked the enrichment of these marks at the TSS of protein-coding transcripts in group-1, group-2 and the control group-3. We observed distinct patterns of histone modifications across TSS overlapping protein-coding genes in the different groups. As predicted, protein-coding genes in group-1 showed higher enrichment of repressive H3K27me3 mark (Figure 4B, Supplemental Table S3). H3K27ac and H3K4me3 marks did not show as significant enrichment at the TSS of protein-coding genes in the group-1 as group-2. The relative lack of H3K4me3 and H3K27ac marks and the enrichment of H3K27me3 mark suggests that group-1 protein-coding partners are not transcribed. In contrast, the histone modification marks at the group-2 genes were distinct. The protein-coding genes in group-2 showed much lower enrichment for H3K27me3 repressive mark (Figure 4B, Supplemental Table S3). However, unlike group-1, there was a significant enrichment of H3K27ac and H3K4me3 marks across their TSSs (Figure 4B right hand

panels). The histone modification patterns at group-2 promoters are very similar to that observed at average protein-coding genes (Supplemental Fig S6A), again reflecting that the group-2 probably represents genes with housekeeping functions as shown in our analysis of functional categories (Figure 4A). On the other hand, group-3 protein-coding genes did not show any distinctive enrichment for any histone modifications (Supplemental Fig S6B), indicating they do not display a particular pattern of expression.

Evolutionary conservation of genes or genomic elements is often used as a proxy for their functional importance. Developmental genes are often more conserved than average genes. We assessed evolutionary conservation of the gene promoters, 3' UTR and 5' UTR in the two categories (Figure 4C, Supplemental Fig S6C & S6D). As would be expected for developmental genes, the group-1 showed higher conservation score at the promoters (*t*-test; p-value < 0.05) compared to the group-2 and the control group-3 (Figure 4C). Since the polycomb group of genes have GC-rich promoters, we further analyzed the GC content at the promoters of different groups. Our analysis suggests that the group-1 protein-coding genes have promoters with higher G+C content as compared to group-2 genes and control group-3 (Figure 4D). The G+C content at 5' UTR and 3' UTRs, however, was not different between the protein-coding genes in the three classes (Supplemental Fig S6E & F).

These observations support that the protein-coding genes in group-1 are polycomb targeted developmental genes. On the other hand, group-2 and non-correlated groups were likely house-keeping genes suggesting a different mechanism of regulation.

NAT expression is regulated in a group-specific manner

Given that we observed differences in protein-coding genes in different groups, we also sought to understand the differences in the transcription regulation of NATs. First, we plotted histone modifications and chromatin accessibility at TSS of different category of NATs. As observed in the case of protein-coding genes, distinct patterns of histone modifications were observed across TSS of NATs in the different groups (Figure 5A, Supplemental Table S3). Unlike group-1 protein-coding genes, NATs in the group-1 showed little or no enrichment of H3K27me3 mark around the TSS indicating that they were probably not repressed by polycomb machinery. In addition, H3K27ac and H3K4me3 marks were less enriched at group-1 NAT TSSs as compared to group-2. This is in accordance with our prediction that group-1 NATs are maternally deposited rather than zygotically

transcribed. This was also reflected in open chromatin status as measured by ATAC-seq method (Figure 5B), which showed that the group-1 NAT promoters do not show presence of open chromatin supporting absence of transcription. In contrast to group-1, group-2 NATs showed enrichment of H3K27ac and H3K4me3 marks across their TSSs (Figure 5A right hand panels). This was supported by ATAC-seq data showing more open chromatin than group-1. On the other hand, group-3 NATs did not show enrichment for any histone modifications (Supplemental Fig S6B top panels), indicating they do not display a particular pattern of expression.

The distinct expression patterns of NATs and functional roles of overlapping protein-coding genes in the three categories point to differences in NAT transcription regulation. To further dissect the nature of transcription regulation of NATs, we analysed the sequence motifs associated with the promoters of NATs in these three categories. As expected, NATs in the three categories showed enrichment of distinct transcription factor (TF) motifs (Figure 5C). We observed that promoters of group-1 NATs showed enrichment for TF motifs such as Mtf1 and Runx3 which are known to be essential for normal vertebrate development (Burns et al. 2002; Oates and Ho 2002; Hogstrand et al. 2008). The promoters of group-2 category NATs were enriched instead in Her1 and Irf-6 transcription factor motifs that are also involved in zebrafish development (Figure 5C).

Another indication of transcription regulation is provided by the promoter width (Haberle et al. 2014). The existence of sharp TSS in case of maternal transcripts and broad TSS in case of zygotic transcripts has been reported (Haberle et al. 2014; Haberle et al. 2015). The promoter width can be effectively calculated by mapping TSS using Cap Analysis Gene Expression or CAGE analysis. Using CAGE changes in TSS and its relative usage can be measured at single nucleotide resolution (Kodzius et al. 2006). CAGE gives us the information for the start sites of capped RNAs which in turn can be used as an indicator of promoter organization. The width of TSSs can be calculated based on the mapped CAGE tags(Haberle et al. 2015). For this analysis, we deep sequenced the CAGE tags from six stages of early zebrafish development and calculated the promoter width (Figure 5D). In this analysis, group-2 NATs appeared to have much broader promoters as compared to promoters of group-1 NATs and the control group-3. This observation further supports maternal nature of group-1 NATs and housekeeping nature of group-2 NATs.

NATs show differences in cellular localization

Subcellular localization of RNAs can give an indication regarding their functional mechanism(Cabili et al. 2015). Studies on yeast (Long et al. 1997) and flies (Johnstone and Lasko 2001) have indicated very specific subcellular localization of mRNAs to be important in yeast and fly development. It can be envisioned that the specific subcellular localization of NATs during zebrafish embryogenesis might also be linked to their cellular function. For example, lncRNAs involved in regulation of chromatin-modification and transcription are localised in the nucleus and lncRNAs involved in post-transcriptional regulation are localised in cytoplasm (Cabili et al. 2015). Keeping this in mind and to get an idea regarding the functional mechanisms of group-1 and group-2 NATs, we sought to identify the subcellular localization of different category of NATs during early stages of zebrafish embryos. To achieve this, we carried out deep sequencing of RNAs from cytoplasmic and nuclear fractions collected from different stages of development. An RNA was categorized as more enriched in the nuclear fraction, if its abundance was 1.5-fold or more when compared to the cytosolic fraction. Similarly, an RNA was considered more enriched in the cytosol, if its abundance was 1.5-fold or more in the cytoplasm than that in the nuclear fraction. RNAs with abundance between 0 to 1.5-fold were considered to be equally present in both cellular fractions.

To verify if our protocol correctly identifies subcellular localisation of RNAs, we first analyzed the localization of mRNAs. The mRNAs were identified as predominantly equally present in both cellular fractions or cytosolic showing that our protocol for identifying subcellular localization of RNAs worked well (Supplemental Fig S7A).

We then analyzed cellular localization of NATs in the three groups. We ranked NATs based on their abundance and then calculated the $\log_2$ ratio to determine their localization. Group-1 NATs were more enriched in the cytosolic fraction during pre-MZT stages further supporting maternal deposition and become more enriched in the nuclear fraction in the shield stage presumably indicating beginning of zygotic transcription (Figure 6A). The majority of group-2 NATs were equally enriched in both the fractions at 64 cell stage, however, after 64-cell stage, they were more enriched in the nucleus (Figure 6A). The control group NATs were in the cytoplasm at 64-cell stage but did not show any particular enrichment after 64-cell stage (Supplemental Fig S7B).

Previous studies have indicated that cytosolic NATs that downregulate the overlapping protein-coding genes, act by forming an RNA-RNA hybridization leading post-transcriptional changes in mRNA stability (Villegas and Zaphiropoulos 2015). Given the cytosolic localization of group-1 NATs, a post-

transcriptional mechanism involving RNA-RNA hybridization can be envisioned. Given the negative correlation of group-1 NATs with the levels of their overlapping mRNAs, it can be speculated that they form RNA:RNA hybrids with the sense-strand mRNA partners and decrease their stability. To check this possibility, we assessed the sequence complementarity between NAT transcripts and corresponding mRNAs in the pairs in the three groups (Figure 6B). In this analysis, group-1 NATs showed higher complementarity as reflected in the alignment score as compared to the control group-3 (Figure 6B; p-value < 0.0001). The cytosolic localization and higher complementarity indicate a post-transcriptional mechanism for group-1 NATs for regulation of mRNA expression. We speculate that group-1 NATs, possibly through RNA:RNA interaction, help to repress developmental gene expression (Figure 6C).

DISCUSSION

Antisense transcription is a common feature among all types of organisms ranging from bacteria to mammals with up to 40% of human transcriptome being predicted to show evidence of antisense transcription (Chen et al. 2004; Katayama et al. 2005; Zhang et al. 2006; He et al. 2008; Balbin et al. 2015). A large proportion of protein-coding mRNAs are part of sense-antisense transcript pairs where the antisense transcript is a lncRNA (Chen et al. 2004; Katayama et al. 2005; Engstrom et al. 2006; Derrien et al. 2012; Hon et al. 2017; Zucchelli et al. 2019). In many cases, NATs regulate the overlapping protein-coding gene on sense strand (Villegas and Zaphiropoulos 2015). Despite the widespread nature of antisense transcription, its functional significance remains poorly characterized. The functional relation of NATs to the overlapping sense strand protein-coding genes still remains confusing. One of the reasons behind this is NATs display positive as well as negative correlation to the expression of their sense strand partners (Faghihi and Wahlestedt 2009; Pelechano and Steinmetz 2013; Wight and Werner 2013; Khorkova et al. 2014; Rosikiewicz and Makalowska 2016). In addition, a number of NATs do not affect the expression of overlapping genes. In some cases, the antisense transcription can result merely as a consequence of opening of chromatin due to transcription of the sense strand gene or as a by-product of enhancer activity (Struhl 2007; Onodera et al. 2012). In other instances, they are shown to regulate protein-coding genes at other non-overlapping loci (Roberts and Morris 2013; Rosikiewicz and Makalowska 2016). As a result, it has been difficult to propose a unifying mechanism behind their function.

Here we focused on analysing the relation between NATs and their overlapping protein-coding genes during zebrafish development. Previous studies took an approach where they first divided NATs according to their genomic location vis-à-vis their overlapping protein-coding gene pair and then analyzed the co-expression pattern between NAT/protein-coding pairs (Balbin et al. 2015). Instead, we focused on first grouping NATs according to the co-expression patterns of NAT/protein-coding pairs across eight stages of zebrafish development and then analyzing their characteristics.

Different sets of NATs/protein-coding pairs show distinctive characteristics. The group-1 protein-coding genes are mainly developmental genes while group-2 protein-coding genes show enrichment for house-keeping functions (Figure 4). On the other hand, group-3 did not show enrichment for any particular functional pathways. Group-1 and group-2 NATs were also distinctive in terms of positioning of transcription start and end sites as well as promoter structure indicating that their expression is regulated in distinct manner (Figure 3, 5). This is reflected in their expression pattern during development indicating group-1 NATs are maternally deposited while group-2 RNAs seem to be transcribed from zygotic genome (Figure 2). Our RNA-seq data shows that group-1 RNAs are cytosolic which is also typical of maternal RNAs (Figure 6). Group-2 RNAs on the other hand show more nuclear presence, possibly because they are transcribed from zygotic genome (Figure 6).

Previous studies on NATs have shown they can regulate either transcription or post-transcriptional processing (Faghihi and Wahlestedt 2009; Pelechano and Steinmetz 2013; Wight and Werner 2013; Khorkova et al. 2014; Rosikiewicz and Makalowska 2016). Given the predominantly cytosolic nature of group-1 RNAs, it is unlikely that they are involved in transcriptional regulation. Previous publications have indicated that antisense transcription from intragenic enhancers can have an attenuating effect on transcription of the overlapping protein-coding genes (Cinghu et al. 2017). However, unlike group-1 NATs, these intragenic enhancers were predominantly found to be enriched in chromatin fraction (Cinghu et al. 2017). Besides in our analysis (Figure 5), neither group-1 nor group-2 NATs display classic enhancer signature, enrichment of H3K27ac mark and lack of H3K4me3 mark, indicating that the majority of group-1 and group-2 NATs are not by-products of enhancer activity.

However, based on their cytosolic localization and opposite expression pattern as compared to overlapping developmental mRNAs, we can speculate that group-1 NATs function by changing the stability of developmental mRNAs during early stages of zebrafish development (Figure 6C). They

might be involved in decreasing mRNA stability through formation of RNA:RNA hybrids as reported for other cytosolic NATs. This is corroborated by the higher level of complementarity in the group-1 antisense RNA-mRNA pairs which alludes to a possibility of hybridization between antisense RNA and mRNA. Examples show that formation of RNA duplex between lncRNAs and mRNAs can attract mRNA-degradation machinery as seen in case of STAU1-mediated mRNA degradation(Gong and Maquat 2011b). In addition, lncRNAs are also implicated in generating endogenous siRNAs which can lead to mRNA degradation (Wilusz et al. 2009; Gong and Maquat 2011a). One of the questions that is highlighted from our analysis is the reason behind the higher percentage of group-1 NATs among maternally deposited RNAs. It is possible that they do not have any significant role in regulating the developmental genes. However, their expression specifically from developmental gene loci and their distinct features as compared to the other two groups of NATs point to their significance. It can be speculated that group-1 NATs are needed to curtail unwarranted expression of developmental genes during early stages before MZT which can be detrimental to normal development. However, functional characterization and further experimental validation is needed to confirm these observations.

In addition to functional differences between group-1 and group-2 mRNAs, we also see differences in the genomic location of group-1 and group-2 NATs vis-à-vis overlapping mRNAs. A majority of group-1 NATs start in an intergenic region, away from the TSS of their overlapping protein-coding partner and generally display larger overlap with the sense gene. In contrast, group-2 NATs start much closer to sense gene TSS and they appear to be in a head-to-head or an embedded configuration with respect to the sense protein-coding gene**.** We speculate that specific configuration plays a role in deciding the relationship between NATs and their overlapping protein-coding genes. This kind of relationship has been reported in previous studies and individual examples which showed that head-to-head configuration is associated with positive co-expression patterns between antisense/protein-coding pairs (Balbin et al. 2015). Distinct features observed in negatively correlated group-1 and positively correlated group-2 can be useful in predicting NAT-protein-coding co-expression in future. Further experimental studies are however needed to understand the exact nature of the relationship between NATs and their protein-coding pair and also to verify if these observations are broadly applicable to the development of other vertebrates.

Lastly, for this study we have generated a large transcriptomics dataset (RNA-seq and CAGE-seq) which can provide information regarding RNA enrichment in different cellular compartments. In addition to NATs, these data will be useful to assess roles of other ncRNAs such as enhancer RNAs and circular RNAs during development.

METHODS

Analysis of RNA-Sequencing data

Expression levels of NATs and mRNA was calculated using RNA-sequencing data for 8 zebrafish developmental stages (Pauli et al., 2012). FASTQ files corresponding to raw RNA-sequencing reads for 8 zebrafish developmental stages was downloaded from SRA database (https://www.ncbi.nlm.nih.gov/sra; GEO accession no: GSE32900). They were quality checked and trimmed using FastQC (Andrews 2010) and Trimmomatic v0.27 respectively (Bolger et al. 2014). Trimmed and quality filtered sequencing reads were then mapped back to the Zv9 or danRer7 genome assembly of zebrafish using two different alignment pipelines. In case of the first pipeline, the reads were mapped to the genome using TopHat suit (Kim et al. 2013) and then the aligned reads were used for transcript assembly and expression levels using Cufflinks v2.2.1 (Trapnell et al. 2012). The expression levels were measured by Cufflinks as Fragments Per Kilobase Per Million (FPKM) reads for each transcript. For second pipeline, the raw sequencing reads were also mapped to the Zv9 genome assembly of zebrafish using another alignment programme STAR v2.6.0a (Dobin and Gingeras 2015) and generated alignment files were used to assemble transcripts using programme StringTie (Pertea et al. 2015). The StringTie programme was run using the -B and -b options. The output of StringTie was used for differential expression analysis using the programme Ballgown (Pertea et al. 2016). The transcript and gene levels we calculated as Tags Per Millions (TPM) for different stages of development. The outputs of Cufflinks and StringTie files were used for all our downstream analysis. The stage specific abundance of NATs and lincRNAs was plotted and visualized using UpSet plots in R (Lex et al. 2014). For visualisation purposes, strand specific expression tracks were generated using “genomecov” command in BEDTools package (Quinlan and Hall 2010).

Comparison of LincRNAs and NAT characteristics

All the analysis of lincRNAs and NATs was carried on Ensembl annotations for Zv9, mm10 and hg19 assemblies for zebrafish, mouse and human genomes respectively. Given that this analysis relies on average and comparative characteristics, the genome versions (for example, GRCh38 or Zv10) would not significantly affect the conclusions. LincRNAs and NATs were considered present if their expression level is >=1 FPKM which was calculated as described above. The transcript lengths and exon counts were calculated based on Ensembl annotations.

Identification of protein-coding and NAT pairs

We first paired NATs with overlapping protein-coding mRNA transcripts. For this analysis, we considered 47279 protein-coding transcripts and 682 NATs that are annotated by Ensembl for zebrafish genome version Zv9(Zerbino et al. 2018; Yates et al. 2020). Using the genomic coordinates, we identified all the pairs where the antisense gene overlapped at least 10% of the protein-coding gene. Based on this criterion, we identified 1482 pairs NAT/protein-coding pairs. The percent length of mRNA transcript that was overlapped with antisense RNA was calculated using BEDTools suit's "Intersect" command using -S option (Quinlan and Hall 2010).

Functional analysis of NAT – mRNA pairs

The expression levels of protein-coding and NATs calculated using Cufflinks and StringTie were first normalized with respect to the maximum level of each RNA among the 8 stages. The normalized values were then used to compute linear correlation between each pair. The pairs were categorized based on whether the pairs correlated negatively (group-1), positively (group-2) or did not show any correlation at all (group-3). Only pairs with significant correlation were retained (p-value < 0.05, r >=0.70 for sample size N=8). The analysis of functional annotations was carried out using Database for Annotation, Visualisation and Integrated Discovery (DAVID 6.8) software for functional annotation of these transcripts (Huang da et al. 2009a; Huang da et al. 2009b). For this analysis, categories related to biological process, molecular function and cellular components were considered and a background of all zebrafish genes was used. R package, ggplot2 (Wickham 2016), was used for plotting all of our violin and box plots. The geom_histogram() function in ggplot2 was used to plot the histograms for showing the overlap region and the distance between the TSS of protein-coding genes and the antisense TES. Heatmaps related to NATs and their overlapping protein-coding genes were also plotted using R (R Core Team 2019) and deepTools2.0 (Ramirez et al. 2016). Base-specific

conservation scores (Vertebrate Cons) corresponding to each NAT and mRNA sequence were downloaded from UCSC Genome Browser (Karolchik et al. 2004). The conservation values for NATs were corrected by only considering promoters and exons that overlap < 10% with the protein-coding exons on sense strand.

The G+C content track for zebrafish (danRer7.gc5Base.wig) from UCSC Genome Browser was used to calculate the G+C content at the promoters (Haeussler et al. 2019). We also used EMBOSS geecee analysis to calculate the frequency of G and C nucleotide in the sequences (Rice et al. 2000). The EMBOSS Needle analysis was run with the default options for pairwise sequence alignment of NATs and mRNA sequence. The sequence for mRNA and NATs were downloaded from the UCSC Genome Browser.

ChIP-seq and ATAC-seq analysis

The raw Chromatin ImmunoPrecipitation followed by sequencing (ChIP-seq) data for different histone modifications H3K27me3 (DCD003227SQ), H3K4me3 (DCD003231SQ) and H3K27ac (DCD003287SQ) corresponding to Dome (4hpf) stage of zebrafish development (Vastenhouw et al. 2010; Bogdanovic et al. 2012; Zhang et al. 2014) were obtained from the DANIO-CODE repository at https://danio-code.zfin.org. The ChIP sequencing reads were downloaded in FASTQ format and mapped to Zv9 version of Zebrafish genome using Bowtie2(Langmead and Salzberg 2012). The mapped reads were analyzed using ChIP-seq analysis software HOMER to create tag directories and to annotate peaks using the makeTagDirectory and findPeaks tools (-style factor and -o auto options). The processed ChIP-sequencing data was used to plot heatmaps with the help of deepTools2 (Ramirez et al. 2016) and to visualize the enrichment of different histone modifications within ± 2kb distance of transcription start sites of protein-coding transcripts as well as NATs in the three different groups of NAT/protein-coding pairs. The statistical significance of histone modification enrichments around transcription start sites of different NAT groups was calculated using BEDTools Fisher programme (Quinlan and Hall 2010).

The FASTQ files for ATAC-seq data for different stages (DCD003157SQ, DCD003146SQ and DCD003127SQ) of zebrafish development (Vastenhouw et al. 2010; Bogdanovic et al. 2012; Zhang et al. 2014) were downloaded from the DANIO-CODE repository at https://danio-code.zfin.org. The raw reads were mapped to the Zv9 genome using Bowtie 2(Langmead and Salzberg 2012). The BAM files

generated were used to get bigWig files from deepTools using BAMCoverage (–normalizeUsing BPM option). The bigWig files were mapped to visualize open chromatin within ± 4kb distance of transcription start sites for both NATs and overlapping mRNAs.

Nuclear & cytosolic fractionation of zebrafish embryos in different stages of development

Wild-type male and female zebrafish (AB-strain) were set up in breeding tanks overnight and on the next day, to ascertain their synchronisation, the fertilized eggs were collected with minimum delay (<10 min). About 100-500 (depending upon the stage) embryos were collected for each developmental stage (32 cells, 64 cells, 256 cells, 512 cells, high, shield). The embryos were treated with Pronase (Sigma-Aldrich) to remove the chorion. 1mL of Pronase working solution (1mg/mL) was added to 2-3 ml of fish water containing embryos. Dechorionated embryos were then transferred to a 1.5mL microcentrifuge tube and then washed twice with 1 X PBS (Phosphate buffer solution, Thermo Fisher Scientific). RLN buffer (Sigma-Aldrich) was used to disrupt the yolk sac releasing the cells into the buffer. This was then incubated on ice for 5 mins and centrifuged. The resulting supernatant was collected and labelled as the cytosolic fraction. The pellet was washed twice with 200 µL of RLN buffer and finally collected and labelled as the nuclear fraction. RNA was extracted from the nuclear and cytosolic fractions of the embryos was extracted using the RNeasy mini kit from QIAGEN and was DNase (Sigma-Aldrich) treated to remove traces of genomic DNA. The quality of the RNAs extracted was detected on the RNA TapeStation using Agilent High Sensitivity RNA Screen Tape assay. Sequencing libraries were prepared using the TruSeq Stranded Total RNA library Prep kit with the Illumina Ribo-Zero rRNA removal kit (Human-Mouse-Rat). The quality check, library preparation and sequencing were carried out by University of Birmingham's Genomics facility using Illumina NextSeq 500.

The RNA sequencing reads obtained were mapped to the Zv9 genome assembly of zebrafish using STAR v2.6.0a (Dobin and Gingeras 2015) as explained earlier. Aligned reads were assembled into transcripts using StringTie (Pertea et al. 2015). Further Ballgown program (Pertea et al. 2016) was used to produce differential expression values (in transcript per million or TPM) in different stages of development. These expression values were used for all our downstream analysis.

CAGE sequencing & analysis

CAGE library preparation was carried out using a modified cap trapping protocol (Carninci et al. 1996) for low quantity samples or LQ-ssCAGE protocol (Takahashi H 2020). RNA was extracted using the RNeasy mini kit from QIAGEN with 1µg of RNA per sample as starting material. We divided each sample into four parts, 250ng of RNA to be pooled later after cDNA synthesis. Raw tags from CAGE sequencing were mapped using STAR aligner and the resulting BAM files were used in the bioconductor package CAGEr for downstream analysis including quality filtering, normalization, removal of the 5' end G nucleotide that was added during the CAGE protocol (Haberle et al. 2015). The visualization of strand specific CAGE tracks was carried out using "genomecov" command in BEDTools package (Quinlan and Hall 2010)

Statistical Analysis

Significance levels (p-values) and sample sizes are provided in the text, figure legends or indicated on the figures. Statistical analysis was performed using GraphPad Prism or as a part of the computational tools used. For measuring statistical significance, un-paired *t*-tests or chi-square tests were used to calculate p-value. P-values of <0.05 were considered significant.

DATA ACCESS

The RNA-seq and CAGE-seq data generated in this study have been submitted in the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE143208 and GSE144040 respectively.

ACKNOWLEDGEMENTS

We thank Prof Ferenc Mueller and his lab for useful discussions and help with zebrafish methods. We are grateful to all members of ZENCODE ITN for critical comments on the work. We also thank Nishiyori Hiromi and Miki Kojima from Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences, Japan for their help with LQ-CAGE library preparation. This work was supported by the European Commission's Marie Curie H2020 ITN funding.

DISCLOSURE DECLARATION

Authors declare no conflict of interests.

REFERENCES

Aloia L, Di Stefano B, Di Croce L. 2013. Polycomb complexes in stem cells and embryonic development. *Development* **140**: 2525-2534.

Andrews S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Balbin OA, Malik R, Dhanasekaran SM, Prensner JR, Cao X, Wu YM, Robinson D, Wang R, Chen G, Beer DG et al. 2015. The landscape of antisense gene expression in human cancers. *Genome Res* **25**: 1068-1079.

Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, Bonilla F, de Herreros AG. 2008. A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* **22**: 756-769.

Berteaux N, Aptel N, Cathala G, Genton C, Coll J, Daccache A, Spruyt N, Hondermarck H, Dugimont T, Curgy JJ et al. 2008. A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Mol Cell Biol* **28**: 6731-6745.

Black JC, Van Rechem C, Whetstine JR. 2012. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**: 491-507.

Bogdanovic O, Fernandez-Minan A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gomez-Skarmeta JL. 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* **22**: 2043-2053.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Burns CE, DeBlasio T, Zhou Y, Zhang J, Zon L, Nimer SD. 2002. Isolation and characterization of runxa and runxb, zebrafish members of the runt family of transcriptional regulators. *Exp Hematol* **30**: 1381-1389.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**: 20.

Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, Kamiya M, Shibata K, Sasaki N, Izawa M et al. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**: 327-336.

Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. 2005. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends Genet* **21**: 326-329.

Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD. 2004. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* **32**: 4812-4820.

Chung DW, Rudnicki DD, Yu L, Margolis RL. 2011. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum Mol Genet* **20**: 3467-3477.

Cinghu S, Yang P, Kosak JP, Conway AE, Kumar D, Oldfield AJ, Adelman K, Jothi R. 2017. Intragenic Enhancers Attenuate Host Gene Expression. *Mol Cell* **68**: 104-117 e106.

de Hoon M, Shin JW, Carninci P. 2015. Paradigm shifts in genomics through the FANTOM projects. *Mamm Genome* **26**: 391-402.

Deevy O, Bracken AP. 2019. PRC2 functions in development and congenital disorders. *Development* **146**.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG 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.

Dobin A, Gingeras TR. 2015. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* **51**: 11 14 11-19.

Drummond IA, Davidson AJ. 2016. Zebrafish kidney development. *Methods Cell Biol* **134**: 391-429.

Ebralidze AK, Guibal FC, Steidl U, Zhang P, Lee S, Bartholdy B, Jorda MA, Petkova V, Rosenbauer F, Huang G et al. 2008. PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev* **22**: 2085-2092.

Engstrom PG, Suzuki H, Ninomiya N, Akalin A, Sessa L, Lavorgna G, Brozzi A, Luzi L, Tan SL, Yang L et al. 2006. Complex Loci in human and mouse genomes. *PLoS Genet* **2**: e47.

Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G, 3rd, Kenny PJ, Wahlestedt C. 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* **14**: 723-730.

Faghihi MA, Wahlestedt C. 2009. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* **10**: 637-643.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **15**: 7-21.

Gong C, Maquat LE. 2011a. "Alu"strious long ncRNAs and their role in shortening mRNA half-lives. *Cell Cycle* **10**: 1882-1883.

Gong C, Maquat LE. 2011b. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**: 284-288.

Haberle V, Forrest AR, Hayashizaki Y, Carninci P, Lenhard B. 2015. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res* **43**: e51.

Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, Gehrig J, Dong X, Akalin A, Suzuki AM et al. 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**: 381-385.

Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN et al. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res* **47**: D853-D858.

Haque S, Kaushik K, Leonard VE, Kapoor S, Sivadas A, Joshi A, Scaria V, Sivasubbu S. 2014. Short stories on zebrafish long noncoding RNAs. *Zebrafish* **11**: 499-508.

Hastings ML, Ingle HA, Lazar MA, Munroe SH. 2000. Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J Biol Chem* **275**: 11507-11513.

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. 2008. The antisense transcriptomes of human cells. *Science* **322**: 1855-1857.

Hogstrand C, Zheng D, Feeney G, Cunningham P, Kille P. 2008. Zinc-controlled gene expression by metal-regulatory transcription factor 1 (MTF1) in a model vertebrate, the zebrafish. *Biochem Soc Trans* **36**: 1252-1257.

Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J et al. 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**: 199-204.

Huang da W, Sherman BT, Lempicki RA. 2009a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, Lempicki RA. 2009b. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**: Unit 13 11.

Johnstone O, Lasko P. 2001. Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annu Rev Genet* **35**: 365-406.

Jung HM, Castranova D, Swift MR, Pham VN, Venero Galanternik M, Isogai S, Butler MG, Mulligan TS, Weinstein BM. 2017. Development of the larval lymphatic system in zebrafish. *Development* **144**: 2070-2081.

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**: D493-496.

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564-1566.

Khorkova O, Myers AJ, Hsiao J, Wahlestedt C. 2014. Natural antisense transcripts. *Hum Mol Genet* **23**: R54-63.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M et al. 2006. CAGE: cap analysis of gene expression. *Nat Methods* **3**: 211-222.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Lee MT, Bonneau AR, Giraldez AJ. 2014. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30**: 581-613.

Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* **20**: 1983-1992.

Li B, Hu Y, Li X, Jin G, Chen X, Chen G, Chen Y, Huang S, Liao W, Liao Y et al. 2018. Sirt1 Antisense Long Noncoding RNA Promotes Cardiomyocyte Proliferation by Enhancing the Stability of Sirt1. *J Am Heart Assoc* **7**: e009700.

Li K, Blum Y, Verma A, Liu Z, Pramanik K, Leigh NR, Chun CZ, Samant GV, Zhao B, Garnaas MK et al. 2010. A noncoding antisense RNA in tie-1 locus regulates tie-1 function in vivo. *Blood* **115**: 133-139.

Long RM, Singer RH, Meng X, Gonzalez I, Nasmyth K, Jansen RP. 1997. Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. *Science* **277**: 383-387.

Magistri M, Faghihi MA, St Laurent G, 3rd, Wahlestedt C. 2012. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* **28**: 389-396.

Mandal AK, Pandey R, Jha V, Mukerji M. 2013. Transcriptome-wide expansion of non-coding regulatory switches: evidence from co-occurrence of Alu exonization, antisense and editing. *Nucleic Acids Res* **41**: 2121-2137.

Mork L, Crump G. 2015. Zebrafish Craniofacial Development: A Window into Early Patterning. *Curr Top Dev Biol* **115**: 235-269.

Morrissy AS, Griffith M, Marra MA. 2011. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res* **21**: 1203-1212.

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**: 1717-1720.

Oates AC, Ho RK. 2002. Hairy/E(spl)-related (Her) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling pathway in the formation of anterior segmental boundaries in the zebrafish. *Development* **129**: 2929-2946.

Onodera CS, Underwood JG, Katzman S, Jacobs F, Greenberg D, Salama SR, Haussler D. 2012. Gene isoform specificity through enhancer-associated antisense transcription. *PLoS One* **7**: e43511.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**: 232-246.

Pauli A, Rinn JL, Schier AF. 2011. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* **12**: 136-149.

Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A et al. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* **22**: 577-591.

Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nat Rev Genet* **14**: 880-893.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**: 1650-1667.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290-295.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

R Core Team. 2019. R: A Language and Environment for Statistical Computing https://www.R-project.org.

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160-165.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.

Roberts TC, Morris KV. 2013. Not so pseudo anymore: pseudogenes as therapeutic targets. *Pharmacogenomics* **14**: 2023-2034.

Rosikiewicz W, Makalowska I. 2016. Biological functions of natural antisense transcripts. *Acta Biochim Pol* **63**: 665-673.

Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**: 103-105.

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. *Development* **136**: 3033-3042.

Takahashi H N-SH, Carninci P. 2020. Low Quantity single strand CAGE protocol. *protocolsio* doi:dx.doi.org/10.17504/protocols.io.bbwkipcw.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562-578.

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**: 1537-1550.

Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, Rinn J, Schier AF. 2010. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**: 922-926.

Villegas VE, Zaphiropoulos PG. 2015. Neighboring gene regulation by antisense long non-coding RNAs. *Int J Mol Sci* **16**: 3251-3266.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**: 904-914.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**: 120-124.

Wei N, Pang W, Wang Y, Xiong Y, Xu R, Wu W, Zhao C, Yang G. 2014. Knockdown of PU.1 mRNA and AS lncRNA regulates expression of immune-related genes in zebrafish Danio rerio. *Dev Comp Immunol* **44**: 315-319.

Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wight M, Werner A. 2013. The functions of natural antisense transcripts. *Essays Biochem* **54**: 91-101.

Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23**: 1494-1504.

Xue Y, Shen SQ, Corbo JC, Kefalov VJ. 2015. Circadian and light-driven regulation of rod dark adaptation. *Sci Rep* **5**: 17616.

Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R et al. 2020. Ensembl 2020. *Nucleic Acids Res* **48**: D682-D688.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG et al. 2018. Ensembl 2018. *Nucleic Acids Res* **46**: D754-D761.

Zhang Y, Liu XS, Liu QR, Wei L. 2006. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res* **34**: 3465-3475.

Zhang Y, Vastenhouw NL, Feng J, Fu K, Wang C, Ge Y, Pauli A, van Hummelen P, Schier AF, Liu XS. 2014. Canonical nucleosome organization at promoters forms during genome activation. *Genome Res* **24**: 260-266.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756.

Zucchelli S, Fedele S, Vatta P, Calligaris R, Heutink P, Rizzu P, Itoh M, Persichetti F, Santoro C, Kawaji H et al. 2019. Antisense Transcription in Loci Associated to Hereditary Neurodegenerative Diseases. *Mol Neurobiol* **56**: 5392-5415.

FIGURE LEGENDS

Figure 1: LncRNA dynamics during zebrafish development (A) A bar plot showing the percentage of antisense and lincRNAs present during 8 stages of zebrafish development. The significance values are as * P<0.05; **P<0.005; ***P<0.001; ****P<0.0001; ns: non-significant (Chi-square test). **(B)** A boxplot of abundance levels of NATs and lincRNAs across 8 developmental stages. Each box shows median abundance value (as horizontal lines) and extend from 25th to 75th percentile values for each group. The outliers are shown as dots. The significance values are as * P<0.05; **P<0.005; ***P<0.001; ****P<0.0001 ; ns: non-significant (unpaired, two-tailed *t*-test). **(C)** UpSet diagrams depicting the percentage of NATs (left) and lincRNAs (right) that are common or unique in the 8 stages of zebrafish development. The percentage of lncRNAs is shown on y-axis and the stages in which the lncRNA is present is shown below x-axis. Filled circles represent the stages under consideration for the above bar. **(D)** Bar plots showing Gene Ontology terms associated with the mRNAs that overlap NATs (top) vs mRNAs neighboring to lincRNAs (bottom) in zebrafish, mouse and human. The -log(p-val) values for gene enrichment are plotted on the x-axis and Gene Ontology terms are shown on the y-axis.

Figure 2. Distinct classes of NATs during zebrafish development (A) Overview of the pipeline undertaken to identify and categorize the NATs based on their expression correlation with overlapping mRNAs. There are 127 anti-correlated (group-1), 326 positively correlated (group-2) and 146 no correlation (group-3) NAT-mRNA pairs. **(B)** Line plots showing the average abundance of NATs and overlapping mRNAs in the three categories (negatively correlated, positively correlated and no correlation) during development. **(C)** UpSet plots illustrating the frequency of NATs in negatively correlated group-1 (yellow), positively correlated group-2 (blue) and no correlation group-3 (grey) that are common or unique during zebrafish development. Filled circles represent the stages under consideration for the above bar.

Figure 3: The NATs in three categories are differently transcribed vis-à-vis overlapping protein-coding genes (A) A bar chart showing the overlap region between NATs and overlapping mRNAs in group-1 (yellow), group-2 (blue) and group-3 (grey) **(B)** A bar chart showing the distance between the TSS of NATs and TSS of overlapping mRNAs (log of distance in bp) in the group-1 (yellow), group-2 (blue) and group-3 (grey) **(C)** A bar chart showing the distance between the TES of NATs and TSS of overlapping mRNA (log) in the group-1 (yellow), group-2 (blue) and group-3 (grey). For **(A), (B) and (C)** the schematic above each figure represents how the values were calculated. The bar values represent the mean ± standard deviation. The scatter shows individual values. The significance values are as * P<0.05; **P<0.005; ***P<0.001; ****P<0.0001; ns: non-significant (unpaired, two-tailed *t*-test). (**D)** A bar plot of the percentage of NAT TSSs overlapping different genomic features (exon, intron, intergenic region, 5' UTR and 3' UTR) on the opposite strand. **(E)** A bar plot of the percentage of NAT TESs overlapping different genomic features (exon, intron, intergenic region, 5' UTR and 3' UTR) on the opposite strand. **(F)** Schematic diagram representing

how three different groups of NATs are distributed and localized in the genome with respect to the overlapping protein-coding genes.

Figure 4: Negatively correlated NATs overlap protein-coding genes involved in vertebrate development. (A) Bar plots depicting the Gene Ontology terms associated with mRNAs in the three categories of NAT-protein-coding pairs. In each case, the x-axis displays - log(p-value) for Gene Ontology terms that are shown on the y-axis. **(B)** Heatmaps displaying the distribution of ChIP-seq reads for H3K27ac (red), H3K4me3 (red) and H3K27me3 (blue) histone modifications across the TSS of mRNAs overlapping group-1 and group-2 NATs. **(C)** Density plots showing the conservation score (phyloP8) at the promoters of mRNAs. Group-1 mRNAs are more conserved in comparison to other two groups **(D)** Heatmaps displaying G+C content at the promoters of mRNAs in the three categories.

Figure 5. Analysis of transcription regulation of NATs in the three different categories. (A) Heatmaps displaying the distribution of ChIP-seq reads for H3K27ac (red), H3K4me3 (red) and H3K27me3 (blue) histone modifications across the TSS of group-1 and group-2 NATs (B) Heatmaps of the distribution of ATAC-seq reads displaying the open chromatin at the TSS of group-1 and group-2 NATs **(C)** The figure shows the different motifs associated with in the promoters of group-1, group-2 and group-3 NATs. **(D)** Graph of promoter-width (x-axis) distribution of the NATs in the three groups.

Figure 6. Localization of NATs during zebrafish development (A) Heatmaps showing the localization of group-1 and group-2 NATs during zebrafish development. The scales are from -2 (red, cytosolic) to 2 (blue, nuclear) and 0 (white, both). **(B)** A boxplot of the alignment score between NATs and overlapping mRNA sequence in the group-1 (yellow), group-2 (blue) and group-3 (grey) group. Each box shows median value (as horizontal lines) and extend from 25^{th} to 75^{th} percentile values for each group. The outliers are shown as dots. The significance values are as * P<0.05; **P<0.005; ***P<0.001; ****P<0.0001; ns: non-significant (unpaired, two-tailed *t*-test). **(C)** A schematic model showing probable mechanism of group-1 NATs in downregulating developmental genes.

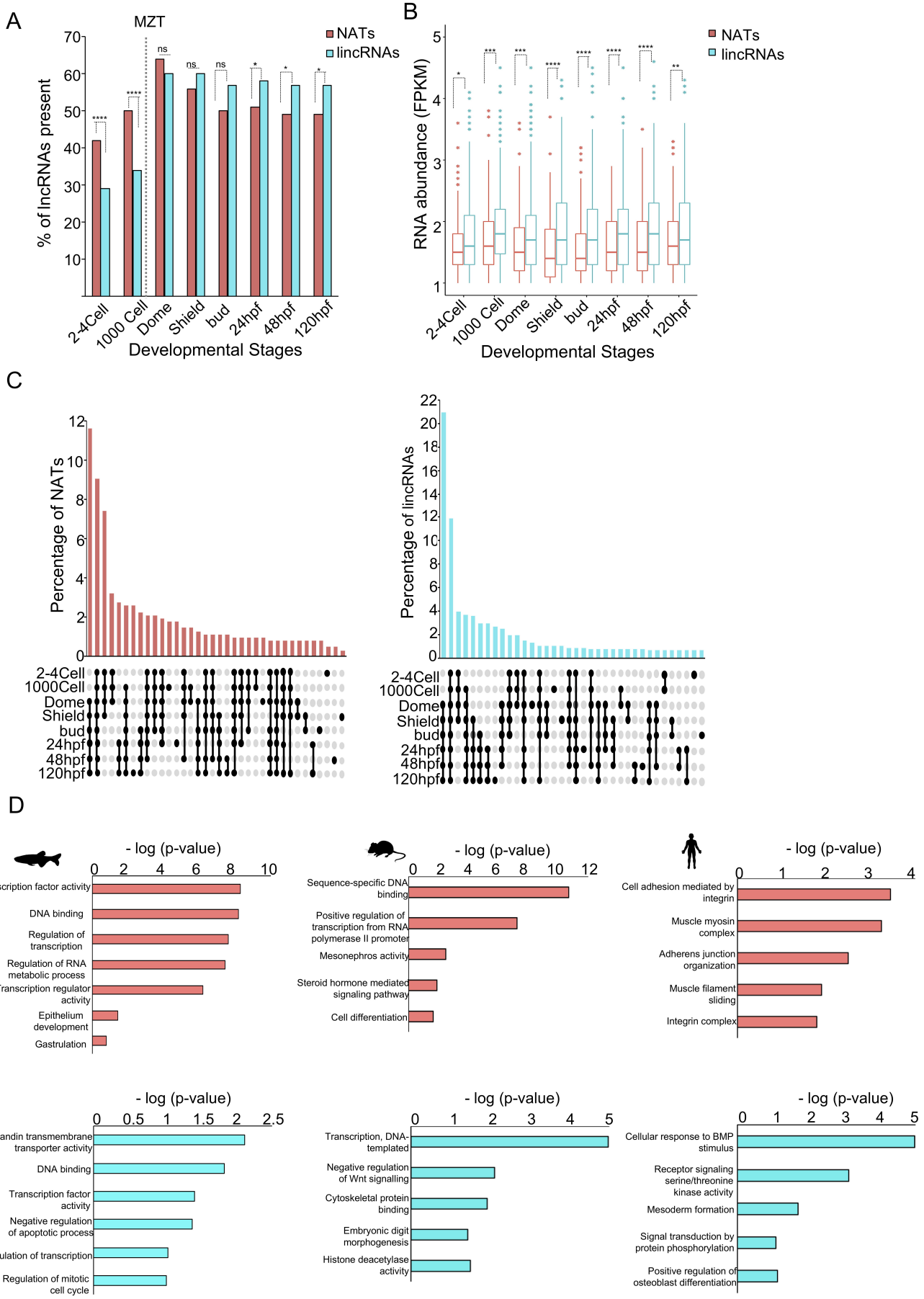

A
MZT
NATs
lincRNAs
% of lncRNAs present
2-4Cell
1000 Cell
Dome
Shield
bud
24hpf
48hpf
120hpf
Developmental Stages
B
RNA abundance (FPKM)
C
Percentage of NATs
Percentage of lincRNAs
D
- log (p-value)
Transcription factor activity
DNA binding
Regulation of transcription
Regulation of RNA metabolic process
Transcription regulator activity
Epithelium development
Gastrulation
Sequence-specific DNA binding
Positive regulation of transcription from RNA polymerase II promoter
Mesonephros activity
Steroid hormone mediated signaling pathway
Cell differentiation
Cell adhesion mediated by integrin
Muscle myosin complex
Adherens junction organization
Muscle filament sliding
Integrin complex
Prostaglandin transmembrane transporter activity
DNA binding
Transcription factor activity
Negative regulation of apoptotic process
Regulation of transcription
Regulation of mitotic cell cycle
Transcription, DNA-templated
Negative regulation of Wnt signalling
Cytoskeletal protein binding
Embryonic digit morphogenesis
Histone deacetylase activity
Cellular response to BMP stimulus
Receptor signaling serine/threonine kinase activity
Mesoderm formation
Signal transduction by protein phosphorylation
Positive regulation of osteoblast differentiation

Figure 1

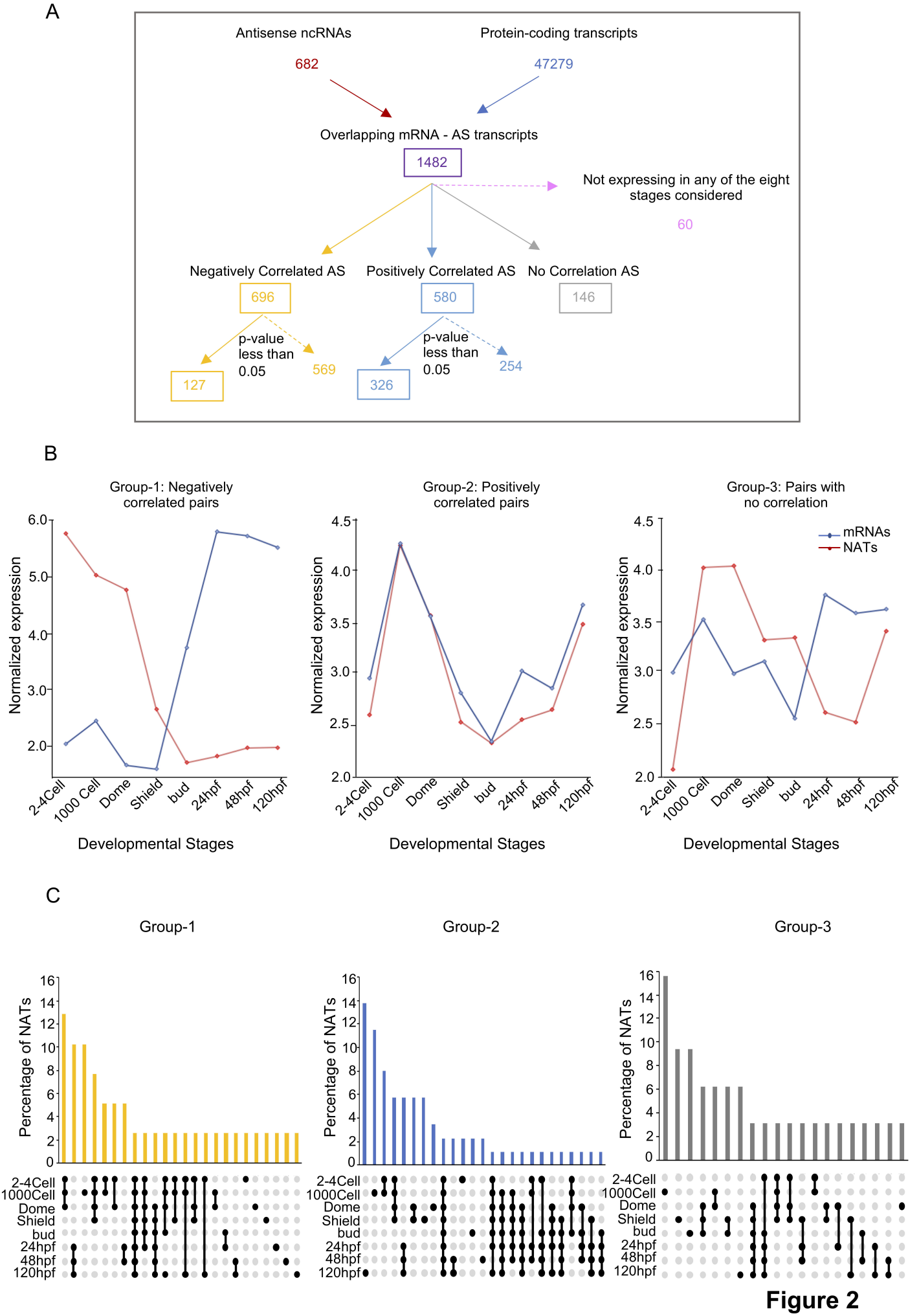

A
Antisense ncRNAs
682
Protein-coding transcripts
47279
Overlapping mRNA - AS transcripts
1482
Not expressing in any of the eight stages considered
60
Negatively Correlated AS
696
Positively Correlated AS
580
No Correlation AS
146
p-value less than 0.05
127
569
p-value less than 0.05
326
254
B
Group-1: Negatively correlated pairs
Group-2: Positively correlated pairs
Group-3: Pairs with no correlation
mRNAs
NATs
Normalized expression
Developmental Stages
2-4Cell
1000 Cell
Dome
Shield
bud
24hpf
48hpf
120hpf
C
Group-1
Group-2
Group-3
Percentage of NATs
2-4Cell
1000Cell
Dome
Shield
bud
24hpf
48hpf
120hpf

Figure 2

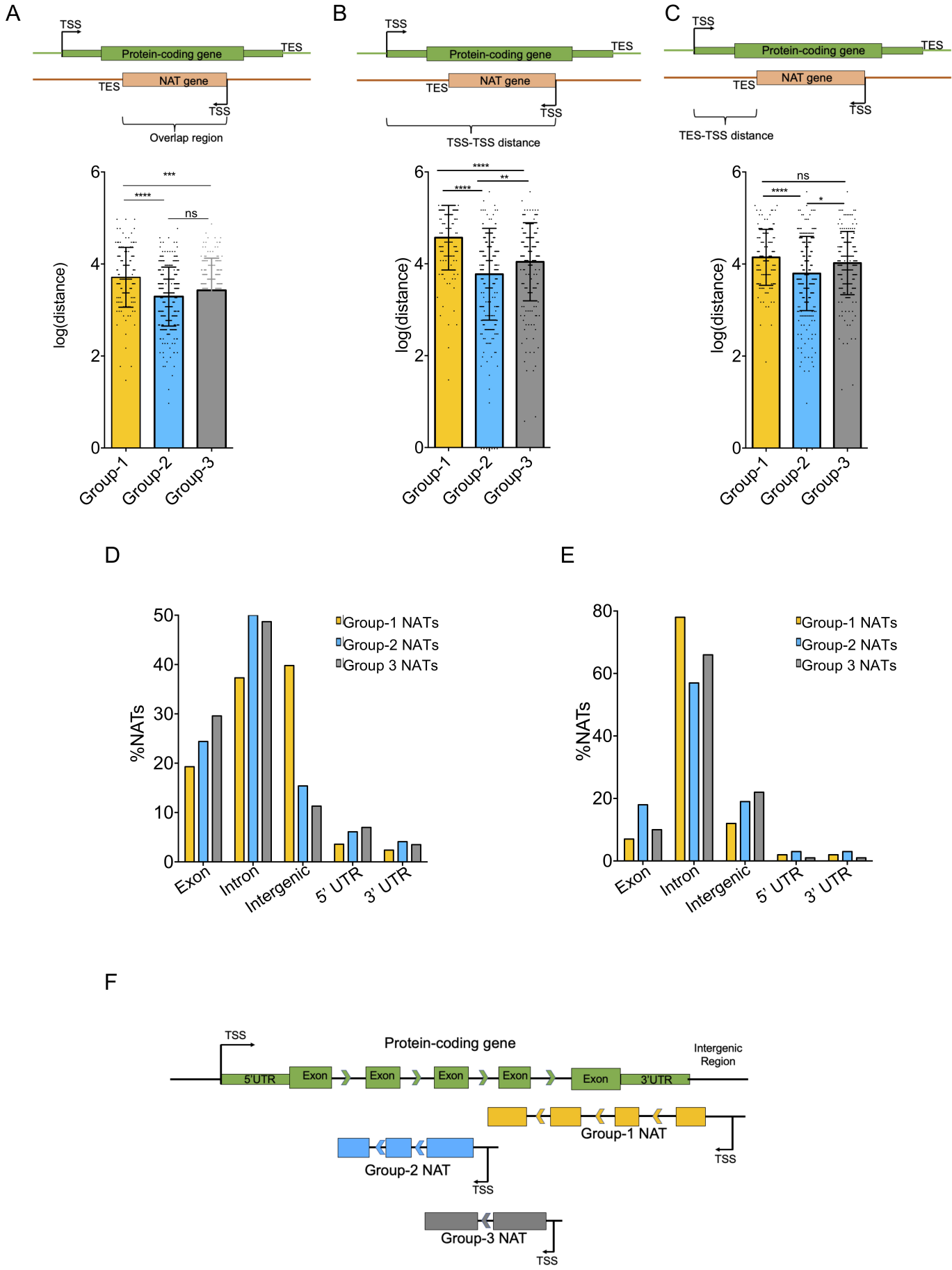

A
TSS
Protein-coding gene
TES
TES
NAT gene
TSS
Overlap region
log(distance)
***
****
ns
Group-1
Group-2
Group-3
B
TSS
Protein-coding gene
TES
TES
NAT gene
TSS
TSS-TSS distance
log(distance)
****
****
**
Group-1
Group-2
Group-3
C
TSS
Protein-coding gene
TES
TES
NAT gene
TSS
TES-TSS distance
log(distance)
ns
****
*
Group-1
Group-2
Group-3
D
%NATs
Group-1 NATs
Group-2 NATs
Group 3 NATs
Exon
Intron
Intergenic
5' UTR
3' UTR
E
%NATs
Group-1 NATs
Group-2 NATs
Group 3 NATs
Exon
Intron
Intergenic
5' UTR
3' UTR
F
TSS
Protein-coding gene
Intergenic Region
5'UTR
Exon
Exon
Exon
Exon
Exon
3'UTR
Group-1 NAT
TSS
Group-2 NAT
TSS
Group-3 NAT
TSS

Figure 3

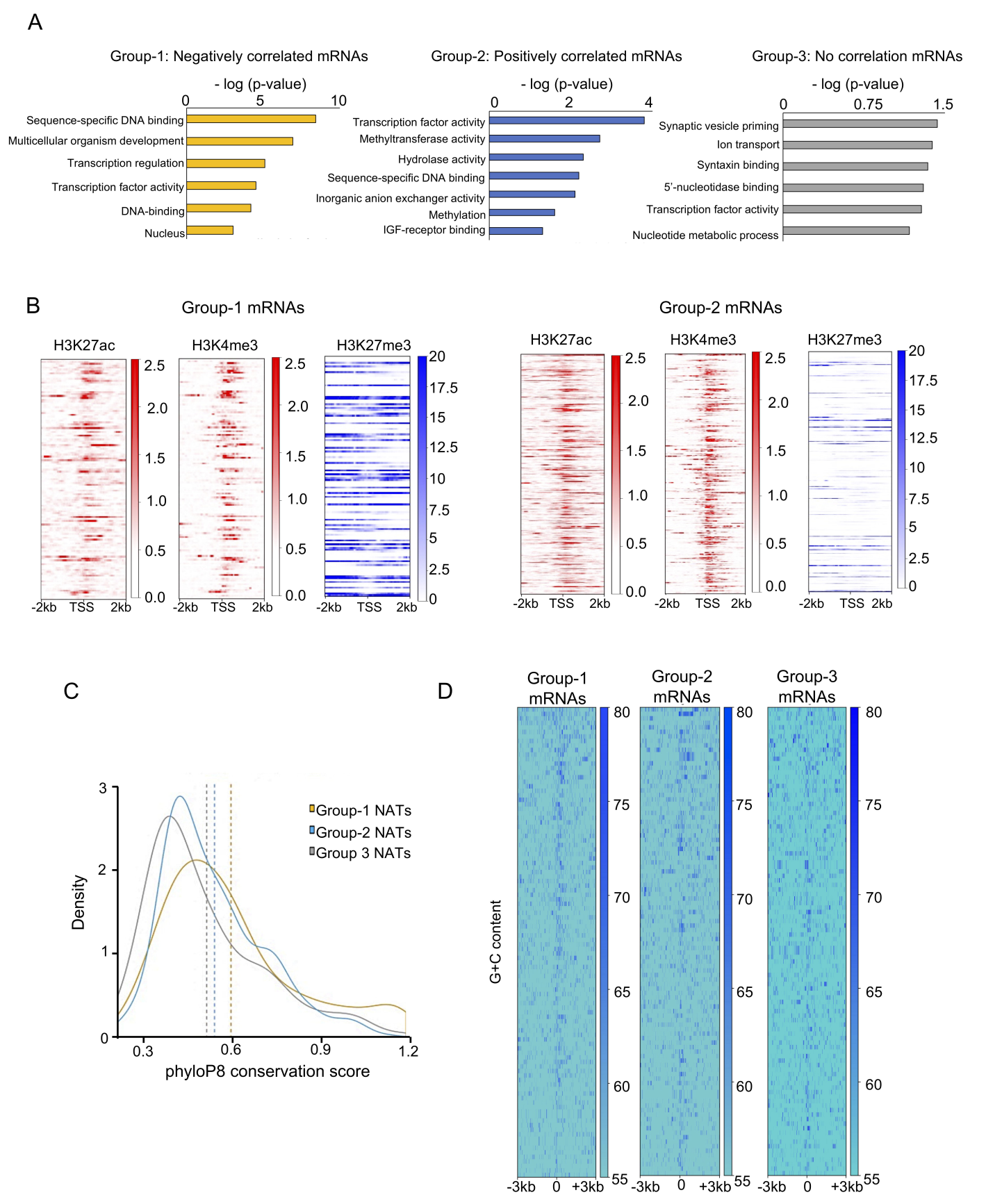

A
Group-1: Negatively correlated mRNAs
- log (p-value)
0
5
10
Sequence-specific DNA binding
Multicellular organism development
Transcription regulation
Transcription factor activity
DNA-binding
Nucleus
Group-2: Positively correlated mRNAs
- log (p-value)
0
2
4
Transcription factor activity
Methyltransferase activity
Hydrolase activity
Sequence-specific DNA binding
Inorganic anion exchanger activity
Methylation
IGF-receptor binding
Group-3: No correlation mRNAs
- log (p-value)
0
0.75
1.5
Synaptic vesicle priming
Ion transport
Syntaxin binding
5'-nucleotidase binding
Transcription factor activity
Nucleotide metabolic process
B
Group-1 mRNAs
H3K27ac
H3K4me3
H3K27me3
-2kb TSS 2kb
Group-2 mRNAs
H3K27ac
H3K4me3
H3K27me3
-2kb TSS 2kb
C
Density
Group-1 NATs
Group-2 NATs
Group 3 NATs
phyloP8 conservation score
D
Group-1 mRNAs
Group-2 mRNAs
Group-3 mRNAs
G+C content
-3kb 0 +3kb

Figure 4

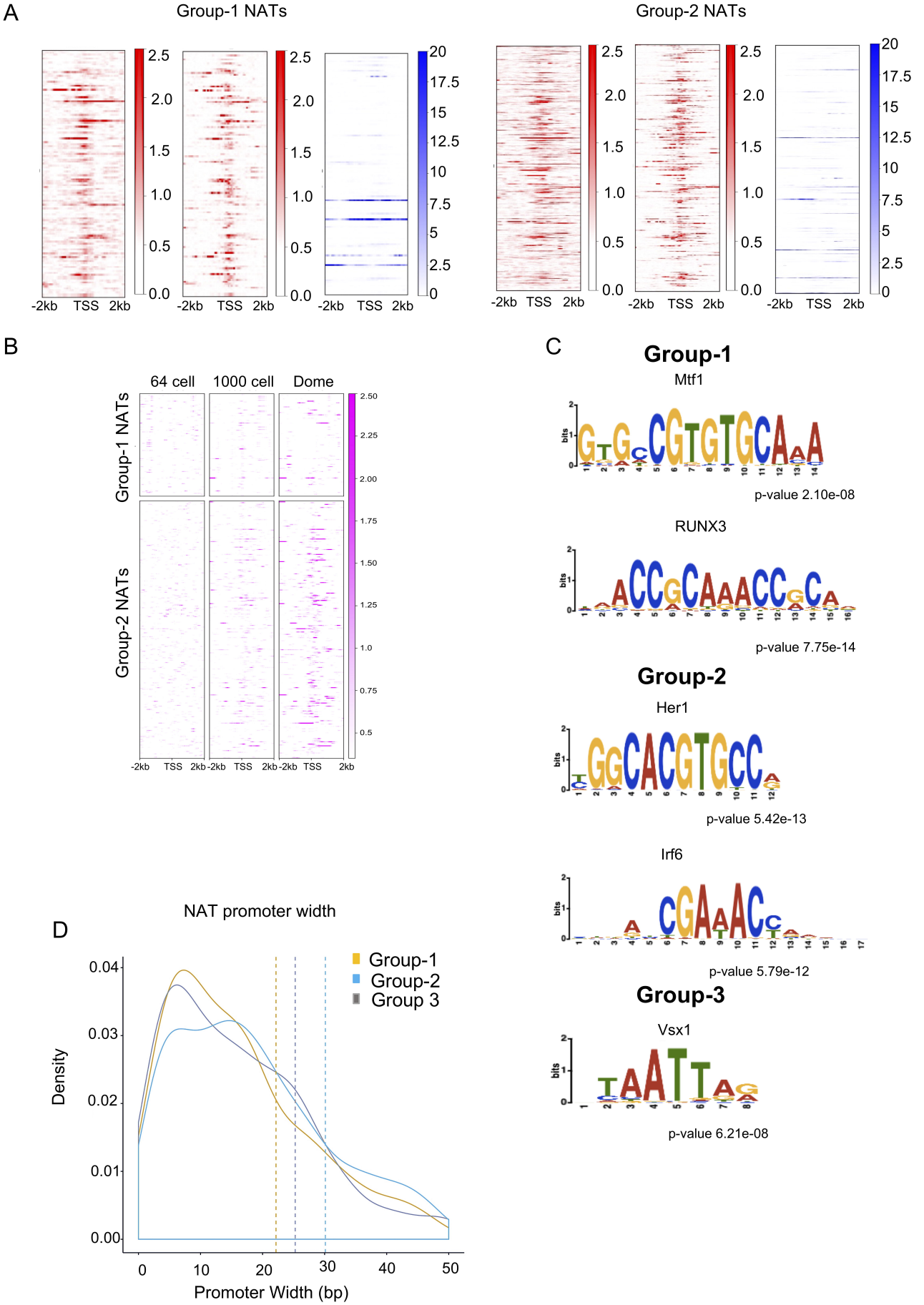
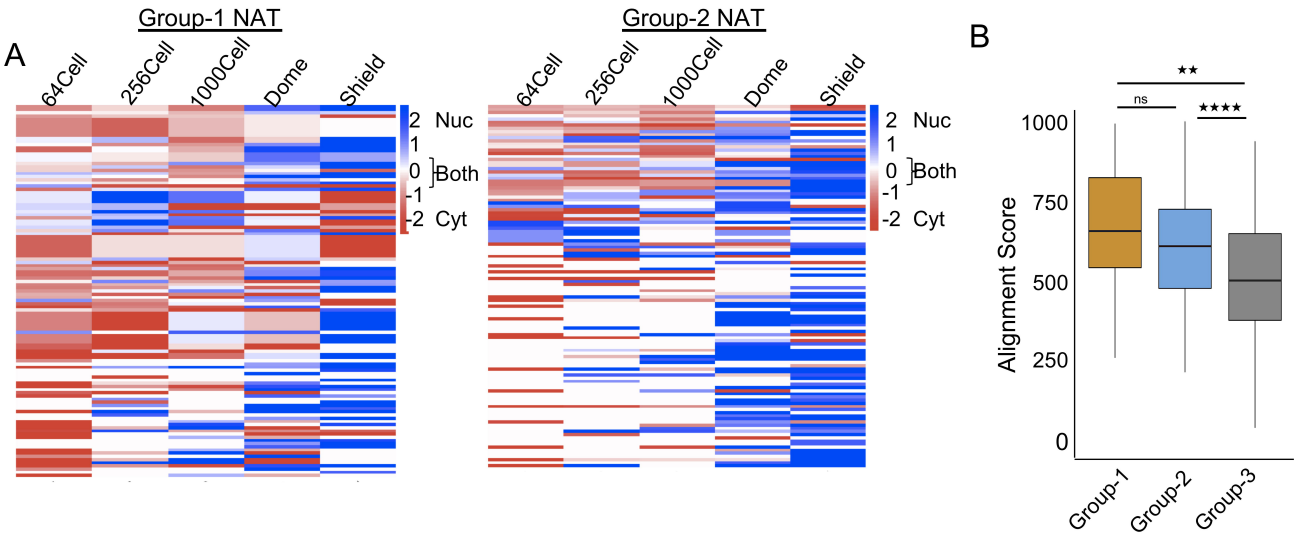

A
Group-1 NATs
Group-2 NATs
-2kb TSS 2kb
B
64 cell
1000 cell
Dome
Group-1 NATs
Group-2 NATs
C
Group-1
Mtf1
p-value 2.10e-08
RUNX3
p-value 7.75e-14
Group-2
Her1
p-value 5.42e-13
Irf6
p-value 5.79e-12
Group-3
Vsx1
p-value 6.21e-08
D
NAT promoter width
Group-1
Group-2
Group 3
Density
Promoter Width (bp)

Figure 5


A
Group-1 NAT
Group-2 NAT
64Cell
256Cell
1000Cell
Dome
Shield
2 Nuc
1
0 Both
-1
-2 Cyt
B
Alignment Score
1000
750
500
250
0
ns
**
****
Group-1
Group-2
Group-3

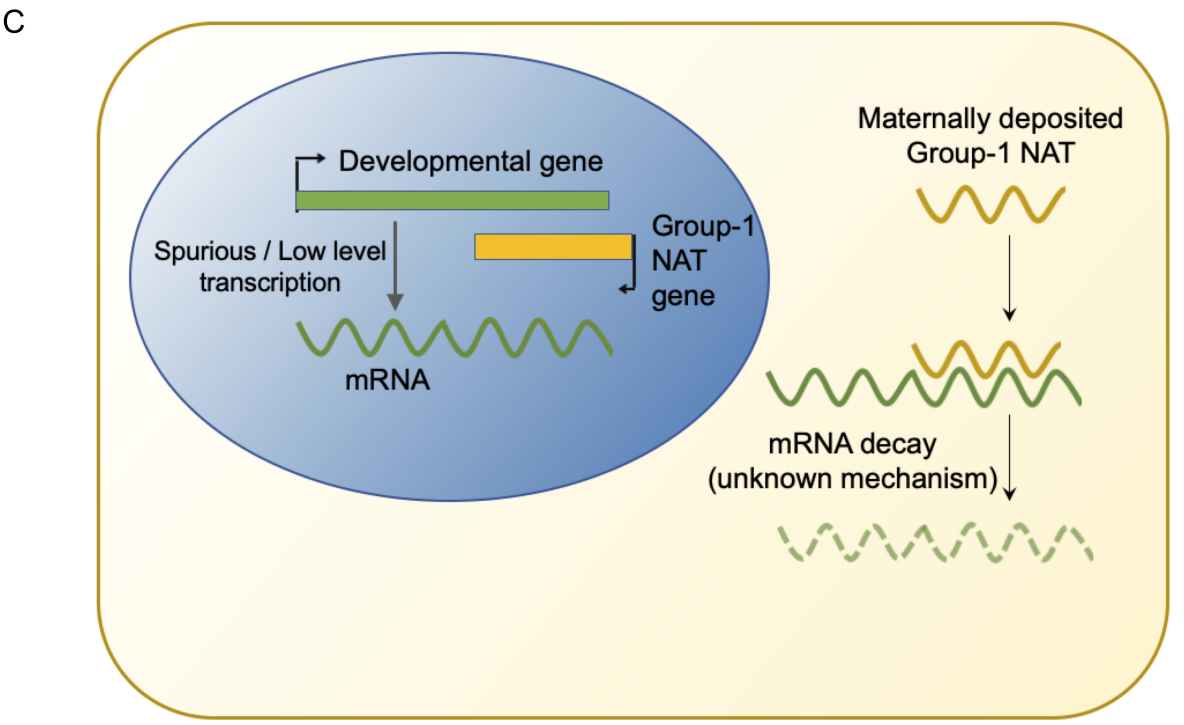

C
Developmental gene
Spurious / Low level transcription
Group-1 NAT gene
mRNA
Maternally deposited Group-1 NAT
mRNA decay (unknown mechanism)

Figure 6