Potential sperm contributions to the murine zygote predicted by \textit{in silico} analysis

Panagiotis Ntostis\textsuperscript{1,2}, Deborah Carter\textsuperscript{3}, David Iles\textsuperscript{1}, John Huntriss\textsuperscript{1}, Maria Tzetis\textsuperscript{2} and David Miller\textsuperscript{1}

\textsuperscript{1}Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, West Yorkshire, UK, \textsuperscript{2}Department of Medical Genetics, St. Sophia’s Children Hospital, School of Medicine, National and Kapodistrian University of Athens, Athens, Attiki, Greece and \textsuperscript{3}Leeds Institute of Molecular Medicine, University of Leeds, Leeds, West Yorkshire, UK

Correspondence should be addressed to D Miller; Email: d.miller@leeds.ac.uk

Abstract

Paternal contributions to the zygote are thought to extend beyond delivery of the genome and paternal RNAs have been linked to epigenetic transgenerational inheritance in different species. In addition, sperm–egg fusion activates several downstream processes that contribute to zygote formation, including PLC zeta-mediated egg activation and maternal RNA clearance. Since a third of the preimplantation developmental period in the mouse occurs prior to the first cleavage stage, there is ample time for paternal RNAs or their encoded proteins potentially to interact and participate in early zygotic activities. To investigate this possibility, a bespoke next-generation RNA sequencing pipeline was employed for the first time to characterise and compare transcripts obtained from isolated murine sperm, MII eggs and pre-cleavage stage zygotes. Gene network analysis was then employed to identify potential interactions between paternally and maternally derived factors during the murine egg-to-zygote transition involving RNA clearance, protein clearance and post-transcriptional regulation of gene expression. Our \textit{in silico} approach looked for factors in sperm, eggs and zygotes that could potentially interact co-operatively and synergistically during zygote formation. At least five sperm RNAs (Hdac11, Fbxa2, Map1lc3a, Pcbp4 and Zfp821) met these requirements for a paternal contribution, which with complementary maternal co-factors suggest a wider potential for extra-genomic paternal involvement in the developing zygote.

Reproduction (2017) 154 777–788

Introduction

Assuming fertilisation is successful, spermatozoal entry into the egg triggers a series of events that ends with the transformation of the terminally differentiated egg into the totipotent zygote. Alongside the paternal genome, the sperm also delivers non-genomic factors including the microtubule organising centre or centriole (excepting rodents), the oocyte-activating factor, PLC-zeta (Saunders et al. 2002, Barroso et al. 2009) and a complex repertoire of RNAs to the egg (Ostermeier et al. 2004, Yuan et al. 2015). Both sperm and egg are transcriptionally silent (Braun 2000, Richter & Lasko 2011) and the egg-to-zygote transition (EZT) occurs in the absence of transcription (Evsikov et al. 2006). While somatic nuclear cloning (Gurdon & Melton 2008) and the generation of viable gynogenetic mice (Kono et al. 2004) suggest that maternal factors alone are sufficient to guide early embryo development, these processes are grossly inefficient and structural or signaling factors from the sperm may complement maternal factors that could participate in and aid the early programming of embryonic development (Jodar et al. 2015, Miller 2015).

RNAs can epigenetically affect transgenerational inheritance through specific small non-coding RNAs (sncRNAs) and associated RNA-binding proteins (Rodgers et al. 2015, Chen et al. 2016). In \textit{Caenorhabditis elegans}, a hybrid strain crossing showed that approximately 10% of embryonic RNA is of paternal origin with functional importance during EZT and possibly embryogenesis (Stoeckius et al. 2014b). An equivalent paternal contribution to the mammalian zygote will be small by comparison, but evidence of the potential for sperm RNAs (or their translated proteins) to contribute to and participate in zygote formation is strong and worthy of further investigation.

Early molecular processes in the zygote can be classified into three main categories including maternal clearance, chromatin remodeling and eventually zygotic genome activation (ZGA) (Lee et al. 2013). Maternal clearance is the process of removing maternal factors including RNAs and proteins essential for oogenesis that become surplus to requirements after fertilisation.
(Tadros & Lipshitz 2009). Post-transcriptional regulation plays a role during EZT and includes the destruction of maternal mRNAs guided by their 3’ untranslated (3’ UTR) sequences (Giraldez 2010). Compared with approximately 2000 proteins reported in pre cleavage stage zygotes of M. musculus, over 3500 proteins have been identified in metaphase II eggs (Wang et al. 2010, Yurttas et al. 2010). During the embryonic development, this removal of maternal factors is guided mainly by ubiquitin-dependent degradation pathways and by autophagy (Marlow 2010).

While transcriptionally inert, MII eggs and zygotes are likely to be translationally active (Potireddy et al. 2006, Fang et al. 2014), leaving open the possibility for sperm RNAs to be translated into proteins following their introduction to the egg (Fang et al. 2014). We reasoned that a potential non-genomic paternal contribution would most likely involve interactions with maternal factors responsible for the regulation of gene expression prior to the EZT and the clearance of maternal factors prior to embryonic genome activation. The main objective of the study, therefore, was to see if in silico analysis of RNA sequencing data obtained from sperm, MII eggs and pre cleavage stage zygotes (PCZ but henceforth referred to as zygotes) using an identical bespoke protocol, could highlight potential interactions between paternal and maternal cofactors brought together by fertilisation. Herein, we focus on five, full-length mRNAs present at high levels in murine sperm that with associated maternal cofactors, fit the requirements for a potential paternal, non-genomic metabolic contribution to the zygote.

Materials and methods

Study design

RNAs isolated from pooled murine spermatozoa, single MII eggs and single zygotes were sequenced and compared. To help identify paternal RNAs with the potential to participate in EZT events, we looked firstly for candidate RNAs that were highly represented in sperm, thus increasing the probability of being delivered to and translated by the zygote or being translated into protein during the late stages of spermatogenesis and delivered to the MII egg at fertilisation. Secondly, considering the highly fragmented nature of sperm RNA, no less than 80% of the exonic regions of at least one gene isoform of a ‘candidate’ paternal cofactor had to be covered by RNA-seq reads, giving a greater potential for the RNA to be functional. Thirdly, sperm RNAs with good sequence coverage were only considered further if their ontological descriptions suggested functions other than spermatogenesis.

A bespoke identical library preparation method and sequencing pipeline were applied to all samples allowing accurate comparative assessment of RNAs across the different samples. The library kit used (Ovation single cell RNA-Sequencing system, NuGEN, CA, USA) has a mix of oligo-dT and random primers targeting a wider range of transcripts, including those with varying poly(A) tail lengths, typically encountered in gamete and zygotic mRNAs (Paranipe et al. 2013). Cytoscape’s GeneMANIA module (see below) was then employed for the in silico analysis investigating potential interactions between gene products of paternal and maternal origin (Warde-Farley et al. 2010).

Ethics

Experiments involving the use of animals were regulated under the Home Office, UK Animals Scientific Procedures Act (ASPA) under license service PPL 40/3391 approved by the University of Leeds AWERC (Animal Welfare Ethical Review Committee). All animals were culled using cervical dislocation in accordance to Schedule 1 of the ASPA.

Mouse gamete and zygote harvest

Groups of C57BL/6J females were super-ovulated with a 5 IU dose of pregnant mares’ serum gonadotropin (PMSG) (Sigma Aldrich) injected intraperitoneally (1P) on day 1, followed at day 3 by a 5 IU dose of human chorionic gonadotropin (HCG) (Sigma Aldrich) 1P and mated with vasectomised males to provide MII eggs. The zygote groups were mated to proven C57BL/6J stud males immediately after HCG dosing and checked the following day for post coital plugs. Plugged females were pooled and used for zygote harvest. Both zygote and MII egg groups were sacrificed on day 4. Oviducts from the zygote and egg groups were harvested separately and suspended in M2 media (Sigma Aldrich). Dissected oviducts were placed into a pre-heated dish of synthetic Human Tubal Fluid (HTF) media (Irvine Scientific, CA, USA) with bovine serum albumin (BSA) (Sigma Aldrich). Cumulus masses were released into the HTF/BSA medium and transferred into a drop of hyaluronidase (Sigma Aldrich) following which, a wide bore pipette was used to strip the eggs and zygotes of their cumulus cells. These were in turn collected by mouth pipette and washed through sequential drops of M2 media (Sigma Aldrich).

Sperm harvest

The epididymides of fertile C57BL/6J males were dissected out and transferred into pre-warmed HTF (Irvine Scientific). Using a sterile 26G needle, small incisions were made in the cauda and sperm were allowed to swim out before collection by gentle aspiration. Spermatozoa were washed in HTF (Irvine), filtered through an 80-micron mesh (Sigma Aldrich) and centrifuged at 500g prior to resuspension and centrifugation through a two-layer (65–50%) discontinuous percoll gradient (GE Healthcare Biosciences) at 300g, employing the ProInsert technology (Nidacon International AB, Gothenburg, Sweden) to facilitate the selective isolation of pelleted spermatozoa while preventing possible contamination by somatic cells (Fourie et al. 2012). Spermatozoa were pelleted at 500g and washed twice in Dulbecco’s phosphate-buffered saline (DPBS) (Thermo Scientific). Approximately 1 million spermatozoa were harvested before the second wash and Giemsa stained (Sigma Aldrich) to visually confirm lack of other cell types using a Leica Leitz DMRB microscope (Mazurek Optical Services, Southam, UK).
RNA isolation and library construction

Sperm RNA was extracted using the method described by Goodrich (Goodrich et al. 2013) with modifications. Briefly, 10⁷ spermatozoa were placed in RLT buffer (Qiagen) with 1.5% β-mercaptoethanol (Sigma Aldrich) and 0.5 mm nuclease free stainless steel beads. Following homogenisation with a DisruptorGenie cell disruptor (Thermofisher Scientific), an equal volume of chloroform was added followed by centrifugation at 12,000g (4°C), allowing recovery of the RNA. Prior to library construction, any residual genomic DNA was removed from the samples by digestion with Turbo DNase (ThermoFisher Scientific) following the manufacturer’s instructions. Quantitative Real-Time PCR (qRT-PCR) using Prm2 and Map1lc3a intron spanning primers with SybrGreen PCR mastermix (Applied Biosystems) was employed to monitor for DNA contamination.

Sperm RNA quality assessment was carried out using the RNA-6000 pico assay (Agilent) on a 2100 Bioanalyzer (Agilent), where the absence of clearly defined peaks from 18S and 28S ribosomal RNAs (low RIN score) indicates corresponding absence of contaminating somatic cell RNA (Supplementary Fig. 1, see section on supplementary data given at the end of this article). As additional QC, Real-Time qPCR primers for the Melanoma-Associated Antigen D2 (Maged2) were used to confirm potential contaminating somatic cell RNA in these preparations (principally from Leydig and Sertoli cells; Figure 1 Venn diagrams for the cross-representation of sperm, MII egg and zygote RNAs. Panel A shows the overlaps between all RNAs, aggregated from all biological replicates at ≥10CPM from each of the three sources. Panel B shows similar overlaps, except that the selected lists for MII eggs and zygotes were obtained after testing with Edge R, which normalises the data and identifies differentially expressed RNAs that are significantly ‘up’ in MII eggs or in zygotes. These over-represented RNAs are then added to the list of RNAs common to both MII eggs and zygotes. Complete RNA lists are provided in the Supplementary data.

Table 1 Top biological processes for source and differentially expressed RNAs.

<table>
<thead>
<tr>
<th>Term</th>
<th>P value</th>
<th>Benjamini</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spermatozoa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spermatogenesis</td>
<td>9.7E-9</td>
<td>4.8E-6</td>
<td>1.4E-5</td>
</tr>
<tr>
<td>Sperm motility</td>
<td>9.6E-7</td>
<td>2.4E-4</td>
<td>1.4E-3</td>
</tr>
<tr>
<td>Cell differentiation</td>
<td>2.3E-5</td>
<td>3.7E-3</td>
<td>3.3E-2</td>
</tr>
<tr>
<td>Multicellular organism development</td>
<td>1.4E-3</td>
<td>1.5E-1</td>
<td>2.0E0</td>
</tr>
<tr>
<td>Cellular response to hormone stimulus</td>
<td>1.5E-3</td>
<td>1.4E-1</td>
<td>2.2E0</td>
</tr>
<tr>
<td>Lipid biosynthetic process</td>
<td>2.1E-3</td>
<td>1.5E-1</td>
<td>2.9E0</td>
</tr>
<tr>
<td>Chromosome condensation</td>
<td>2.7E-3</td>
<td>1.7E-1</td>
<td>3.8E0</td>
</tr>
<tr>
<td>Penetration of zona pellucida</td>
<td>3.4E-2</td>
<td>8.8E-1</td>
<td>3.9E1</td>
</tr>
<tr>
<td>Fatty acid metabolic process</td>
<td>4.1E-2</td>
<td>9.0E-1</td>
<td>4.5E1</td>
</tr>
<tr>
<td>Intracellular signal transduction</td>
<td>4.6E-2</td>
<td>9.0E-1</td>
<td>4.9E1</td>
</tr>
<tr>
<td>MII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell cycle</td>
<td>1.6E-39</td>
<td>1.2E-35</td>
<td>3.1E-36</td>
</tr>
<tr>
<td>Cell division</td>
<td>5.1E-36</td>
<td>1.9E-32</td>
<td>1.0E-32</td>
</tr>
<tr>
<td>Protein transport</td>
<td>1.5E-35</td>
<td>3.7E-32</td>
<td>3.0E-32</td>
</tr>
<tr>
<td>Mitotic nuclear division</td>
<td>1.7E-29</td>
<td>3.1E-26</td>
<td>3.3E-26</td>
</tr>
<tr>
<td>DNA repair</td>
<td>1.2E-26</td>
<td>1.7E-23</td>
<td>2.3E-23</td>
</tr>
<tr>
<td>mRNA processing</td>
<td>8.8E-25</td>
<td>1.1E-21</td>
<td>1.7E-21</td>
</tr>
<tr>
<td>Cellular response to DNA damage stimulus</td>
<td>2.5E-23</td>
<td>2.7E-20</td>
<td>5.0E-20</td>
</tr>
<tr>
<td>Transcription, DNA-templated</td>
<td>4.5E-20</td>
<td>4.2E-17</td>
<td>8.9E-17</td>
</tr>
<tr>
<td>RNA splicing</td>
<td>2.0E-18</td>
<td>1.6E-15</td>
<td>3.9E-15</td>
</tr>
<tr>
<td>Transport</td>
<td>5.2E-18</td>
<td>3.8E-15</td>
<td>1.0E-14</td>
</tr>
<tr>
<td>Zygotes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell cycle</td>
<td>6.3E-43</td>
<td>4.6E-39</td>
<td>1.3E-39</td>
</tr>
<tr>
<td>Cell division</td>
<td>1.1E-39</td>
<td>4.6E-36</td>
<td>2.2E-36</td>
</tr>
<tr>
<td>Protein transport</td>
<td>4E-34</td>
<td>9.6E-31</td>
<td>7.8E-31</td>
</tr>
<tr>
<td>Mitotic nuclear division</td>
<td>2.9E-31</td>
<td>5.2E-28</td>
<td>5.7E-28</td>
</tr>
<tr>
<td>DNA repair</td>
<td>7.7E-29</td>
<td>1.1E-25</td>
<td>1.5E-25</td>
</tr>
<tr>
<td>Cellular response to DNA damage stimulus</td>
<td>4.9E-23</td>
<td>5.9E-20</td>
<td>9.7E-20</td>
</tr>
<tr>
<td>mRNA processing</td>
<td>8.2E-23</td>
<td>8.5E-20</td>
<td>1.6E-19</td>
</tr>
<tr>
<td>Transcription, DNA-templated</td>
<td>8.9E-20</td>
<td>8.1E-17</td>
<td>1.8E-16</td>
</tr>
<tr>
<td>RNA splicing</td>
<td>8.1E-19</td>
<td>6.6E-16</td>
<td>1.6E-15</td>
</tr>
<tr>
<td>Regulation of transcription, DNA-templated</td>
<td>3.1E-15</td>
<td>2.3E-12</td>
<td>6.2E-12</td>
</tr>
</tbody>
</table>

RNA lists from each source (sperm, MII eggs and zygotes) and differentially expressed MII-Zygote genes flagged by EdgeR analysis (MII up and PCZ UP) were submitted to DAVID for ontological analysis. Bioprocesses are reported alongside uncorrected P values, Benjamini corrected P values and False Discovery Rates (FDRs).
Chamel et al. 2007). Only spermatozoal cDNAs shown to be free of genomic DNA and somatic RNA contamination were used for library construction.

Mouse eggs and zygotes were processed at the single cell level. Each cell was transferred by mouth pipette into lysis buffer, after being immersed in DPBS (LIFE Technologies) in a washing step containing 0.1% BSA (Sigma Aldrich). Following first and second strand synthesis and processing using the Ovation single cell RNA-Seq system (NuGEN), Illumina adaptor sequences were ligated to the sperm, egg and zygote cDNAs. Two rounds of library amplification were carried out and the fragment distribution was checked using the Agilent high sensitivity DNA assay on the 2100 Bioanalyzer (Agilent Technologies). The libraries were quantified using Picogreen assay (Thermo Scientific) on a FLUOstar Galaxy plate reader (MTX Lab Systems, Bradenton, FL, USA) and pooled. The Illumina HiSeq 2500 and 3000 platforms were employed for RNA-Seq.

Bioinformatics analysis

Spermatozoa, MII eggs and zygotes from a minimum of three biological replicates each were sequenced using either 50bp (single-ended) or 150bp paired-end reads. RNA-Seq data underwent automated adapter and quality trimming using Trimgalore! v0.4, ignoring reads with MAPQ<20 (Krueger 2015). The filtered reads were mapped to the Mus musculus reference genome (mm10) using the subjunc function of the Rsubread package version 1.20.3 (Liao et al. 2013b). The output BAM format files were sorted using Samtools version 1.3 (Li et al. 2009) and duplicate reads removed using the Picard MarkDuplicates tool version 2.1.1 (Broad Institute, 2010), available online at http://broadinstitute.github.io/picard). BedGraph and bigwig files were generated using Bedtools version 2.25.0 (Quinlan & Hall 2010), and the function bedGraphToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/macOSX.x86_64/bedGraphToBigWig). After removal of duplicate unmapped and incorrectly paired reads using Samtools version 1.3 (Li et al. 2009), the reads were visualized on the UCSC genome browser (Kent et al. 2002). The numbers of reads assigned to genomic features were counted using the featureCounts function of Rsubread (Liao et al. 2013a). For paired-end libraries, we required both read mates to be uniquely mapped in the correct orientation. All remaining options were set to featureCounts default. Differential RNA representation in MII egg and zygote RNA-Seq libraries was tested using the edgeR exact test (Robinson et al. 2010) and only genes represented at levels ≥10 counts-per-million reads (CPM) in at least 6 out of 7 MII and zygote libraries were included in the downstream analysis. The only exception to this rule was for five maternal transcripts with reads just below 10CPM, represented across all exons, that were also included. Data normalisation was based on the trimmed mean of M value (TMM) using the calcNormFactors function (Robinson & Oshlack 2010).

Table 2 Paternally-derived factors and their potential maternal cofactors in Mus musculus.

<table>
<thead>
<tr>
<th>Spermatozoal gene</th>
<th>Associated gene name and MGI accession number</th>
<th>Function (UniProt)</th>
<th>Potential maternal interactions (GeneMANIA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fbxo2</td>
<td>F-Box Protein 2 (Source: MGI Symbol; Acc: MGI:2446216)</td>
<td>Component of E3 ubiquitin-protein ligase complex</td>
<td>Cul1, Fbx1, Fbx3, Fbx5, Fbxo34, Fbxo5, lgb1, Rbx1, Skp1a</td>
</tr>
<tr>
<td>Map1lc3a</td>
<td>Microtubule-associated protein 1 light chain 3 alpha (Source: MGI Symbol; Acc: MGI:1915661)</td>
<td>Ubiquitin-like protein involved in autophagosome formation</td>
<td>Atg10, Atg3, Babap2p, Kit15, Man1b, Map1lc3b, Pias4, Tab2a, Wdfy3</td>
</tr>
<tr>
<td>Pcbp4</td>
<td>Poly(C) binding protein 4 (Source: MGI Symbol; Acc: MGI:1890471)</td>
<td>Binds to single stranded nucleic acids and specifically to oligo(dC)</td>
<td>Pchp1, Pchp2, Qk, Hnmpk</td>
</tr>
<tr>
<td>Zip821</td>
<td>Zinc finger protein 821 (Source: MGI Symbol; Acc: MGI:1923121)</td>
<td>Maybe involved in transcription regulation through DNA binding</td>
<td>Agrp1, Chchd3, Fam13c, Fchsd2, Mob1a, Rimk1b, Tpy44, Zic2b</td>
</tr>
<tr>
<td>Hdac11</td>
<td>Histone deacetylase 11 (Source: MGI Symbol; Acc: MGI:2385252)</td>
<td>Deacetylases core histones and plays a role in transcription regulation, cell cycle, epigenetic repression and development</td>
<td>Aamp, Bub1b, Cdc20, Hda2c, Hdac8, Mett11b, Nedcd</td>
</tr>
</tbody>
</table>

Column 1 shows spermatozoal factors with good exonic representation as revealed by RNA-Seq and UCSC browsing. Column 2 gives the associated gene name and MGI accession number. Column 3 briefly depicts their functionality as described in UniProt. Column 4 shows the potentially interacting maternal factors as predicted by GeneMANIA.

Ontological analysis, gene networks and molecular interactions

Ontological descriptions of RNAs from sperm, MII eggs and zygotes were derived by DAVID v6.8 (Huang et al. 2009), with a subsequent focus on biological processes. Gene networks involving candidate spermatozoal and maternal factors were identified by the Cytoscape module GeneMANIA v3.4.1; (Warde-Farley et al. 2010). GeneMANIA uses publically available data sets, encompassing physical and molecular interactions, co-expression, co-localisation and molecular pathways.

NGS validation using quantitative real-time PCR

Following first-strand cDNA synthesis, the cDNA of mouse MII egg, zygote, spermatozoal and testicular RNA (positive control) was amplified by long distance PCR, using the SMART-Seq v4 ultra low input RNA kit (Clontech). Validatory quantitative real-time PCR was carried out as required using gene-specific primers (Supplementary Table 1) and SYBR green on an ABI 7900HT Real-time PCR system (Applied Biosystems) over 40 cycles according to the manufacturer’s instructions. The annealing temperature per primer pair ranged between 59°C and 62°C.
Results

RNA characterisation and ontological profiles

The average correctly paired and mapped reads per sperm RNA sample was calculated at $20 \pm 2 \times 10^6$. The average number for MII eggs was $18 \pm 1.5 \times 10^6$ and for the zygotes $20 \pm 1.5 \times 10^6$ per sample. RefSeq IDs for sperm, MII eggs and zygotes, alongside differentially expressed RNAs using EdgeR for MII eggs and zygotes are listed in the Supplementary data (RNA lists). While we cannot be certain that RNAs common to sperm and zygotes originated in the fertilising sperm, our initial approach was to look for sperm RNAs that were absent in MII eggs but present in zygotes. Figure 1A shows Venn diagrams for overlaps between sperm, MII egg and zygote RNAs $\geq 10$ CPM. In aggregate, 5368, 5148 and 1918 RNAs were reported, respectively from MII eggs, zygotes and sperm with 75 shared between sperm and zygotes that were either absent altogether or present in MII eggs at well below threshold reporting levels. The same sperm list compared with EdgeR normalised reads for MII egg and zygote RNAs yielded 56 RNAs shared between sperm and zygotes (Fig. 1B). Closer scrutiny of the read data, however, showed all but four of these ‘shared’ RNAs were detected in MII eggs albeit at low levels of expression ($<10$ CPM). The four absent in MII eggs were more fragmented in sperm than in zygotes, suggesting they were not sperm-specific.

An alternative approach focused simply on highly abundant sperm RNAs with good exon coverage that were essentially absent in both MII eggs and zygotes. To help narrow down the list of hundreds of possible RNAs to pursue in this regard, functional annotation clustering (Supplementary data; FAC sheets) was employed to provide a general overview of MII egg and zygote RNAs using the lists of differentially expressed MII egg and zygote RNAs generated by EdgeR alongside the list of sperm RNAs selected on the basis of high representation and good exon coverage. A graphical representation of the numbers of genes in the main ontological annotations (biological processes) for sperm, MII eggs and zygotes is shown in Supplementary Fig. 2.

As shown in Table 1 and in Supplementary data (BP; EdgeR sheet), the expected top sperm annotation related to spermatogenesis as a differentiation process, with associated weaker enrichments in processes associated with lipid metabolism and DNA condensation. Enrichment for annotation relating to the control of transcription dominated the ontological descriptions for both MII eggs and zygotes, which was expected considering the similarity between them (Table 1 and Supplementary data; BP EdgeR sheet). Focusing on differential expression between the two, however, revealed interesting differences (Supplementary data; BP EdgeR MII or PCZ Up sheet). MII eggs showed specific enrichments in activities relating to mRNA processing, while the cell cycle showed the strongest enrichment in zygotes. Processes relating to ubiquitination and transcription were more apparent in differentially up-regulated RNAs from zygotes but not MII eggs, suggesting that clearance activity and perhaps renewed RNA processing triggered by fertilisation may have already commenced in zygotes at the time of harvesting.

Using the functional annotation of highly expressed RNAs as guidance alongside closer inspection of the selected gene lists from sperm, MII eggs and zygotes,
Figure 3 Continued.
Figure 3: Alignment of RNA sequencing reads (pile ups) across representative components of the E3 ubiquitin ligase complex. As tracked on the UCSC genome browser for all spermatozoal (sperm 1–3), egg (MII 1–4) and zygote (PCZ 1–3) biological replicates, reads for components of the E3 ubiquitin ligase complex are shown for $\textit{Fbxo2}^\sigma$ (A) alongside $\textit{Cul1}^\Omega$ (B), $\textit{Rbx1}^\Omega$ (C) and $\textit{Skp1a}^\Omega$ (D). Genes are depicted at the foot of each diagram with exons shown as filled blocks.
potential interacting partners relating to clearance of maternal factors were revealed. Considering the RNAs' high expression in sperm compared with MII eggs and zygotes, their exclusion from the dominant spermatogenesis ontology and their relative freedom from fragmentation as assessed by exon coverage, five sperm RNAs were selected for follow up (Table 2). These include the histone deacetylase 11 (Hdac11), the Rbx1-SCF E3 ubiquitin-ligase component F-box protein 2 (Fbxo2), the microtubule-associated protein 1A/1B light chain 3A (Map1lc3a), the poly (rC)-binding protein 4 (Pcbp4) and the zinc finger protein 821 (Zfp821). These five sperm RNAs were in turn interrogated using GeneMANIA for all known interacting partners, which returned approximately 100 genes of which 37 were either present in the up-regulated zygotic transcripts or in the list of shared (MII & zygote) maternal RNAs (Table 2). Together, these paternal and maternal RNAs comprise the gene network profile shown in Fig. 2.

The network's functional annotation was dominated by strong enrichment in processes related to ubiquitin-mediated degradation pathways (Supplementary data; BP EdgeR sheet), reflecting the ontology of up-regulated transcripts in zygotes.

Predicting and providing evidence for potential parental interactions

Network analysis (Fig. 2) suggested that paternal (♂) and maternal (♀) cofactors could interact in pathways leading to EZT. An example is illustrated in Fig. 3 for the gamete-specific cofactors of the multiple component SCF E3 ubiquitin ligase complex which includes Fbxo2 (also known as Fbs1♂; A) alongside Cullin 1 (Cul1♀; B), Ring Box 1 (Rbx1♀; C) and S-Phase Kinase-Associated Protein 1A (Skp1a♀; D). These genes are indicated in Fig. 2 by boxes. Note that reads covering all exons for Fbxo2, were strongly represented in the sperm RNA libraries but with few or no reads from either MII egg or zygote libraries. In contrast, with the exception of Fbxo2, RNAs encoding the other cofactors of the SCF-E3 ubiquitin ligase complex were highly represented in both MII eggs and zygotes but not in sperm. Additional predicted interactions between Fbxo2♂ and the maternally expressed Fbxo5♀ and Fbxo34♀ were also suggested (Fig. 2 and Table 2; boxes). Real-time qPCR confirmed the expression of paternal factors in sperm and testis (Fig. 4) and although products were generated for Map1lc3a from all sources (panel A), Ct data confirmed that the RNA was considerably more abundant in sperm (see panel B). A 142 bp product from Maged 2 was only detected in RNA from eggs and zygotes, indicating that sperm libraries were free of contamination by RNAs from testis-derived somatic cells. All PCR products were obtained from samples after 40 PCR cycles and so the products shown in panel A are only qualitative. The corresponding Ct values give more quantitative assessments.

UCSC tracks are shown in Supplementary Fig. 3 for a number of additional, GeneMANIA suggested potential paternal-maternal interacting cofactors. Hdac11♂ (A), which was highly expressed in sperm, could interact
with the Mitotic Checkpoint Serine/Threonine Kinase B, (Bub1; B) and cell division cycle protein 20 (Cdc20; Fig. 2; ovals). GeneMANIA also suggested potential co-localisation and co-expression between Hdac11, Hdac2 and Hdac8 (Fig. 2; ovals) with Hdac11 and Hdac2 having shared protein domains. Predicted interactions with the Nlcle2 and Aamp were also highlighted (Fig. 2; ovals). Reads from a long terminal repeat (LTR) region located within the 8th intron of Hdac11, which could potentially be expressed independently of Hdac11 RNA were also noted (Supplementary Fig. 3A).

The ubiquitin-like protein Microtubule Associated protein 1, light chain 3 alpha (Map1lc3a; C) is involved in autophagosome formation and GeneMania indicated functional interactions with several maternal factors, including Map1b (D), Atg3 (E) and Atg10 (Fig. 2, boxes with course dashed lines) of which Atg3 and Map1b showed high levels of expression in MII eggs and zygotes with good exonic representation and were absent in sperm. In addition, Pcbp4 (F), which may complement the heterogeneous nuclear ribonucleoprotein K (Hnmpk; G) and Quaking (Qk; H; and Fig. 2, pentagons) were highly represented in MII eggs and zygotes but not in sperm. The corresponding maternal proteins are involved in post-transcriptional regulation of gene expression, protein degradation and the cell cycle. Pcbp4 also has predicted interactions with Pcbp1 and co-localises with the Pcbp2 isoform. Finally, GeneMANIA reported predicted interactions between Zfp821 (I), which may be involved in transcriptional regulation, Fchsd2 and Rimk8 (Fig. 2; boxes with fine dashed lines). Zfp821 is highly expressed in sperm but not MII eggs and zygotes.

Discussion

Existing sequencing germ line and zygote datasets are not fully complementary and are therefore difficult to compare (they either omit sperm or MIImsgs or zygotes from their analysis) and are derived from library construction methods that differ between the various cell types (Tang et al. 2010, Xue et al. 2013, Johnson et al. 2015). To avoid introducing methodological effects and bias, we used a bespoke pipeline that included construction of our own libraries for sequencing and analysis. Sperm contain far less RNA than either MII eggs or zygotes; therefore, sperm libraries were unavoidably derived from sperm-specific pooled samples while MII egg and zygote libraries were from individual cells. The equivalent read counts obtained from the three sources demonstrate the care taken to assure quantitative equivalence of input RNA. We found, however, that relying on comparisons between the three to select sperm-specific factors were problematic, because although representation may have been too low to report the RNA as present in MII eggs or zygotes (for example), we frequently encountered reads indicating fragmented RNA in both regardless. We focused, therefore, on highly abundant RNAs with full-length transcripts in sperm, MII eggs and zygotes and with demonstrably reciprocal representation (in sperm but not MII eggs or zygotes and the reverse) following inspection of RNA-Seq tracks on the UCSC browser and where necessary, confirming by qRT-PCR.

In silico analysis of our RNA sequencing data supports the possible complementation of maternal with paternal factors introduced at fertilization. Five highly expressed sperm RNAs were considered based...
on their relative low abundance or absence in MII eggs or zygotes. All factors potentially interacting with translated products from these RNAs were mapped out using the pathway and network analyses tools in GeneMANIA. These factors were then matched to corresponding maternal cofactors to help identify those with a greater potential to participate in EZT pathways. As the predicted interactions were more likely to be between proteins, where possible we checked for a corresponding proteomic record of the RNAs in question (Wang et al. 2010, 2013, Skerget et al. 2015). Sperm RNAs could either be translated into proteins during late spermatogenesis or if delivered to the MII egg, in the pre cleavage stage zygote (Fang et al. 2014). We also searched for recombination effects of existing knockout models for the corresponding genes of parental factors where available (Supplementary Table 2).

The current study provides evidence of a novel role for paternally introduced factors in murine zygotic RNA/protein clearance (Stitzel & Seydoux 2007, Sato & Sato 2013). The RBX1-SCF E3 ubiquitin ligase complex, for example, plays an important role during gametogenesis and mouse embryogenesis, catalysing the ubiquitination of proteins during cytoplasmic turnover, which are then destined for proteasomal degradation (Jia & Sun 2009, Sato & Sato 2013). The F-box family includes FBXO2, which is an E3 ligase adaptor protein targeting glycosylated proteins for degradation. Our network analysis, showed that all RBX1-SCF E3 ubiquitin ligase components but one (FBXO2δ) were maternally expressed (McCall et al. 2005) and on fertilisation, FBXO2δ could complete the complex and be active in the EZT. The FBXO2 protein is also present in mature sperm (Wang et al. 2013) and has been linked to idiopathic male infertility (Bieniek et al. 2016). Similarly, MAP1LC3Aδ is a ubiquitin-like modifier (Cherra et al. 2010) with potential autophagic interactions with ATG3φ and MAP1Bβ. MAP1B protein is also present in eggs and zygotes. Both Atg3φ and Map1b KO studies show lethality one day after birth (Supplementary Table 2).

Quaking (OQKφ), HNRNPKφ and PCBP1/2/4φ/δ showed predicted in silico interactions as part of the post-transcription regulatory process. In C. elegans, the first wave of degradation of egg factors involves PES4 (Stoeckius et al. 2014a), a member of the PCBP family of RNA-binding proteins that post-transcriptionally regulate alternative polyadenylation at a global level (Ji et al. 2013). Both members of the Poly-(rC) binding protein family, PCBP4φ and PCBP1/2φ, detected in our analysis, are mammalian orthologues of the nematode PES4 protein. Potential interactions between HDAC11δ, BUB1β and CDC20γ were predicted by our analysis and both Hdac11 and Bub1b were detected in high levels in sperm and eggs/zygotes, respectively. HDAC11γ is involved in epigenetic repression, transcriptional regulation and embryonic development (Haberland et al. 2009, Bagui et al. 2013, Sahakian et al. 2015). HDAC11 activates BUB1B by deacetylation, which in turn lifts the inhibition of the CDC20/APC complex, activating its ubiquitin ligase activity (Watanabe 2014). Although the fertility rate of mice homozygous for Hdac11 deletion has not been reported, Cdc20 KO mice showed 2-cell embryo arrest and Bub1b KO mice show developmental arrest in early gestation (E8.5 (Supplementary Table 2). A particularly interesting finding was the expression of a long terminal repeat (LTR) transposable element, located entirely within the 8th intron of Hdac11 (intragenic) in sperm, which was absent in both MII eggs and zygotes. LTR RNAs are expressed abundantly in mouse eggs and zygotes where they are thought to augment the regulation of host gene expression (Peaston et al. 2004, Göke et al. 2015). Spermatozoal LTRs transferred into the oocyte during fertilization, could lead to new retrotransposition events and possibly genetic alterations (Kitsou et al. 2016). Paternally derived Hdac11 LTRs may have maternal targets that together participate in the regulation of zygotic gene expression.

In conclusion, our data support the argument favouring extra-genomic contributions by the fertilising sperm to the zygote. In addition to the inheritance of acquired traits propagating transgenerationally via sperm RNA (Gapp et al. 2014, Chen et al. 2016), our data and its analysis provide evidence for a role of paternal RNAs or proteins in maternal clearance during EZT. Sperm may deliver signals or factors that can potentially interact locally with maternal cofactors and act, perhaps as a ‘last minute’ checkpoint or gateway for embryonic genome activation (EGA). The hypothesis of confrontation and consolidation with regard to the uniquely invasive nature of sperm entry to the egg falls into this latter category (Miller 2015, Bourc’his & Voinnet 2010). Figure 5 shows a model for how a sperm factor introduced at fertilisation might complement a maternal cofactor or pathway required for the EZT. At least one such sperm-borne factor, PLC zetaφ is already known to activate the oocyte (Saunders et al. 2002). A similar approach to ours could be employed to investigate paternal/maternal interactions in humans. However, to confirm the potential biological relevance of the suggested interacting cofactors reported in this study, additional work such as RNA knock down upon or conditional gene knock out prior to fertilization would require the mouse model. In view of the renewed concern over rising human male infertility (Levine 2017) and the rapid rise and expansion of infertility treatment by ICSI, further research into extra-genomic paternal contributions using model systems is fully justified.

**Supplementary data**

This is linked to the online version of the paper at http://dx.doi.org/10.1530/REP-17-0097.

Reproduction (2017) 154 777–788
Declarations of interest

The authors declare that they have no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work was supported by a FP7, Marie Curie International Reproductive Biology Early Research Training (Reprotrain) network fellowship to Panagiotis Ntostis (Grant number 289980, 2012). P Ntostis carried out the experimental work and wrote the original manuscript, and P Ntostis and D Iles carried out the bioinformatic analysis reported in this paper. D Carter assisted with the animal work, J Huntriss and M Tzetis provided critical reading of the manuscript and D Miller and D Iles designed the original experimental plan and D Miller revised and wrote the final manuscript.

References


Haberland M, Montgomery RL & Olson EN 2009 The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nature Reviews Genetics 10 32–42. (doi:10.1038/nrg2485)


Krueger F 2015 Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.


www.reproduction-online.org

Reproduction (2017) 154 777–788

Downloaded from Bioscientifica.com at 11/01/2023 10:32:17AM via free access