

RESEARCH ARTICLE

Open Access

# Differential transcript isoform usage pre- and post-zygotic genome activation in zebrafish

Håvard Aanes<sup>1</sup>, Olga Østrup<sup>2,3</sup>, Ingrid S Andersen<sup>2,3</sup>, Lars F Moen<sup>1</sup>, Sinnakaruppan Mathavan<sup>4</sup>, Philippe Collas<sup>2,3</sup> and Peter Alestrom<sup>1\*</sup>

## Abstract

**Background:** Zebrafish embryos are transcriptionally silent until activation of the zygotic genome during the 10<sup>th</sup> cell cycle. Onset of transcription is followed by cellular and morphological changes involving cell speciation and gastrulation. Previous genome-wide surveys of transcriptional changes only assessed gene expression levels; however, recent studies have shown the necessity to map isoform-specific transcriptional changes. Here, we perform isoform discovery and quantification on transcriptome sequences from before and after zebrafish zygotic genome activation (ZGA).

**Results:** We identify novel isoforms and isoform switches during ZGA for genes related to cell adhesion, pluripotency and DNA methylation. Isoform switching events include alternative splicing and changes in transcriptional start sites and in 3' untranslated regions. New isoforms are identified even for well-characterized genes such as *pou5f1*, *sall4* and *dnmt1*. Genes involved in cell-cell interactions such as *f11r* and *magi1* display isoform switches with alterations of coding sequences. We also detect over 1000 transcripts that acquire a longer 3' terminal exon when transcribed by the zygote compared to their maternal transcript counterparts. ChIP-sequencing data mapped onto skipped exon events reveal a correlation between histone H3K36 trimethylation peaks and skipped exons, suggesting epigenetic marks being part of alternative splicing regulation.

**Conclusions:** The novel isoforms and isoform switches reported here include regulators of transcriptional, cellular and morphological changes taking place around ZGA. Our data display an array of isoform-related functional changes and represent a valuable resource complementary to existing early embryo transcriptomes.

**Keywords:** Zebrafish, Mid-blastula Transition, Zygotic Genome Activation, Alternative Splicing, Transcriptional Start Site, 3'UTR

## Background

During the first ten cell cycles after fertilization, the zebrafish embryo is transcriptionally silent and consists of undifferentiated and rapidly dividing blastomeres [1]. Initiation of transcription at the 10<sup>th</sup> cell cycle is termed zygotic genome activation (ZGA). At the time of ZGA, development proceeds from a control by mRNAs synthesized during oogenesis and stored in the egg (maternal transcripts) to mRNAs produced by the embryo's own genome. Following ZGA, blastomeres divide less frequently and more asynchronously, they start to differentiate and migrate to form the three germ layers of the

gastrulating embryo [1-4]. Collectively, these transformations characterize the mid-blastula transition (MBT). These fundamental cellular and functional changes occurring during development make zebrafish an attractive model to study transcriptional changes governing between pre-MBT and post-MBT development.

Previous studies have shown essential roles of activation and degradation of maternal transcripts in regulating the MBT and ZGA [5-7]. Signaling pathways involving Bmp, Nodal, Fgf, Wnt and maternal  $\beta$ -catenin are essential for the formation of germ layers and body axis [8]. We and others have shown that the establishment of post-transcriptionally modified histones on specific genomic sites and DNA methylation play a role in transcriptional regulation around the time of ZGA by patterning developmental gene expression [9-12]. However, although several early

\* Correspondence: [petera@nvh.no](mailto:petera@nvh.no)

<sup>1</sup>BasAM, Norwegian School of Veterinary Science, 0033 Dep, Oslo, Norway  
Full list of author information is available at the end of the article

transcriptomes have recently been published [5,13,14], little is known on the isoform-specific dynamics governing developmental transitions around the MBT.

The notion that each gene can give rise to multiple mRNAs has evolved from being reported as a rare phenomenon [15,16] to include virtually all loci in man [17,18], and has been shown to be crucial for differentiation [19], development [20] and human disease [21]. Isoform switches, defined as a change in the isoform composition of gene products between two conditions (e.g. two developmental stages) are of particular interest since such events are critical for differentiation [19]. The differences between transcript isoforms can affect the coding sequence (CDS) and/or untranslated regions (UTRs) 3' or 5' of the CDS. The former is likely to affect protein function, while the latter may affect translational efficiency, mRNA degradation kinetics and spatial distribution of transcripts [22,23]. The mechanisms regulating splicing and determining the production of specific transcript isoforms have started to be unveiled and involve *cis*-elements and *trans*-acting factors, as well as epigenetic modifications in the proximity of spliced exons [24-26]. A genome-wide landscape of transcript isoforms synthesized during early development, and particularly of the switches occurring between maternal and zygotic isoforms at the time the embryo initiates its own transcriptional program, has been lacking and has hampered a comprehensive appreciation of the transcriptional dynamics occurring at the time of ZGA.

We have used an isoform prediction and quantification program, Cufflinks [27], to detect novel isoforms and quantify isoform-specific changes from RNA-sequencing (RNA-seq) reads before and after ZGA in zebrafish. We identify numerous novel isoforms related to shifts in transcription start site (TSS), alternative splicing (AS) events and transcription termination sites (TTS), when comparing transcripts of maternal and zygotic origin coming from the same gene. These include transcripts of genes involved in cell-cell interactions, pluripotency control and DNA methylation. Using H3K4 and H3K36 trimethylation (me3) data acquired by chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq), we find that H3K4me3 can form relatively broad domains which cannot distinguish between closely spaced alternative TSSs, unless TSSs are linked to alternative promoters at distant locations. Skipped exons are enriched in H3K36me3 and H3K4me3, extending previous reports on the involvement of these histone modifications in the regulation of isoform usage.

## Results and discussion

### Mapping of RNA-seq reads achieved from combined use of two aligners

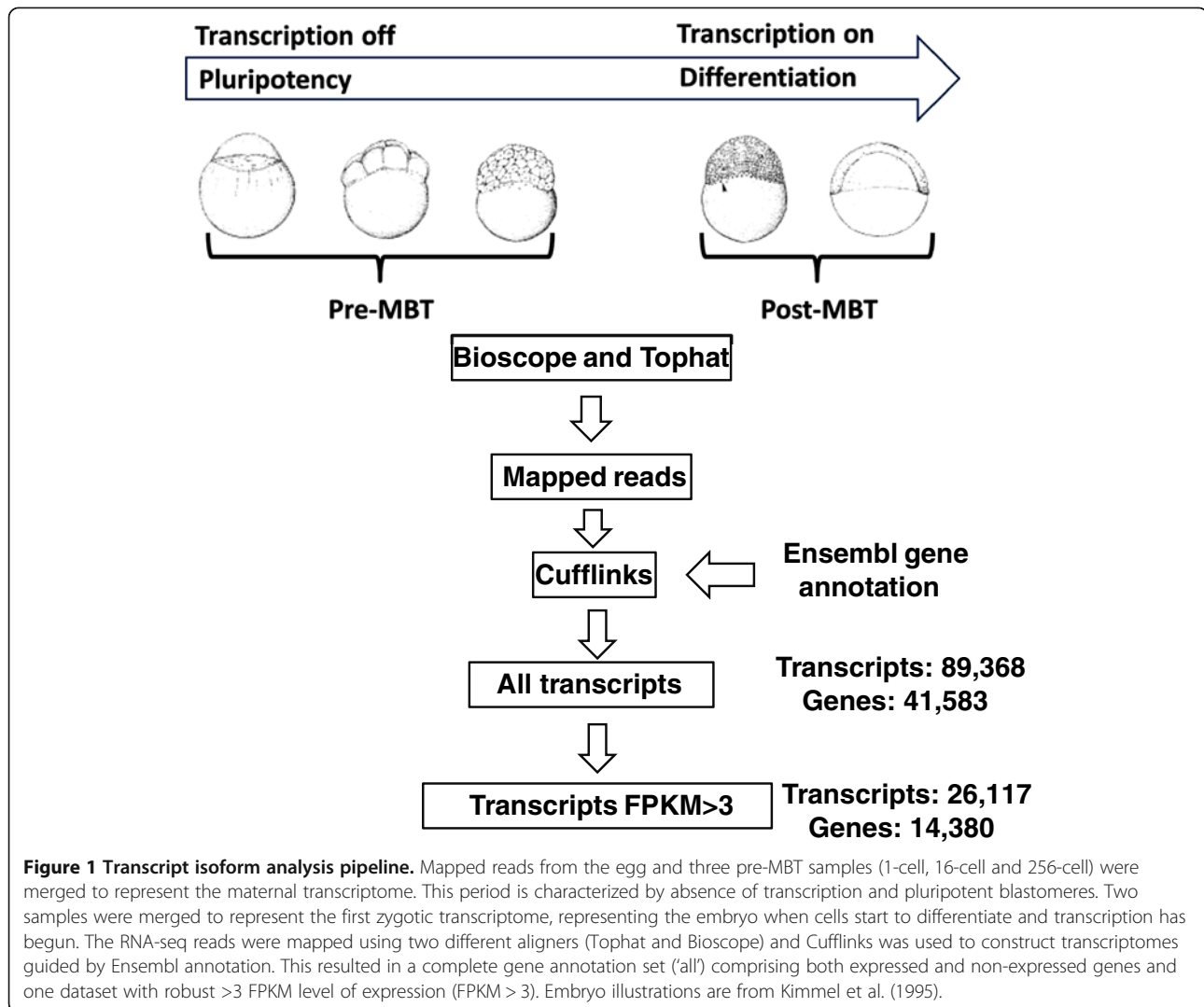
Using two different short RNA-seq read aligning programs, namely Bioscope (Life Technologies, USA, Carlsbad, CA)

and TopHat [28], we remapped all reads generated by RNA-seq in a recently published study [5]. This 'two aligner' approach yields a higher number of mapped reads than each aligner alone and proves to be complementary compared to using either aligner separately (Figure S1 and S2 in Additional file 1). From our RNA-seq data, four datasets from pre-ZGA zebrafish developmental stages (pre-MBT) and two from post-ZGA stages (post-MBT) were merged, respectively generating one maternal transcriptome dataset, and one combined late maternal-early zygotic (henceforth termed zygotic) dataset (Figure 1). This strategy yields an opportunity to examine the embryonic transcriptome before and after ZGA, and provides more sequencing depth and thus higher accuracy in constructing transcripts. A total of 71.5 and 49 million RNA-seq reads were thus mapped from the pre-MBT and post-MBT developmental stages, respectively.

### Transcriptome assembly identifies novel genes and isoforms

We used the Cufflinks software and associated tools [27,29] to assemble transcriptomes based on the mapped reads (Figure 1). Cufflinks was run in Reference Annotation Based Transcript assembly (RABT)-mode [30], using annotations from Ensembl as a guide (Zv9, version 64). We created two datasets. i) A set of all Ensembl annotations and all transcripts assembled was generated, referred to as the 'all' gene annotation; this set contains 89,368 transcripts from 41,583 genes (Additional file 2). ii) Expression values were obtained for the 'all' dataset (Additional file 3) and these were used to filter the 'all' gene annotation to generate a file of robustly expressed transcripts, defined as fragments per kilobase per million mapped fragments with a value  $>3$  (FPKM  $>3$ ) (Additional file 4). This dataset contains 26,117 isoforms from 14,380 genes (Figure 1). Of these transcripts, the vast majority (22,600) have multiple exons (Figure 2a). These multiple-exon transcripts arise from 10,863 genes (Figure 2a), thus multi-exon loci produce on average nearly 2 isoforms. Altogether, we identify 22,453 TSSs (Figure 2b) and 20,562 transcription termination sites (TTSs) in the FPKM  $>3$  gene annotation dataset (Figure 2c). This indicates that a large fraction of the isoforms arise from differential use of TSSs or TTSs. However, the fact that isoforms from the same locus with different TSSs also frequently differ in their exon content is a reflection of Cufflinks' parsimonious nature, because Cufflinks proposes as few transcript models as necessary to explain the aligned reads [27].

To identify novel isoforms and genes, the predicted isoforms in the FPKM  $>3$  dataset were compared to Ensembl annotations (Zv9, version 64). This classifies 8,108 of the transcripts as identical, 12,100 as new isoforms, and 4,174 as novel transcribed regions (NTRs)



i.e. possibly novel genes (Figure 2d). This analysis indicates that the assembled zebrafish transcriptome contains numerous novel transcripts, some of which may be encoded by novel genes.

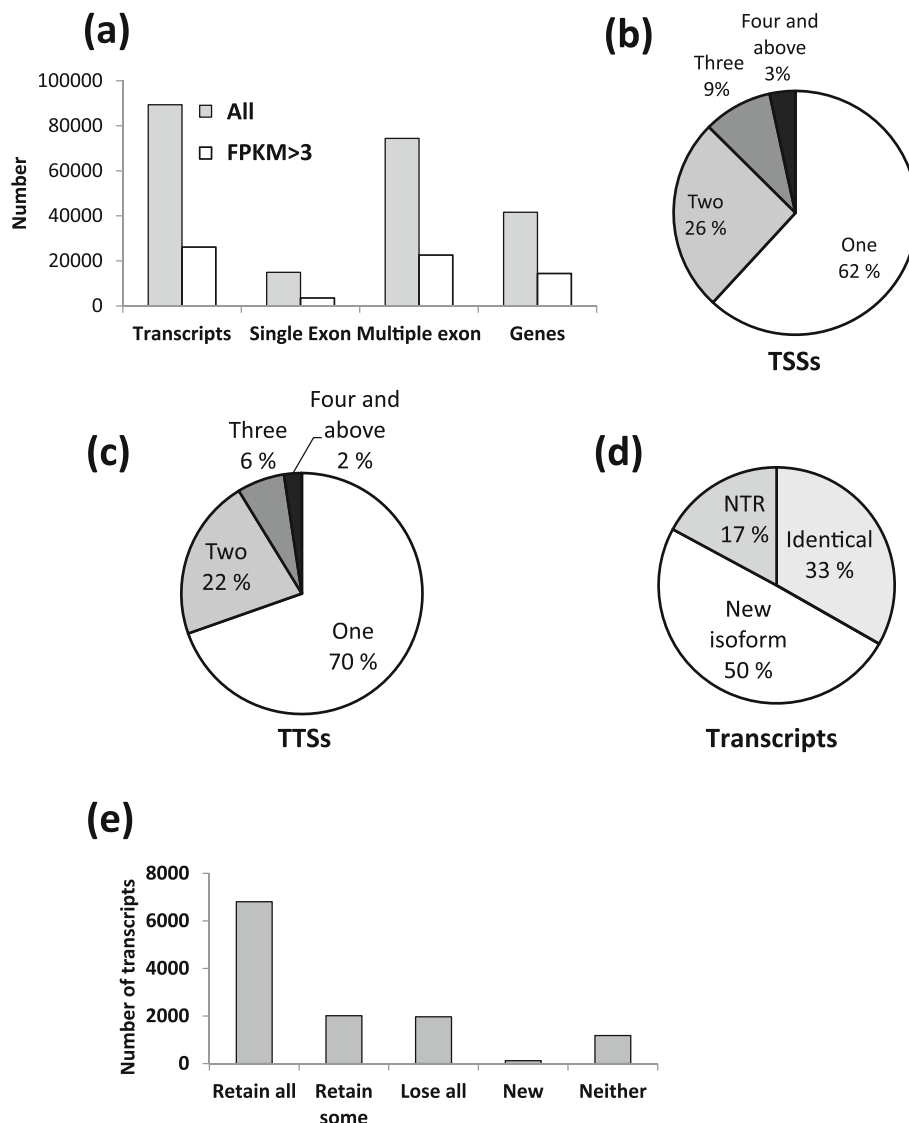
#### Functional differences between annotated and novel transcripts

The large number of novel isoforms warranted further validation and characterization. We investigated the coding potential of new isoforms and found that nearly all have an open reading frame (ORF) (11,627/12,100; 96%) of >50 amino acids in length (starting with an AUG start codon and ending with a stop codon). Increasing the ORF length requirement results in a slight decrease in the number of ORFs detected (100 amino acids, 89%; 150 amino acids, 81%).

We then compared both the transcript sequence and the inferred coding sequence between novel isoforms and their closest matching annotated transcript from the

same locus (as defined by Cufflinks). On average, the novel isoforms have shorter coding sequences than the corresponding annotated coding sequences with a mean difference of -230 bp (median -109 bp). Of 11,529 comparisons, 1,695 of the novel isoforms differ by one amino acid or less, indicating identical coding sequences (this corresponds to 15% of the transcripts with an ORF of ≥50 amino acids). Transcript isoforms with no difference in coding sequences length have on average longer 5' and/or 3' UTRs, with a mean length difference of 549 bp (median 256 bp). Therefore, the novel isoforms can be broadly divided into two groups, those with no change in coding sequence length but with longer 5' and/or 3' UTR, and those with an alteration in coding sequence length.

We next examined any functional effect these transcript isoforms may have, and used the predicted protein sequences for a domain scan using Pfam, a database of protein families [31]. The detection of a protein domain has been shown to be suggestive of protein functionality



**Figure 2** Characterization of the pre-ZGA and post-ZGA embryo transcriptomes. **(a)** Number of genes, transcripts and transcripts with one or more exons in the all and FPKM > 3 datasets. **(b)** Percentages of loci with 1, 2, 3 or more TSSs. **(c)** Percentages of loci with 1, 2, 3 or more TTSs. **(d)** Percentage of isoforms classified as NTR, identical (known) or new in the FPKM > 3 dataset. **(e)** Difference between the novel and annotated transcripts on functional domain(s) content. Most novel isoforms had retained all functional domains as compared to their closest matching annotated counterpart.

[32]. Of the novel isoforms analyzed, 8,949 (74%) contain a sequence with high similarity to an identified protein domain. In comparison, the closest matching annotated transcripts ( $n = 5,908$ ) have a substantially larger fraction of Pfam domain-containing sequences (5,196; 88%). Further, most of the novel isoforms encode the same functional domains as the annotated isoforms (Figure 2e). However, a substantial fraction loses either all functional domains ( $n = 1,971$ ), or some of them ( $n = 2,017$ ; Figure 2e). Also, a small subgroup ( $n = 123$ ) of the novel isoforms contains functional domains absent from the annotated isoform (Figure 2e). Taken together, these results suggest that the

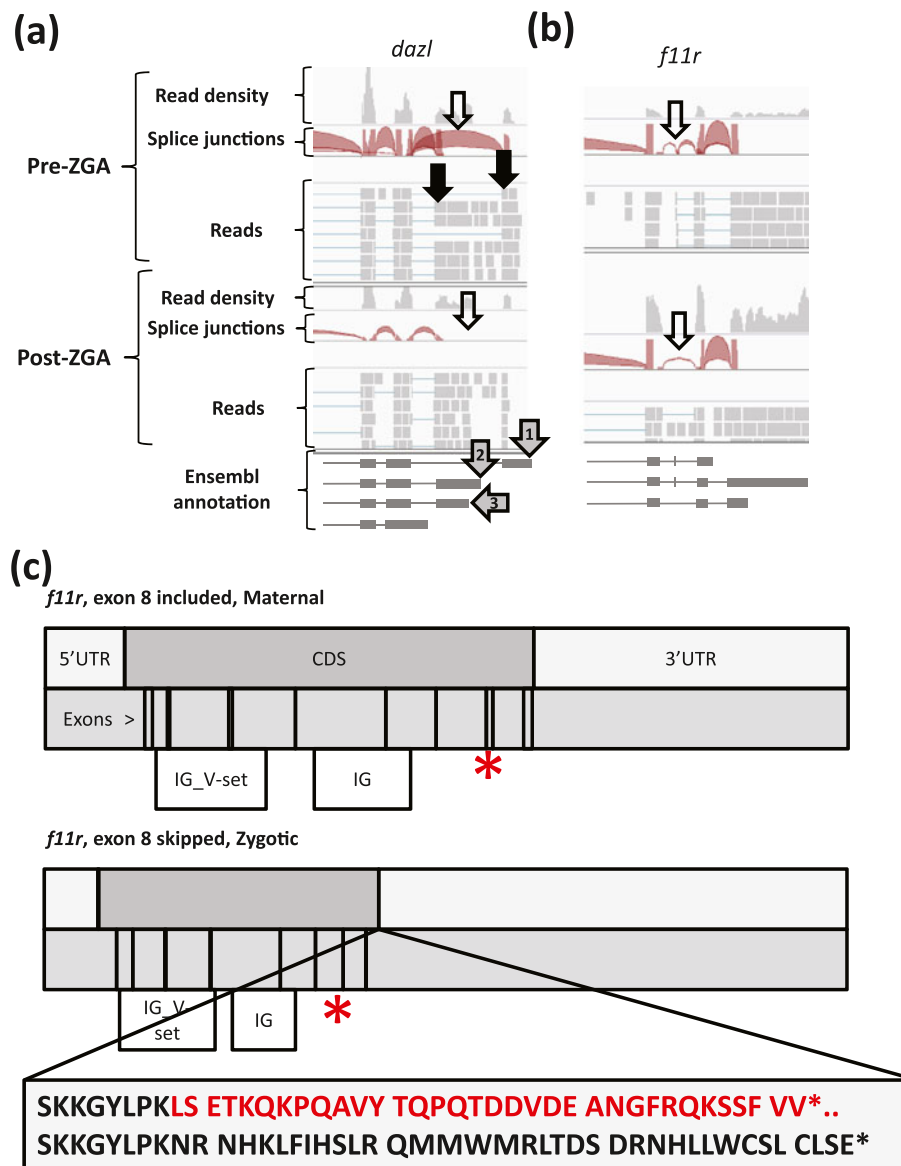
majority of the novel transcript isoforms are functional, with roles related to their annotated counterparts. However, some lose or gain functional domains, suggesting that they might exhibit different biological functions.

#### TSS usage and switching

Transcription is initiated at TSSs and differential usage of TSSs is often associated with tissue-specific expression or distinct protein products [33]. Using Cufflinks, we identify 513 loci ( $q$ -value < 0.001; Additional file 5) with a change in TSS usage between pre- and post-ZGA stages. A typical example is the isoforms of *dazl* encoding an RNA-binding

protein crucial for germ cell development, oocyte-to-zygote transition and maintenance of pluripotency [34-36]. The biological function of the Dazl protein is at the level of regulation of translation [37] and recent experiments indicate that Dazl protects transcripts from degradation, possibly through antagonizing the de-adenylating effect of miRNAs [38]. Our data show that the transcript isoforms

of *dazl* are expressed from at least two promoters maternally (Figure 3a; promoter 1 = grey arrow 1 (most upstream); promoter 2 = grey arrows 2 and 3, black arrows in the top panel point towards reads supporting expression from both TSSs pre-ZGA). In the post-ZGA transcriptome, the lack of reads supporting usage of promoter 1 (Figure 3a; white arrow, bottom panel) demonstrates



**Figure 3 Dazl TSS switching and f11r exon skipping switch.** (a) A screenshot from Integrative Genomics Viewer shows that *dazl* has been transcribed from at least two TSS's in the maternal transcriptome; grey arrows 1-3 at bottom indicate Ensembl annotated TSSs. Black arrows in the upper panel indicate reads supporting promoter 1 (rightmost black arrow) and promoter 2/3 (leftmost black arrow; our data cannot distinguish between TSS 2 and 3). A white arrow in the upper panel points to a splice junction 'bridge' (red), symbolizing the splice junction for the most distal first exon. The thickness of splice junction symbols is correlated with the number of reads supporting the junction. Post-ZGA transcripts lack the most 5' exon as can be observed from the absence of reads supporting the splice junction originating from this exon (white arrow, bottom panel). (b) Pre-ZGA an *f11r* isoform with exon 8 included is present (white arrow, top panel). Post-ZGA a different *f11r* isoform, this one with exon 8 lacking, has been transcribed (white arrow, bottom panel). (c) The *f11r* exon skipping (red asterisk) leads to a frame shift resulting in a 6 nt longer ORF, including the removal of two C-terminal valine residues. Two functional domains, immunoglobulin V-set domain (PF07686) and immunoglobulin domain (PF00047) (white bars) are intact in both splice isoforms.

degradation of the maternal-specific 5' distal isoform during the MBT (Figure 3a). RT-qPCR confirms the trend observed by RNA-seq (Figure S3 in Additional file 1). Both *dazl* isoforms are present in Ensembl (pre-ZGA: ENSDART00000147911; post-ZGA: ENSDART00000137590). The CDS of the maternal-specific transcript displays a 17 amino acid shorter N-terminus compared to the zygotic transcript. However, the functional implication of this difference is not straightforward because the RNA-recognition motif of the Dazl protein is located 30 amino acids downstream and is not affected, unless induced by configuration change. Considering the diverse and temporarily restricted roles of Dazl in germ cell development and oocyte-to-zygote transition [34,35], it is tempting to speculate that the maternal and zygotic isoforms may be responsible for two biological processes: the 'maternal' version may act in the oocyte-to-zygote transition, whereas the zygotic variant may act in germ cell development.

Among the 513 genes which display a change in TSS pattern between the pre- and post-ZGA, *dazl* represents a subset where the TSSs are defined by different first exons. This type of TSS switch is likely to affect developmental functions as illustrated with *dazls* changes in the 5'UTR and N-terminus of the protein.

#### Epigenetic regulation of TSS usage and expression

H3K4me3 enrichment over the TSS is considered as a histone modification associated with transcription permissiveness. While virtually all expressed genes are marked by H3K4me3 at the TSS [39,40], many genes that are not expressed also harbor H3K4me3 [9,41]. H3K27me3 promoter occupancy is on the other hand associated with transcriptional repression [42]. To identify a relationship between TSS usage and potentially associated chromatin marks, we used published ChIP-seq peak calling data for H3K4me3 and H3K27me3 in post-MBT stage embryos [13], and investigated the correlation between these histone modifications and the TSSs mapped in our study.

We find that a subset of 22,534 TSSs from 13,542 genes from our 'all' gene annotation dataset are enriched in H3K4me3. Most peaks (18,547/23,778; 78%) were within  $\pm 500$  bp of a gene. The number of TSSs in the dataset and the number of peaks covering them indicates that multiple TSSs are included within each H3K4me3 peak, as exemplified by *dazl* and *sall4* (Figure S4a,b in Additional file 1). On rare occasions, we observe that H3K4me3 peaks are located over the most frequently used TSS (major TSS) (e.g. *actl6a*; Figure S4c in Additional file 1). Therefore, at present H3K4me3 peaks cannot be used to predict the major TSS since the mark can overlap multiple TSSs. More accurate determination of TSS usage, e.g. using approaches such as cap analysis of gene

expression (CAGE) together with H3K4me3 enrichment data may improve the usefulness of this histone modification in predicting major active start sites among closely located TSSs.

Our RNA-seq data show that 21% of the H3K4me3-marked genes are expressed at very low levels, with 2,822 loci displaying an FPKM value below 1 in both pre- and post-ZGA samples (Additional file 3). For the subgroup of bidirectional promoters identified using the H3K4me3 peaks ( $n = 1,433$  promoter pairs) (see Methods) we find no significant correlation in expression between the bidirectional gene pairs (Pearson's product-moment correlation  $< 0.01$ ), again supporting H3K4me3 as a mark for predisposition of promoter activation rather than having a direct activating effect, even when presumably sharing the same promoter. However, on average, H3K4me3-marked genes are more strongly expressed than genes without H3K4me3, as described earlier for RefSeq genes (data not shown) [10].

An important question is whether epigenetic marks can regulate particular subgroups of genes, functionally and spatially. To address the first question, analysis in DAVID of gene ontology (GO) terms for biological functions were obtained using H3K4me3- and H3K27me3-marked genes. This was significant only for one term for the H3K4me3-marked genes; cell cycle (FDR  $< 0.001$ ; Additional file 6). This is in contrast to H3K27me3-marked genes which were enriched for 21 GO terms with the same cut off (i.e. regulation of transcription, embryonic morphogenesis, pattern specification process and regionalization; Additional file 7). Thus, whereas H3K4me3 mark TSSs of genes linked to many biological functions, H3K27me3 marks mostly the TSSs of genes involved in developmental functions. These results are consistent with a H3K4me3 pre-patterning of promoters of a specific subset of genes pre-MBT (involved in housekeeping and homeostatic functions), and with a view of H3K27me3 repressing developmental genes prior to their activation at or after the MBT [9,10].

Another intriguing question is whether TSS marking with both H3K4me3 and H3K27me3 exists (so-called 'bivalency'). We utilized the extensive annotation database of where zebrafish genes are expressed anatomically (anatomical ontology provided by the zebrafish model organism database; ZFIN [43]) to bring further clarity to this issue. H3K4me3-marked genes are mostly expressed in the entire organism, in contrast to H3K27me3-marked genes which are predominantly annotated with expression in distinct tissue types (in particular central nervous system structures; Additional file 8 and Additional file 9). This is consistent with H3K27me3 analysis of dissected *Xenopus* embryos [44], and not supportive of bivalency in zebrafish embryos, as others have suggested [9].

Taken together, our results show that during zebrafish development, H3K4me3 is a mark of TSSs independent of transcription status. The marked genes show no pronounced tissue-specific preferences and few specialized biological function as a group. In contrast, H3K27me3 marked genes have specific biological functions and are expressed in specialized cells. Our results are in line with the view that H3K4me3 contributes to a transcriptionally permissive chromatin environment for TSS usage; however, the repressive marks seem to contribute to lineage commitment and tissue-specific differentiation [45]. The use of ChIP-seq data and a comprehensive gene annotation based on existing and novel annotations confirm the trends from previous studies based on older annotation and ChIP-chip data [9,10].

### Shifts in alternative splicing during the switch from maternal to zygotic transcripts

Alternative splicing (AS) allows the production of multiple mature mRNAs from the same primary transcript, increasing genetic diversity of a genome by several-fold. We used Cufflinks to identify shifts in AS isoform patterns before and after the ZGA. A total of 2,160 primary transcripts exhibited a significant change in the splice isoform composition (examples include *foxh1*, *hdac5*, *dnmt3* and *gsk3a*; Additional file 10). Quality control by visual inspection revealed mostly subtle changes among the significantly changing primary transcripts. Importantly we find, also by visual inspection, that several obvious isoform switches do not pass the significance criteria of Cufflinks. For example, we report *f11r*, encoding a trans-membrane adhesion protein found in association with tight junctions [46,47], which displays a switch between the maternal and zygotic stages, with exon 8 skipped only in the zygotic transcripts (Figure 3b). Using reverse transcription (RT)-qPCR, we validated the expression pattern of *f11r* (Figure S5 in Additional file 1). Exon 8 of *f11r* is only 13 bp long and is included in the maternal isoform of the *f11r* transcript (Figure 3c, upper panel). Skipping of exon 8, detected in the zygotic transcriptome (Figure 3c, lower panel), causes a frame-shift with an alteration of the C-terminal 36 amino acids; however two known functional domains in the F11r protein are not directly affected, although indirect structural effects cannot be excluded (Figure 3c). The change in the intracellular C-terminal portion of the protein includes the removal of two terminal valine residues which are involved in binding to PDZ-domain containing proteins [46]. In zebrafish, interference with F11r function causes defects in migration at later stages of development [48]. Interestingly, the shift in *f11r* isoform usage between pre- and post-MBT stages coincides with the initiation of cell migration taking place after the MBT [1,2]. It is tempting to speculate that attenuation of

binding potential to PDZ domain-containing proteins enables cells to alter their migration patterns [49,50]. However, functional studies are required to test this hypothesis.

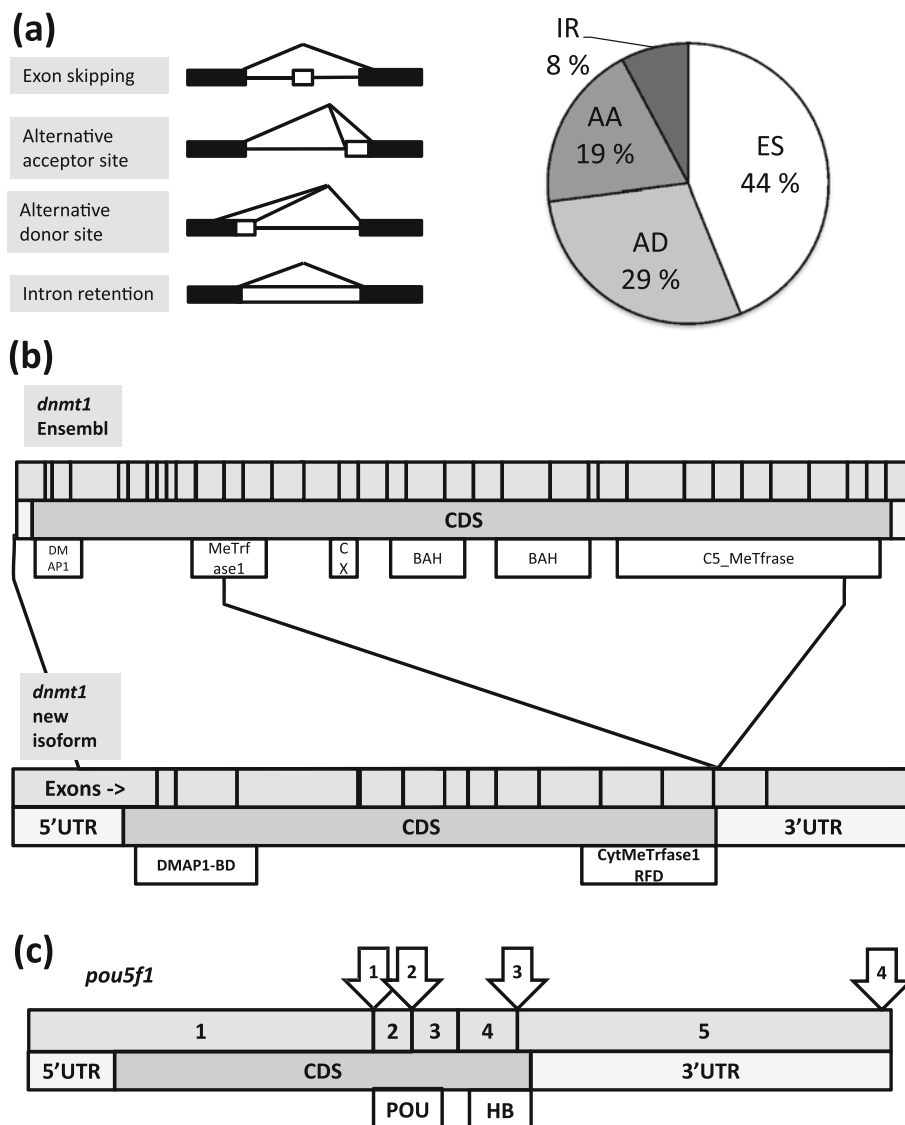
From our dataset, it appears therefore that thousands of splice isoform switches occur between the maternal and zygotic transcriptomes; nonetheless, it is important to emphasize that the majority constitutes subtle changes rather than obvious functional on/off events.

### Quantification of AS events

We used the software ASTALAVISTA [51] to quantify the frequency of four different types of AS events; exon skipping (ES), alternative donor (AD) and acceptor (AA) sites and intron retention (IR) (Figure 4a). Using the FPKM > 3 transcript subset, we identify a total of 10,950 AS events linked to 3,843 loci (36% of the multi-exon loci). AA and AD sites make up 2,107 (19%) and 3,185 (29%) of these events, respectively, and we find 852 IR and 4,806 ES occurrences (Figure 4a). Noteworthy, only 39% of the ES events involve the skipping of a single exon, while the majority of these events involve the skipping of  $\geq 2$  exons (61%).

An example of the latter includes *dnmt1*, encoding DNA methyltransferase 1 responsible for maintenance of DNA methylation patterns, particularly during development (reviewed in [52]). *dnmt1* has approximately 34 exons with three isoforms annotated in zebrafish; however, our analysis detects several additional isoforms (Figure S6 in Additional file 1). Cufflinks predicts 9 *dnmt1* isoforms (FPKM > 3), but without full-length sequence data the number of biologically relevant isoforms remains uncertain. One new isoform of *dnmt1* is robustly expressed maternally (Additional file 3) and truncated due to skipping of 19 exons (validated by cDNA cloning and sequencing; data shown in Additional file 11). This skipping event removes most of the functional domains involved in DNA methyltransferase activity, but it retains a DMAP1 binding domain (Figure 4b). This domain binds the co-repressor DMAP1 and thereby confers DNMT1 with transcriptional repressor activity [53,54]. Interestingly, in *Xenopus* it has been shown that Dnmt1 is needed to repress transcription prior to the MBT [55] and this function does not rely on the cytosine methyltransferase catalytic activity of Dnmt1 [56]. Zebrafish may therefore be a valuable candidate to shed more light on the functional role of *dnmt1* and its isoforms. The novel *dnmt1* isoforms discovered in this study may bring clarity to the enigmatic issue of transcriptional repression during cleavage stages in zebrafish [11,57] and other species [58].

For the pluripotency-associated gene *pou5f1* (also known as *oct4*) we identify several new splicing events, all of which were confirmed by cDNA cloning and sequencing (sequences are shown in Additional file 12).



**Figure 4 Alternative splicing in the early embryo.** (a) The frequency of 4 different alternative splicing events: Exon skipping (ES), alternative acceptor site (AA), alternative donor site (AD) and intron retention (IR). ES is the most frequent AS event. In most ES events multiple exons are skipped. (b) Schematic representation of an annotated version (top) and a new (bottom) isoform for *dnmt1*. Black lines between the isoforms represent the conserved area. Functional domains (white bars) were defined by Pfam: DMAP1 binding domain (DMAP1-BD, PF06464), Cytosine specific methyltransferase replication foci domain (Cyt MeTrfase1 RFD, PF12047), CXXC zinc finger domain ("CX": Znf CXXC, PF02008), Bromo-adjacent homology domain (BAH, PF01426), C-5 cytosine-specific DNA methylase domain (C5\_ MeTrfase, PF00145). In the new isoform 19 exons are skipped, and all functional domains except a DMAP1 binding domain are lost. The novel isoform also have a longer 5'UTR relative to the annotated version. (c) Novel *pou5f1* isoforms. The annotated version of *pou5f1* has two Pfam domains; Homeobox domain (HB, PF00046) and POU (PF00157). We identify 3 novel alternative acceptor sites (arrows 1–3) all within the CDS of *pou5f1*. The first leads to a 3 nt deletion and removal of one glutamic acid, upstream of the POU; the second a 19 nt insertion causing a frame shift and a truncated Pou5f1 protein with both functional domains lost; the third event gives a 4 nt deletion which truncates the HB. We also detect a longer 3' UTR post-ZGA (arrow 4).

In zebrafish, there is only one annotated *pou5f1* transcript, while in human three arise from AS [59]. We observe novel *pou5f1* AS events originating from alternative acceptor sites at exons 2, 3 and 5 (Figure 4c; Figure S7 in Additional file 1). The first event leads to deletion of a nucleotide triplet coding for glutamic acid (validated by one cDNA clone sequence out of 12 random clones

analyzed), yet with no effect on the functional domains. The removal of a charged amino acid is however likely to affect the three-dimensional structure of Pou5f1, although functional consequences remain unknown at this stage. The second AS event is located 19 nt prior to the annotated exon 3 (validated by one cDNA clone sequence). This leads to a frame-shift and a truncated Pou5f1 protein,



removing both functional domains. Lastly, a 4 nt deletion causes a frame-shift in the exon 5 (validated by 5 of 12 cDNA clones sequences, suggesting higher abundance of this isoform) and leads to a truncation of the homeobox domain. Of note, we also find novel isoforms for other genes important for pluripotency, such as *klf4*, *sall4* and *nanog*. This indicates that several well-studied genes with central functions in pluripotency are still not fully annotated in zebrafish, and highlights a likelihood of uncovering novel intricacies about pluripotency control mechanisms.

#### Epigenetic marks associated with exon skipping

AS predominantly occurs co-transcriptionally, which may influence the regulation of splicing by changes of epigenetic signatures such as histone modifications and DNA methylation [24,25,60]. We used ChIP-seq to map H3K36me3 occupancy in post-ZGA embryos, and used published peak calling data for H3K4me3 and H3K27me3 [13] to investigate any association between these modified histones and single ES events. We extracted the number of reads spanning each skipped exon (this is direct evidence for skipping of an exon; see Methods); 1,668 single ES events have one or more reads supporting the skipping event, of which most only with a few reads spanning them. Focusing on 446 exons with >5 reads supporting the skipping of these exons at the post-MBT stage, we examined peaks of H3K4, H3K27 and H3K36 trimethylation in their vicinity (see Methods). H3K36me3 peaks and skipped exons co-localize in 194 of the 446 exons (some exons overlap with multiple H3K36me3 peaks), which represent a slight overrepresentation relative to a random set of regions (~145 overlaps on average when 446 exons were drawn at random from all exons a thousand times; Chi-squared test,  $p < 0.001$ ). H3K4me3 is also slightly overrepresented (113 overlaps versus 71 at random,  $p < 0.001$ ), whereas H3K27me3 is not overrepresented over skipped exons (~35 peaks overlapped both skipped and non-skipped exons).

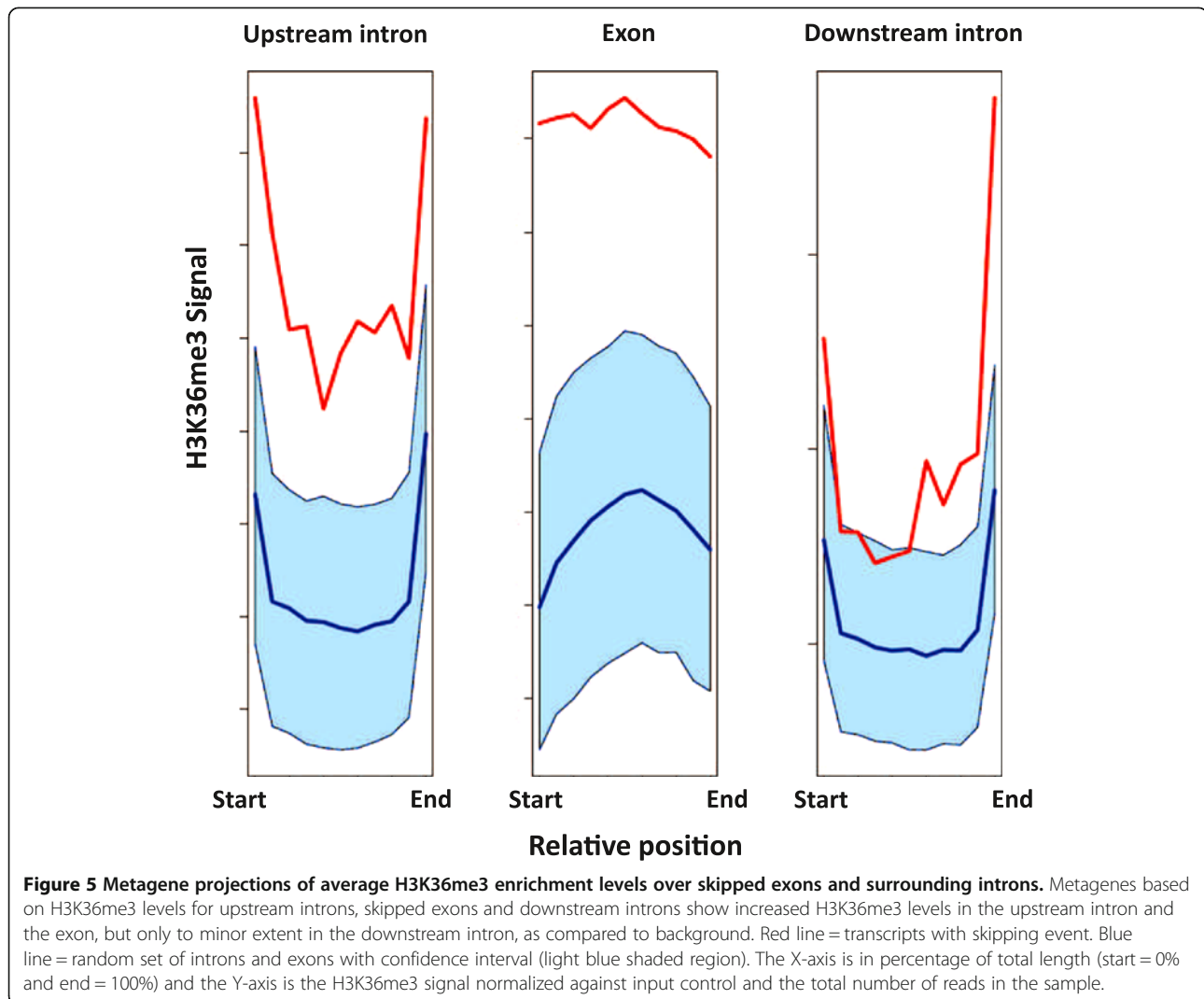
From metagene sections of average enrichment profiles containing the skipped exon as well as introns upstream and downstream of the skipping event, we find that both the upstream intron and the skipped exon display higher levels of H3K36me3 than average, determined from random selections of the entire set of introns and exons (Figure 5). Introns localized downstream of the skipped exons only show slightly higher H3K36me3 enrichment values than the average. Since H3K36me3 associates with expressed genes [61], a possible explanation for H3K36me3 enrichment on the skipped exons may be that the exons are only skipped in a subpopulation of cells in the embryo, reflecting distinct transcript profiles in individual cells. Isoforms with the exon included can thus be more prevalent and account for the observed

enrichment in H3K36me3. However as expected, globally in the cell populations examined, the skipped exons have on average substantially lower number of reads mapped to them than the rest of the exons (Figure S8 in Additional file 1). This shows that the H3K36me3 enrichment at skipped exons is not due to expression of an isoform with the exon included in the majority of the cells. Collectively, our observations are consistent with a role of H3K36me3 and H3K4me3 as possible facilitators, rather than determinants, of increased exon skipping [24,62].

#### Longer 3'UTRs in zygotic than in maternal transcripts

The TTS can be regulated either through termination of transcription [63] or via the 3'-end processing machinery [64]. Regardless of the mechanism, sequence motifs in the 3'UTR play an important role in mRNA stability, spatial distribution and translational efficiency [65]. We observe that zygotic transcripts frequently have longer 3'UTRs than corresponding maternal transcripts, as exemplified by *pou5f1* (Figure 4d, arrow 4; black and white arrow in Figure S7 in Additional file 1). To investigate the genome-wide frequency of longer 3'UTR in post-ZGA relative to pre-ZGA embryos, we extracted all last exons and counted the number of reads mapping to the first and second half of the last exon. Using these counts, we calculated a numeric score where higher values indicate a shift towards a longer 3' terminal exon post-ZGA (see Methods). We find that 1,249 last exons increase in length post-ZGA relative to pre-ZGA (Additional file 13). A metagene projection for these 1,249 last exons using RNA-seq coverage confirms the detection of more reads in the last half of the exons post-ZGA (Figure 6a). The observation of longer 3'UTRs can be interpreted in light of a recent report of shorter 3'UTRs of highly expressed transcripts [66]. We may infer from this report and our results that post-ZGA embryos produce transcripts with a more diversified control of turn-over kinetics, together with spatial and cell-type specific expression, than the oocyte. In *Drosophila melanogaster*, extension of 3'UTRs after the maternal-to-zygotic transition (coinciding with the MBT) have been reported [67]. Similarly in the mouse, the 3'UTR becomes progressively longer during embryonic development [68]. Furthermore, while this manuscript was in preparation, Ulitsky and colleagues [69] reported extended 3'UTRs in 419 genes after ZGA in zebrafish embryos, using poly(A)-position profiling by sequencing. Thus our findings validate previous reports in zebrafish and other species.

In addition to the extension of last exons, we also observe alternative last exon usage, which may lead to changes in the CDS, the 3'UTR or both. An example of the latter is apparent in *magi1*, encoding a PDZ domain protein; due to a switch in last exon usage between pre-

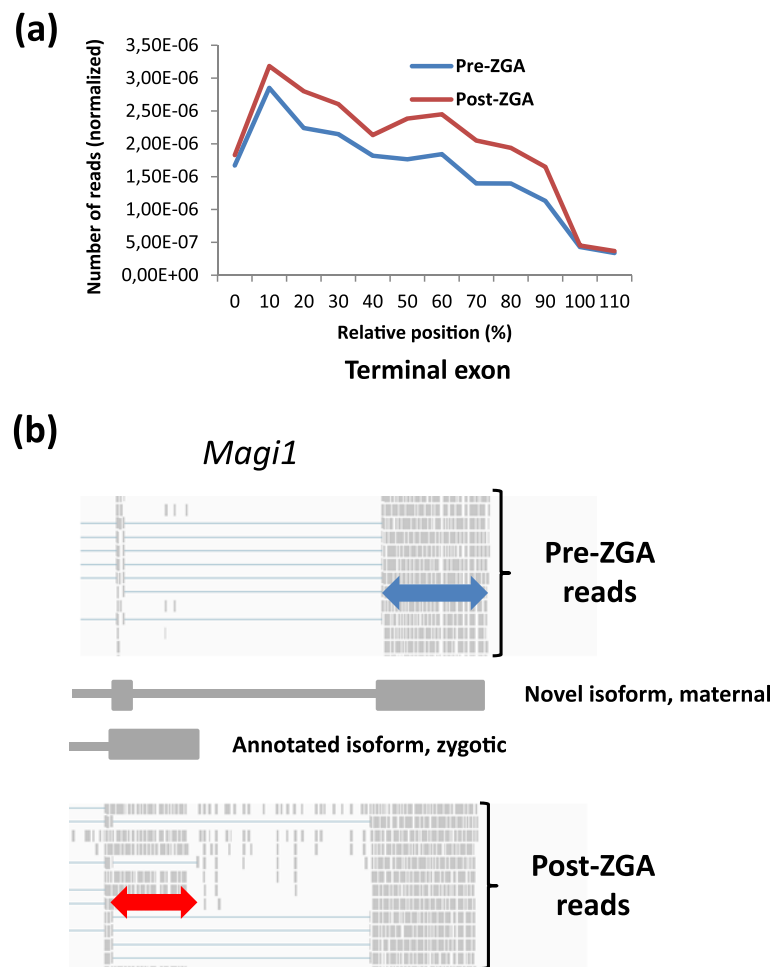


and post-ZGA, both the CDS and the 3'UTR are changed (Figure 6b; see also RT-qPCR validation in Figure S9 in Additional file 1). The 3'UTR of the post-ZGA-specific isoform contains three additional sequence domains absent in the maternal version, namely an internal ribosome entry site (IRES), a sex-lethal binding site (SXL-BS) and a K-box element. The latter two are known to often co-occur to create spatial and temporal expression patterns through negative regulation of translation [70,71]. In addition, although the PDZ domains are not directly affected, alteration in protein sequence might affect the protein 3D-structure and in particular the fifth PDZ domain located proximal to the C-terminus [72]. Interestingly, Magi1 interacts with  $\beta$ -catenin [73], another regulator of cell adhesion crucial for left-right asymmetry of the embryo [8,74]. Each PDZ domain often displays distinct protein specificity and strikingly,  $\beta$ -catenin binding to MAGI1 depends on the 5<sup>th</sup> PDZ domain, thus potentially linking the isoform switch

reported in our study to the regulation of Wnt signaling through  $\beta$ -catenin [75].

### Conclusions

The transcriptomes assembled in this study contain a high degree of complexity pertaining to TSSs, AS and TTSs. Most of the described novel isoforms encode ORFs with predicted significant functional protein domains, suggesting functional and valid transcripts with a role in development. We report several loci which display a switch-like behavior in isoforms expressed pre-ZGA and post-ZGA stages; these notably include *dazl* (TSS switch), *f11r* (exon skipping) and *magi1* (alternative last exon usage). Their changes in expression pattern coincide with key developmental processes such as onset of transcription upon ZGA, cell speciation and migration, suggesting functional roles for these isoforms. We also identify validated novel isoforms for *dnmt1* and *pou5f1*, important genes in development and pluripotency control.



**Figure 6 Extended 3' terminal exons and alternative last exons.** (a) Metagene representation of RNA-seq reads for the last exon in a set of transcripts detected with more coverage in the proximal half of the terminal exon post-ZGA (red line) as compared to pre-ZGA (blue line). X-axis is in percentage of total length and Y-axis is number of reads normalized against total number of reads in the samples. (b) The gene *magi1* displays a change in the expression of its isoforms pre- and post-ZGA with an alternative 3'UTR appearing post-ZGA (red double-arrow). Prior to the ZGA a novel isoform is expressed (blue double-arrow). This isoform is expressed also after the MBT but then together with the novel isoform.

Moreover, we find increased length of the 3' terminal exon and 3'UTR after the ZGA in over 1,000 transcripts. Analysis of histone marks suggests that H3K4me3 has little value in marking the major TSS in loci with several start sites. However, we find higher levels of H3K36me3 over skipped exons relative to all exons, suggesting a role of H3K36me3 in zebrafish AS control.

## Methods

### Bioinformatics

Bioinformatic analyses relied heavily on custom python scripts and the use of publically available tools such as SAMtools [76] and BEDtools [77], as well as general data handling in R [78]. Python scripts critical for the analyses are available as supplementary material

(Additional file 14). Data were visualized in the Integrative Genomics viewer [79].

### RNA-seq mapping and transcriptome assembly

Embryo collection and RNA sequencing was performed as previously described [5]. Raw data files are available from NCBI's short read archive (GSE22830). Reads were mapped using Bioscope 1.3 (with default settings) and TopHat 1.3.1 (with default options except:  $-a = 5$ ,  $-I = 50$ ,  $-F = 0$ ). The resulting SAM-files were merged and a python script (best\_alignment.py) used to ensure that a read were only present once. We performed transcriptome assembly using Cufflinks (version 1.2.1), with Ensembl annotation as guide (Zv9, annotation version 64), using default settings except;  $F = 0.0$ ,  $-j = 0.05$ ,  $-A = 0.08$ ,  $--3-overhang-tolerance = 100$ .

Additional programs in the Cufflinks pipeline were used to merge files, compare transcripts to known annotation and detect differences between the stages [80]. We created additional datasets using scripts; FPKM > 3 using `extract_gtf.py`, TSS's using `get_tss_bed.py` and TTS using `get_tts.py`.

### Sequence analysis

We extracted transcript sequences using `gffread` in Cufflinks. These were subjected to ORF prediction using `getORF` (with settings: `-minsize 150`, `-reverse_sequence No`, `-find 1`) from the EMBOSS suite [81]. The resulting protein sequences were processed by scripts and R handling to retrieve the longest predicted ORF. We used Pfam to identify functional domains (Pfam A) [31] and domains considered significant by Pfam were used. A script (`'domain_compare.py'`) was used to identify if domains were retained, lost or gained relative to the closest matching transcript. Detection of motifs in the 3'UTR was performed using UTRScan [82].

### Alternative splicing

We used ASTALAVISTA version 2.2 [51] to quantify alternative splicing events. This program takes as input a GTF file (FPKM > 3) and outputs all splicing events in the annotation together with a code representing type of event (exon skipping, intron retention, alternative acceptor site, alternative donor site). We used python scripts to extract and count events from the ASTALAVISTA output (`summary.py` and `events.py`).

### Detection of length changes in last exon

We designed a database of non-redundant last exons (`get_last_exon.py`) and after initial filtering (exons > 400 bp, FPKM > 3 in both samples,  $n = 10,349$  last exons) we used a script (`find_extended_utrs.py`) to identify exons with a length change between pre- and post-ZGA. This was achieved by counting the number of reads in the first and latter part of the exon. A score was calculated as follows:

$$Q = (p_2 - p_1) * \log((\text{FPKM}_{\text{pre}} + \text{FPKM}_{\text{post}}) / 2)$$

where  $p_1$  and  $p_2$  are the proportions of reads in the last part of the exon relative to the first part pre- and post-ZGA, respectively. Exons with a shift towards more reads in the last part post-ZGA will have a positive value. We weighted this value using the log-transformed average of the exons FPKM values pre- and post-ZGA, assuming that higher coverage gives a more robust estimate of change.

### Analysis of ChIP-seq data

Embryos for ChIP-seq were collected at 5.3 hpf and chromatin prepared as described [83]. In accordance with institutional, national and international guidelines for early

stage (0–5.3 hpf) zebrafish embryos no ethics committee approval was needed for any of the experiments performed in this study. Immunoprecipitation was performed using H3K36me3 (Diagenode 058–050, Denville, NJ) and ChIP enriched DNA and input samples prepared for sequencing according to Illumina standard protocol (#11257047 Rev.A). ChIP-seq reads were mapped using BWA [84] and peaks detected using CCAT 3.0 (Settings: fragment size = 200, sliding window size = 500, minimum score = 3.0) [85]. Intron and exons from single exon skipping events identified by ASTALAVISTA were extracted with a script (`extract_skipped_exon.py`) and BEDtools used to count the number of reads supporting each event. Events with >5 reads post-ZGA were considered further. We created databases of all introns and exons using the 'all' annotation. The number of overlapping peaks and skipped exons (>1 bp overlap between peak and exon skipping event: from start upstream intron to stop downstream intron) were calculated using BEDtools and the random control using a script (`bootstr_rand.py`). We utilized Repitools to construct meta-gens [86]. H3K4me3 and H3K27me3 peak data were from a recently published paper (Pauli et al., 2012). We identified potential bidirectional promoters using ChIPpeakAnno [87] with the H3K4me3 data [13] and the 'all' gene annotation file. The TSS of either gene in the pair had to be within 1000 bp of the middle of the H3K4me3 peak.

### PCR and cloning

Newly assembled and annotated isoforms were validated by RT-(q)PCR. RNA from pools of embryos at different stages was isolated using TRIzol reagent followed by RNA cleanup using Qiagen RNeasy MiniKit (Hilden, Germany) as previously described [88]. Control RNA of kanamycin (Promega #C1381, Madison, Wisconsin) was added prior to the RNA extraction and used as the reference control by qPCR. RNA was reverse transcribed using iScript Select cDNA synthesis Kit (BioRad #170-8896, Hercules, CA) according to the manufactures instructions. Primers were designed to cover specific sequences resulting from alternative splicing (*f11r*), alternative TSS (*dazl*), and alternative last exon usage (*magi1*) (for details see Supplementary Table 1 in Additional file 15). The dynamics of the isoforms was confirmed by RT-qPCR. qPCR was performed with the iCycler MyiQ real time PCR detection system and SYBR Green (BioRad, Hercules, CA). Primers pairs gave no signal in PCRs lacking template (data not shown). Relative expression was determined by the  $\Delta\Delta\text{-C}_T$  method. For *pou5f1*, both primers were designed in the sequence common for the predicted isoforms (Supplementary Table 1 in Additional file 15). Presence of the isoforms at 3.5 hpf was confirmed by TOPO TA cloning (Invitrogen K45-0001, Carlsberg, CA) and subsequent sequencing of randomly selected clones.

## Data access

ChIP-seq data for H3K36me3 and input are available under [NCBI GEO GSE40629]. Our RNA-seq data are available under [NCBI GEO GSE22830].

## Additional files

**Additional file 1:** Supplementary figures 1–9.

**Additional file 2:** Non-redundant gene annotations file containing a merger of all Ensembl and assembled transcripts.

**Additional file 3:** Expression values pre-ZGA and post-ZGA for all isoforms in the 'all' gene annotation.

**Additional file 4:** Gene annotation for robustly expressed isoform (>3 FPKM pre-ZGA or post-ZGA).

**Additional file 5:** Test results for whether loci with multiple TSSs exhibited a shift in TSS usage between pre-ZGA and post-ZGA.

**Additional file 6:** GO term analysis results from DAVID of genes marked with H3K4me3.

**Additional file 7:** GO term analysis results from DAVID of genes marked with H3K27me3.

**Additional file 8:** Anatomical ontology results from DAVID of genes marked with H3K4me3.

**Additional file 9:** Anatomical ontology results from DAVID of genes marked with H3K27me3.

**Additional file 10:** Results from a statistical test of whether genes exhibit a significant change in alternative splicing between pre-ZGA and post-ZGA.

**Additional file 11:** Results from sequencing of cloned *dnmt1* fragments.

**Additional file 12:** Results from sequencing of cloned *pou5f1* fragments.

**Additional file 13:** All last exons with evidence for an extended 3'UTR post-ZGA relative to the pre-ZGA.

**Additional file 14:** Different scripts used in the analyses.

**Additional file 15:** A list with description of the primers.

## Abbreviations

RNA-seq: RNA-sequencing; ChIP-seq: Chromatin immunoprecipitation sequencing; ZGA: Zygotic genome activation; MBT: Mid-blastula transition; CDS: Coding sequence; UTR: Untranslated region; TSS: Transcription start site; AS: Alternative splicing; TTS: Transcription termination site; FPKM: Fragments per kilobase per million mapped fragments; ORF: Open reading frame; GO: Gene ontology; FDR: False discovery rate; ES: Exon skipping; AD: Alternative donor site; AA: Alternative acceptor site; IR: Intron retention.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

HA designed the study, wrote the manuscript, made the figures and performed all bioinformatics analysis, including writing python scripts. OØ performed RT-qPCR, cDNA cloning and sequencing, contributed to figures and the manuscript. ISA performed ChIP experiments. LFM collected embryos, isolated RNA and made figures. SM designed the study, initiated the RNA-seq experiment and generated all the basic data. PC supervised ChIP-seq work, designed the study and wrote parts of the manuscript. PA designed the study, wrote the manuscript and supervised the work. All authors approved the final manuscript.

## Acknowledgements

We thank Torbjørn Rognes and Andrew Reiner for valuable discussions on bioinformatic analyses, and Timothy Hughes, Susanne Lorenz and Leonardo Meza-Zapeda for assistance on ChIP-seq and mapping of ChIP-seq reads. This work was supported by the Research Council of Norway, the Norwegian

Center for Stem Cell Research and Carlsberg foundation (OØ). HA has a PhD fellowship at the Norwegian School of Veterinary Science.

## Author details

<sup>1</sup>BasAM, Norwegian School of Veterinary Science, 0033 Dep, Oslo, Norway. <sup>2</sup>Stem Cell Epigenetics Laboratory, Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway. <sup>3</sup>Norwegian Center for Stem Cell Research, 0317Oslo, Norway. <sup>4</sup>Stem Cell and Developmental Biology, Genome Institute of Singapore, 60, Biopolis Street, #02-01, Genome 138672, Singapore.

Received: 26 October 2012 Accepted: 25 April 2013

Published: 15 May 2013

## References

1. Kane DA, Kimmel CB: The zebrafish midblastula transition. *Development* 1993, **119**:447–456.
2. Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EH: Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* 2008, **322**:1065–1069.
3. Ho RK, Kimmel CB: Commitment of cell fate in the early zebrafish embryo. *Science* 1993, **261**:109–111.
4. Mathavan S, Ballard WW, Kimmel SR, Ullmann B, Schilling TF: Stages of embryonic development of the zebrafish. *Dev Dyn* 1995, **203**:253–310.
5. Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SG, *et al*: Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* 2011, **21**:1328–1338.
6. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van DS, Inoue K, *et al*: Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 2006, **312**:75–79.
7. Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR, Govindarajan KR, *et al*: Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet* 2005, **1**:260–276.
8. Chan TM, Longabaugh W, Bolouri H, Chen HL, Tseng WF, Chao CH, *et al*: Developmental gene regulatory networks in the zebrafish embryo. *Biochim Biophys Acta* 2009, **1789**:279–298.
9. Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, *et al*: Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* 2010, **464**:922–926.
10. Lindeman LC, Andersen IS, Reiner AH, Li N, Aanes H, Ostrup O, *et al*: Pre-patterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation. *Dev Cell* 2011, **21**:993–1004.
11. Andersen IS, Reiner AH, Aanes H, Alestrom P, Collas P: Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biol* 2012, **13**:R65.
12. Andersen IS, Ostrup O, Lindeman LC, Aanes H, Reiner AH, Mathavan S, *et al*: Epigenetic complexity during the zebrafish mid-blastula transition. *Biochem Biophys Res Commun* 2012, **417**:1139–1144.
13. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, *et al*: Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* 2012, **22**:577–591.
14. Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J: The zebrafish transcriptome during early development. *BMC Dev Biol* 2011, **11**:30.
15. Rosenfeld MG, Amara SG, Roos BA, Ong ES, Evans RM: Altered expression of the calcitonin gene associated with RNA polymorphism. *Nature* 1981, **290**:63–65.
16. Rosenfeld MG, Lin CR, Amara SG, Stolarsky L, Roos BA, Ong ES, *et al*: Calcitonin mRNA polymorphism: peptide switching associated with alternative RNA splicing events. *Proc Natl Acad Sci U S A* 1982, **79**:1717–1721.
17. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008, **40**:1413–1415.
18. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, *et al*: Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008, **456**:470–476.
19. Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, Sung HK, *et al*: An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 2011, **147**:132–146.
20. Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, *et al*: Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res* 2011, **21**:342–348.

21. Cooper TA, Wan L, Dreyfuss G: **RNA and disease.** *Cell* 2009, **136**:777–793.
22. Pickering BM, Willis AE: **The implications of structured 5' untranslated regions on translation and disease.** *Semin Cell Dev Biol* 2005, **16**:39–47.
23. Andreassi C, Riccio A: **To localize or not to localize: mRNA fate is in 3'UTR ends.** *Trends Cell Biol* 2009, **19**:465–474.
24. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T: **Regulation of alternative splicing by histone modifications.** *Science* 2010, **327**:996–1000.
25. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature* 2011, **479**:74–79.
26. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al: **Deciphering the splicing code.** *Nature* 2010, **465**:53–59.
27. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotechnol* 2010, **28**:511–515.
28. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**:1105–1111.
29. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: **Improving RNA-Seq expression estimates by correcting for fragment bias.** *Genome Biol* 2011, **12**:R22.
30. Roberts A, Pimentel H, Trapnell C, Pachter L: **Identification of novel transcripts in annotated genomes using RNA-Seq.** *Bioinformatics* 2011, **27**:2325–2329.
31. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**:D290–D301.
32. Leoni G, Le PL, Ferre F, Raimondo D, Tramontano A: **Coding potential of the products of alternative splicing in human.** *Genome Biol* 2011, **12**:R9.
33. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al: **Genome-wide analysis of mammalian promoter architecture and evolution.** *Nat Genet* 2006, **38**:626–635.
34. Xu EY, Moore FL, Pera RA: **A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans.** *Proc Natl Acad Sci U S A* 2001, **98**:7414–7419.
35. Chen J, Melton C, Suh N, Oh JS, Horner K, Xie F, et al: **Genome-wide analysis of translation reveals a critical role for deleted in azoospermia-like (Dazl) at the oocyte-to-zygote transition.** *Genes Dev* 2011, **25**:755–766.
36. Haston KM, Tung JY, Reijo Pera RA: **Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro.** *PLoS One* 2009, **4**:e5654.
37. Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK: **The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells.** *EMBO J* 2005, **24**:2656–2666.
38. Takeda Y, Mishima Y, Fujiwara T, Sakamoto H, Inoue K: **DAZL relieves miRNA-mediated repression of germline mRNAs by controlling poly(A) tail length in zebrafish.** *PLoS One* 2009, **4**:e7513.
39. Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, et al: **Methylation of histone H3 Lys 4 in coding regions of active genes.** *Proc Natl Acad Sci U S A* 2002, **99**:8695–8700.
40. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, et al: **Active genes are tri-methylated at K4 of histone H3.** *Nature* 2002, **419**:407–411.
41. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**:553–560.
42. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G: **Genome regulation by polycomb and trithorax proteins.** *Cell* 2007, **128**:735–745.
43. Bradford Y, Conlin T, Dunn N, Fashena D, Frazer K, Howe DG, et al: **ZFIN: enhancements and updates to the Zebrafish Model Organism Database.** *Nucleic Acids Res* 2011, **39**:D822–D829.
44. Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Francois KJ, Stunnenberg HG, et al: **A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos.** *Dev Cell* 2009, **17**:425–434.
45. Bogdanovic O, van Heeringen SJ, Veenstra GJ: **The epigenome in early vertebrate development.** *Genesis* 2012, **50**:192–206.
46. Mandell KJ, Parkos CA: **The JAM family of proteins.** *Adv Drug Deliv Rev* 2005, **57**:857–867.
47. Hamazaki Y, Itoh M, Sasaki H, Furuse M, Tsukita S: **Multi-PDZ domain protein 1 (MUPP1) is concentrated at tight junctions through its possible interaction with claudin-1 and junctional adhesion molecule.** *J Biol Chem* 2002, **277**:455–461.
48. Gallardo VE, Liang J, Behra M, Elkhouloun A, Villablanca EJ, Russo V, et al: **Molecular dissection of the migrating posterior lateral line primordium during early development in zebrafish.** *BMC Dev Biol* 2010, **10**:120.
49. Barone V, Heisenberg CP: **Cell adhesion in embryo morphogenesis.** *Curr Opin Cell Biol* 2012, **24**:148–153.
50. Siddiqui M, Sheikh H, Tran C, Bruce AE: **The tight junction component Claudin E is required for zebrafish epiboly.** *Dev Dyn* 2010, **239**:715–722.
51. Foissac S, Sammeth M: **ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets.** *Nucleic Acids Res* 2007, **35**:W297–W299.
52. Jurkowska RZ, Jurkowski TP, Jeltsch A: **Structure and function of mammalian DNA methyltransferases.** *Chembiochem* 2011, **12**:206–222.
53. Rountree MR, Bachman KE, Baylin SB: **DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci.** *Nat Genet* 2000, **25**:269–277.
54. Lee GE, Kim JH, Taylor M, Muller MT: **DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair.** *J Biol Chem* 2010, **285**:37630–37640.
55. Stancheva I, Meehan RR: **Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos.** *Genes Dev* 2000, **14**:313–327.
56. Dunican DS, Ruzov A, Hackett JA, Meehan RR: **xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function.** *Development* 2008, **135**:1295–1302.
57. Schier AF: **The maternal-zygotic transition: death and birth of RNAs.** *Science* 2007, **316**:406–407.
58. Bogdanovic O, Long SW, van Heeringen SJ, Brinkman AB, Gomez-Skarmeta JL, Stunnenberg HG, et al: **Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis.** *Genome Res* 2011, **21**:1313–1327.
59. Wang X, Dai J: **Concise review: isoforms of OCT4 contribute to the confusing diversity in stem cell biology.** *Stem Cells* 2010, **28**:885–893.
60. Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, et al: **Single-molecule imaging of transcriptionally coupled and uncoupled splicing.** *Cell* 2011, **147**:1054–1065.
61. Wagner EJ, Carpenter PB: **Understanding the language of Lys36 methylation at histone H3.** *Nat Rev Mol Cell Biol* 2012, **13**:115–126.
62. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T: **Epigenetics in alternative pre-mRNA splicing.** *Cell* 2011, **144**:16–26.
63. Kuehner JN, Pearson EL, Moore C: **Unravelling the means to an end: RNA polymerase II transcription termination.** *Nat Rev Mol Cell Biol* 2011, **12**:283–294.
64. Millevoi S, Vagner S: **Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation.** *Nucleic Acids Res* 2010, **38**:2757–2774.
65. Proudfoot NJ: **Ending the message: poly(A) signals then and now.** *Genes Dev* 2011, **25**:1770–1782.
66. Ji Z, Luo W, Li W, Hoque M, Pan Z, Zhao Y, et al: **Transcriptional activity regulates alternative cleavage and polyadenylation.** *Mol Syst Biol* 2011, **7**:534.
67. Hilgers V, Perry MW, Hendrix D, Stark A, Levine M, Haley B: **Neural-specific elongation of 3' UTRs during Drosophila development.** *Proc Natl Acad Sci U S A* 2011, **108**:15864–15869.
68. Ji Z, Lee JY, Pan Z, Jiang B, Tian B: **Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.** *Proc Natl Acad Sci U S A* 2009, **106**:7028–7033.
69. Ulitsky I, Shkumatava A, Jan C, Subtelny AO, Koppstein D, Bell G, et al: **Extensive alternative polyadenylation during zebrafish development.** *Genome Res* 2012.
70. Turi A, Loglisci C, Salvemini E, Grillo G, Malerba D, D'Elia D: **Computational annotation of UTR cis-regulatory modules through Frequent Pattern Mining.** *BMC Bioinformatics* 2009, **10**(Suppl 6):S25.
71. Lai EC, Burks C, Posakony JW: **The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts.** *Development* 1998, **125**:4077–4088.
72. Stetefeld J, Ruegg MA: **Structural and functional diversity generated by alternative mRNA splicing.** *Trends Biochem Sci* 2005, **30**:515–521.
73. Dobrosotskaya IY, James GL: **MAGI-1 interacts with beta-catenin and is associated with cell-cell adhesion structures.** *Biochem Biophys Res Commun* 2000, **270**:903–909.

74. Zhang M, Zhang J, Lin SC, Meng A: **beta-Catenin 1 and beta-catenin 2 play similar and distinct roles in left-right asymmetric development of zebrafish embryos.** *Development* 2012, **139**:2009–2019.
75. Ranganathan R, Ross EM: **PDZ domain proteins: scaffolds for signaling complexes.** *Curr Biol* 1997, **7**:R770–R773.
76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078–2079.
77. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26**:841–842.
78. R Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R foundation for statistical computing; 2012. ISBN 3-900051-07-0, URL<http://www.R-project.org/>.
79. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al: **Integrative genomics viewer.** *Nat Biotechnol* 2011, **29**:24–26.
80. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7**:562–578.
81. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276–277.
82. Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, et al: **UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs.** *Nucleic Acids Res* 2010, **38**:D75–D80.
83. Lindeman LC, Vogt-Kielland LT, Alestrom P, Collas P: **Fish'n ChIPs: chromatin immunoprecipitation in the zebrafish embryo.** *Methods Mol Biol* 2009, **567**:75–86.
84. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754–1760.
85. Xu H, Handoko L, Wei X, Ye C, Sheng J, Wei CL, et al: **A signal-noise model for significance analysis of ChIP-seq with negative control.** *Bioinformatics* 2010, **26**:1199–1204.
86. Statham AL, Strbenac D, Coolen MW, Stirzaker C, Clark SJ, Robinson MD: **Repitools: an R package for the analysis of enrichment-based epigenomic data.** *Bioinformatics* 2010, **26**:1662–1663.
87. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, et al: **ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data.** *BMC Bioinformatics* 2010, **11**:237.
88. Peterson SM, Freeman JL: **RNA isolation from embryonic zebrafish and cDNA synthesis for gene expression analysis.** *J Vis Exp* 2009.

doi:10.1186/1471-2164-14-331

**Cite this article as:** Aanes et al.: Differential transcript isoform usage pre- and post-zygotic genome activation in zebrafish. *BMC Genomics* 2013 **14**:331.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

