THE
EMBO
JOURNAL

Manuscript EMBO-2014-90648

The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3' processing

Ken-ichiro Abe, Ryoma Yamamoto, Vedran Franke, Minjun Cao, Yutaka Suzuki, Masataka Suzuki, Kristian Vlahovicek, Petr Svoboda, Richard M. Schultz and Fugaku Aoki

Corresponding author: Fugaku Aoki, The University of Tokyo

Review timeline:

Submission date:	25 November 2014
Editorial Decision:	27 January 2015
Revision received:	13 February 2015
Editorial Decision:	23 February 2015
Revision received:	09 March 2015
Accepted:	19 March 2015

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

Editor: Anne Nielsen

1st Editorial Decision 27 January 2015

Thank you for submitting your manuscript for consideration by the EMBO Journal and my apologies for the very unusual delay in the decision process in this case. We have now finally heard back from all three referees and their comments are shown below.

As you will see from the reports, all referees express interest in the findings reported in your manuscript and support publication in The EMBO Journal following minor revision:

-> Ref#1 raises a few issues on the presentation and format that should be easily amended.
-> Ref#2 asks you to elaborate on the analysis to better contrast the present work with the findings from Park et 2013 (computational work only)
-> Ref#3 has a few comments to the splice analysis and asks for database deposition of the sequencing data obtained (a feature that is in any case required by the journal prior to publication of the work)

Given the referees' positive recommendations, I would like to invite you to submit a revised version of the manuscript, addressing the comments of the reviewers as outlined above. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version.

Thank you for the opportunity to consider your work for publication and my apologies again for the delay in the review process here. I look forward to your revision.

Referee #1:

I have no substantial comments; this is well executed and described piece of research. Results are not all that surprising, promiscuous transcription in one-cell stage mouse embryos was suspected. However this paper provides much needed insights and observation of substantial promoter independent transcription is interesting as is observation of deficient termination and processing. The authors provide plausible speculation to explain this phenomenon though one suspects that transcription simply occurs because it can in the condition of global chromatin reorganization.
Minor comments:
1. This is not so minor, references need complete editing to confirm to the format of the journal. It is just as easy to present references properly and one wonders why in this otherwise well presented manuscript the references are completely messed up. EndNote glitch?
2. Page 19, line 5; delete "by"
3. Page 37, line 5; there are no asterisks in the Figure 4E, please add.
4. Page 37, lines 18 - 24 and Figure 5C; The authors repeatedly imply that MII oocytes and 1-cell embryos treated with DRB are rather similar in terms of transcription. Yet number of non-overlapping, specific transcripts (upregulated in nontreated 1-cell embryos) is essentially identical to the number of overlapping transcripts (2501 v. 2688). Any comments? What does the black circle in Figure 5C represents?
5. Page 37, lines 24 - 25 and Figure 5D; more details in description are needed. Where in Venn diagram are these two genes? In line 25 delete "and".
6. Page 37, line 25 and page 38, line 1. See comment 5.
7. Page 39, line 12; "mas" should be "maps"

Referee #2:

Review of EMBO Journal EMBOJ-2014-90648 by K. Abe et al., (Aoki lab, and colleagues)

The authors are interested in understanding zygotic genome activation in the mouse. They provide here new work involving comparisons of RNAseq datasets derived from DRB treated or untreated 1-cell embryos, and comparisons to MII oocytes and later stages of embryogenesis. A key feature of the work is the isolation of total RNA, rather than polyA-selected, which allows them to explore features related to splicing and 3' end formation. They also examine parthenotes, allowing them to determine which features in the work involve solely the maternal genome, and further test the impact of inhibiting transcription or replication, through the use of inhibiting drugs. Although there are some very strong aspects of the design and execution of this work, it must also be considered in light of a prior study by Shirahige and colleagues (Park et al., Genes Dev 2013), whose design and execution had many similar features. However, the authors of the current paper conduct several unique modes of analysis that I think provides a set of unique insights of interest, especially in regard to splicing and 3' end formation.

Overall, the work is well designed and executed. The manuscript is well written, the Discussion thoughtful, and the display items are clear and relevant. The key to this paper is the extent to which it provides new insights that extend beyond the Park et al., paper. The first insight is their analysis of intergenic transcription that is DRB sensitive - the 'grass in the forest' - as they call it, and the notion that it may be due to the lack of repressive chromatin that has formed. Here, I would like the authors to check whether there is a statistical overlap between these loci in the 1-cell stage, and the extensive intergenic transcription seen during gametogenesis in the male, where chromatin is also likely less repressive or being removed (e.g. in spermatocytes and round-to-elongating spermatids). There are a couple of papers and datasets available that perform RNAseq at those stages, and a comparison would be interesting. It might be that a subset relies on the ATF and CREM factors (in the male) and a subset represents the same 'lack of promoter element' transcription seen here. In addition, is there statistical overlap here between these sequences and the ever growing locations of enhancer RNAs (eRNA) transcription? Related to this, Park et al., paper noted statistical enrichment of Foxd1, Nkx2-5, Sox18, Myod1, and Runx1 binding sites at 1-cell driven genes - do the authors see these upstream of their intergenic transcription? It would also add to the paper to have an analysis/discussion of satellite sequence transcription, as it is associated with heterochromatin regulation.

I know that the authors are emphasizing the unique aspects of their paper, but it would be helpful to have (perhaps in the Supplemental Data) more of a comparison to the Park et al., paper, just to know where there is high agreement or differences. It would be a service to those interested.

Overall, this is a strong paper both in datasets and analysis. The orchestration of zygotic genome activation is an area of general and high interest, and the authors have revealed new features that should be of real interest to the many - especially the lack of full splicing and 3' end formation, and the notable spurious intergenic transcription, which may create noncoding RNAs with functions (among many possibilities). For these reasons, I view the paper favorably for EMBO Journal, and request the only the minor/moderate additional analyses listed above.

Referee #3:

In this study, Abe and colleagues employed total RNA sequencing in metaphase II-arrested eggs, 1-cell embryos and 2-cell and 4-cell embryos, morulae, and blastocysts to identify sequences transcribed in 1-cell embryos and to obtain an insight into mechanisms that govern their expression. The authors show that pervasive transcription occurs in intergenic regions
including many transposons and their genomic flanks. Interestingly, transcription can occur independently of defined core-promoter elements. Furthermore Abe and coworkers show that mRNAs transcribed at the 1-cell stage are mostly non-functional because their 3' end processing and splicing are highly inefficient.

This is a well written manuscript covering an interesting and timely topic, regulation of gene expression, in early mammalian development. The experimental and analytical work is sound well executed. The authors confirm with their high sequencing data of total RNA some features of previously published observations that 1-cell embryos are transcriptionally permissive. In addition, Abe and coworkers find widespread occurrence of individual, rarely overlapping, low coverage reads (referred to as CPMs) in 1-cell embryos. The CPM reads appeared in gene-rich regions but were also found in intergenic regions. The CPM reads are reproducible and their presence is strongly reduced inhibition of replication. The generated sequencing data seems to date back some time since the first analysis was done with Illumina 36nt and 76nt reads. Today's technology would most certainly provide more coverage. Nevertheless, the authors made a remarkable observation using reporter constructs in the 1-cell embryo. On one hand the vector contains a cryptic promoter sequence upstream of the luciferase gene and secondly TSS sites of the alternatively initiated transcripts was located upstream of a transcriptional pause site with a consensus polyA signal suggesting inefficient transcriptional termination. Using additional reporter assays the authors show that cryptic initiation of transcription can occur without a specific promoter element in 1-cell embryos.
In addition, Abe and coworkers provide evidence that transcripts from genes transcribed in 1-cell embryos are not processed properly (i.e., neither spliced nor terminated correctly). Splicing deficiency was confirmed by RT-PCR and microinjected Ftz pre-mRNA.
Are the transcribed flanking regions of MuERV-L transposons mostly downstream? Could this observation be correlated to the 3' processing deficiency.

Minor point:
Since the HTS data likely is an excellent resource, the authors should ensure that the submitted sequencing files carry the same name or ID as the samples listed in Table S1. Unless the reviewer missed it, an accession number for the submission of HTS data is lacking.

1st Revision - authors' response 13 February 2015

Response to reviewers

Referee #1:

1. *This is not so minor, references need complete editing to confirm to the format of the journal. It is just as easy to present references properly and one wonders why in this otherwise well presented manuscript the references are completely messed up. EndNote glitch?*

We apologize for overlooking inconsistent use of full and abbreviated journal names – the problem has been corrected. The references were imported into Endnote directly from Pubmed Online and formatted with the EndNote default EMBO J. filter. The problem seemed to be in the default option in the journal abbreviation style.

2. Page 19, line 5; delete "by"

Corrected.

3. Page 37, line 5; there are no asterisks in the Figure 4E, please add.

Asterisks were added.

4. Page 37, lines 18 - 24 and Figure 5C; The authors repeatedly imply that MII oocytes and 1-cell embryos treated with DRB are rather similar in terms of transcription. Yet number of non-overlapping, specific transcripts (upregulated in nontreated 1-cell embryos) is essentially identical to the number of overlapping transcripts (2501 v. 2688). Any comments? What does the black circle in Figure 5C represents?

We wish to clarify the Venn diagram in Fig. 5C and similarity of MII and 1C+DRB samples:

(1) We realized an error in the red circle label. Which should have read 1C/MII. This error is now corrected and we apologize for the mistake.

(2) The Venn diagram in Fig. 5C shows relative differences with a selected fold-change cut-off, hence these data are not accurately depicting similarity between MII oocytes and 1-cell embryos treated with DRB in terms of transcription. To visualize similarity among samples in terms of transcription, we used repeat-masked intron-mapping data and produced a heatmap depicting the 1-spearman correlation coefficient for all samples (now shown as Fig. S6B). The rationale was to make a comparison based on putative nascent transcripts to eliminate the impact of mature maternal mRNAs and their degradation (as shown in Fig. 1B, which is based on exon-mapping reads).

(3) Data **outside of the black circle** pointed out by the reviewer (1602+899=2501 vs. 2688) are noisy. We observed that using a 4-fold-change calculated from completely unrestricted intronic RPKMs yielded data with considerable noise. Therefore, we empirically determined filtering values for low-level expression observed close to the background. Accordingly, we filtered for genes with a minimal intronic signal in MII oocytes (genes with ≤0.04 intron RPKM in MII). However, employing just this restriction was problematic because when an intronic RPKM for a gene in MII (or 1C+DRB) was 0 or very close to it, a gene could easily produce 4-fold increase in intronic RPKM values even with almost no intronic reads. Therefore, we empirically determined a minimal value 0.12 RPKM for intronic reads in 1-cell embryos. Under these filtering conditions, there is an excellent overlap between genes showing in 1C/MII and 1C/1C+DRB ≥ 4-fold up-regulation (4039 genes) whereas there are only 131 non-overlapping genes.

(4) We slightly modified the labeling of the black circle to be consistent with labeling of blue and red circles. The labeling of the black circle now reads:

1C ≥0.12 (intron RPKM)
MII ≤ 0.04 (intron RPKM)

5. Page 37, lines 24 - 25 and Figure 5D; more details in description are needed. Where in Venn diagram are these two genes? In line 25 delete "and".

Sord and Slc10a1 are among the 4039 genes, see also Table S3 (genes highlighted in red). The word "and" was deleted and the figure legend was modified to indicate that the genes are among the 4039 genes contained in the center of the Venn diagram.

6. Page 37, line 25 and page 38, line 1. See comment 5.

The figure legend was modified.

7. Page 39, line 12; "mas" should be "maps"

Corrected

Referee #2:

1. I would like the authors to check whether there is a statistical overlap between these loci in the 1-cell stage, and the extensive intergenic transcription seen during gametogenesis in the male, where chromatin is also likely less repressive or being removed (e.g. in spermatocytes and round-to-elongating spermatids). There are a couple of papers and datasets available that perform RNAseq at those stages, and a comparison would be interesting. It might be that a subset relies on the ATF and CREM factors (in the male) and a subset represents the same 'lack of promoter element' transcription seen here.

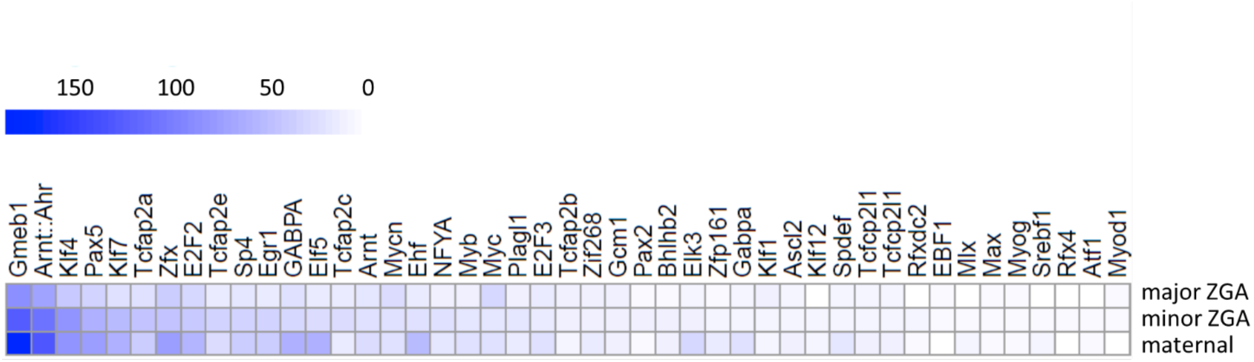
We examined distribution of transcripts identified in spermatocytes and spermatids by Hamoud et al. (Cell Stem Cell 2014, 15, 239–253) who performed at ~200M depth 50nt single-end, stranded, ribo- high throughput sequencing. In agreement with reviewer's comment, at this depth, we observed genome-wide intergenic transcription, which was more pronounced in spermatids. However, the character of spermatogenic intergenic transcription (intergenic localized bursts of transcription, long transcribed regions, and retrotransposon association) was different from that observed in 1-cell embryos. Given the technical differences between the Hamoud et al. and our datasets, and our inability to precisely define transcription start sites in 1-cell embryos, we believe that a comprehensive comparative analysis of transcription in spermatids and 1-cell embryos is beyond the scope of this revision.

2. Is there statistical overlap here between these sequences and the ever growing locations of enhancer RNAs (eRNA) transcription?

We attempted to analyze enhancers in several different ways but we could not find any sound statistical overlap between 1-cell transcribed loci and enhancers. There are several reasons why this analysis could not yield sound statistically significant data. First, our data cannot be used to predict bi-directional enhancer transcription because our sequencing is not strand-specific and the depth is insufficient to define reliably transcription start sites for intergenic transcription. Second, we cannot correlate intergenic sequence tags with positions of enhancers for early development because they are not mapped. We tested for a correlation of intergenic sequence tags with putative liver and ESC enhancers but did not obtain any useful insights.

3. Park et al., paper noted statistical enrichment of Foxd1, Nkx2-5, Sox18, Myod1, and Runx1 binding sites at 1-cell driven genes - do the authors see these upstream of their intergenic transcription?

Unfortunately, we could not analyze statistical enrichment of transcription factor binding sites upstream of intergenic transcription because the depth of our sequencing data does not allow us to distinguish between true transcription start sites and termini of sequenced fragments of longer intergenic transcripts. Instead, we used the 4039 gene list (for which we at least could obtain putative promoter coordinates) and analyzed TF binding sites using JASPAR TF-binding motifs (see below). The analysis could not be replicated exactly as in Park et al. because their paper does not contain sufficient detail to exactly reproduce their results.



Promotor binding sites analysis. Shown is a JASPAR TF-binding motif analysis for maternal genes (4000 genes with the highest expression in MII), minor ZGA (4039 genes from Fig. 5C), and major ZGA (4000 genes with the highest 2-cell/MII expression ratio). The color-scale indicates -log10(p-value). Of note is that except for Myc, there is not strong zygotic enrichment of specific TF-binding motifs.

Thus, taken together, our data is of insufficient depth to perform a robust analysis of transcription factor binding upstream of intergenic transcripts.

4. It would also add to the paper to have an analysis/discussion of satellite sequence transcription, as it is associated with heterochromatin regulation.

Analysis of satellite sequences turned out to be problematic and inconclusive. Satellite sequences were not present in the mm9 genome, which was used for mapping the 76PE data. Although the satellite class is present in the Repeatmasker filter, these sequences do not match Genbank mouse satellite repeat sequences (X06893-9 from Nucleic Acids Res, 1988, 16(3): 1201 and AY439017-8 from Chromosome Res, 2005, 13 (1), 9-25). Repeat masker-identified satellite sequences showed a steady decline; their abundance was ~ 730 RPMs in MII eggs and then declined (630, 660, 530, and 320 RPM in 1C, 1C+DRB, 2C, and 4C libraries, respectively). In parallel, we mapped 76PE data on Genbank sequences listed above: AY439017-8 yielded no sequence hits, and X06893-9 yielded 225, 147, 67, 398, and 441 sequence hits in MII, 1C, 1C+DRB, 2C, and 4C libraries. These data suggest that a maternal pool of satellite sequences is being degraded and replaced with zygotic transcripts. Because of MII numbers exceed those in 1-cell embryos, we cannot formally distinguish between 1-cell transcription (as DRB treatment would suggest) and accelerated satellite degradation in the absence of transcription (an alternative explanation). Therefore, we decided not to include satellite repeat analysis into the manuscript because it would not contribute new knowledge to the story.

5. I know that the authors are emphasizing the unique aspects of their paper, but it would be helpful to have (perhaps in the Supplemental Data) more of a comparison to the Park et al., paper, just to know where there is high agreement or differences. It would be a service to those interested.

We appreciate the Reviewer's sentiment, which is directed to helping the reader, but believe that the manuscript does contain a large volume of data comparing our results with Park's paper in the supplement. What follow is a list of these comparisons:

Fig. S1B – read distribution along the chromosome 1 in MII and 1-cell samples (SOLiD 50SE = Park, Illumina 76PE = our data).

Fig. S2B – 50SE rows are derived from sequencing data from Park et al.

Fig. S7B – 50SE SOLiD rows are data derived from sequencing data from Park et al.

Fig. S8 – this is the same analysis as in Fig. 6A-C but performed on sequencing data from Park et al.

For the revision, we added a set of transcriptome changes comparisons between Park et al and our data. These are included as Fig.S1C. We opted for comparing fold-changes identified within each sequencing platform to reduce the effect of the sequencing platform as much as possible.

Referee #3:

Since the HTS data likely is an excellent resource, the authors should ensure that the submitted sequencing files carry the same name or ID as the samples listed in Table S1. Unless the reviewer missed it, an accession number for the submission of HTS data is lacking.

The data were deposited into the ArrayExpress database under reference # E-MTAB-2950 and will be released when the paper is published. We apologize for omitting the accession number from the submitted manuscript; the accession number is now added in the Methods section of the manuscript.

2nd Editorial Decision 23 February 2015

Thank you for submitting a revised version of your manuscript, addressing and commenting on the concerns raised by the three referees. I am happy to inform you that your study can now in principle be accepted for publication in The EMBO Journal. However, before we can officially accept the manuscript and transfer your files to our production team there are a few editorial issues concerning text and figures that I would ask you to address:

-> Please provide a short paragraph in the manuscript text file stating author contributions and possible conflict of interest.

-> Please extend the discussion to reflect/include comments on the suggestions made by ref #2 to look at overlap with enhancer transcription and transcription factor binding sites. I do realize the negative outcome of the analysis and understand that you do not wish to overstate claims from your dataset, but I think it would nonetheless be useful to the reader to explain that you have attempted to map these different elements in your HTS data.

-> I noticed a few instances in the discussion where you are referring to 'unpublished results' regarding histone mobility in the 1-cell embryo as well as growth arrest at the 2-cell stage in the presence of DRB. We generally do not allow references to 'data not shown' in papers published in The EMBO Journal and I would therefore ask you to include the relevant data as supplemental information, if possible.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal, I look forward to your final revision.