A conserved set of maternal genes?

Citation for published version:

Digital Object Identifier (DOI):
10.1387/ijdb.140121ad

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published in:
The International journal of developmental biology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
A conserved set of maternal genes?
Insights from a molluscan transcriptome

M. MAUREEN LIU1,2, JOHN W. DAVEY3,4, DANIEL J. JACKSON5, MARK L. BLAXTER3,6 and ANGUS DAVISON*1

1School of Life Sciences, University of Nottingham, University Park, Nottingham, UK, 2Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, UK, 3Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, 4Department of Zoology, University of Cambridge, Downing Street, Cambridge, UK, 5Courant Research Centre for Geobiology, University of Göttingen, Göttingen, Germany and 6Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

ABSTRACT The early animal embryo is entirely reliant on maternal gene products for a ‘jump-start’ that transforms a transcriptionally inactive embryo into a fully functioning zygote. Despite extensive work on model species, it has not been possible to perform a comprehensive comparison of maternally-provisioned transcripts across the Bilateria because of the absence of a suitable dataset from the Lophotrochozoa. As part of an ongoing effort to identify the maternal gene that determines left-right asymmetry in snails, we have generated transcriptome data from 1 to 2-cell and ~32-cell pond snail (Lymnaea stagnalis) embryos. Here, we compare these data to maternal transcript datasets from other bilaterian metazoan groups, including representatives of the Ecdysozoa and Deuterostomia. We found that between 5 and 10% of all L. stagnalis maternal transcripts (~300-400 genes) are also present in the equivalent arthropod (Drosophila melanogaster), nematode (Caenorhabditis elegans), urochordate (Ciona intestinalis) and chordate (Homo sapiens, Mus musculus, Danio rerio) datasets. While the majority of these conserved maternal transcripts (“COMATs”) have housekeeping gene functions, they are a non-random subset of all housekeeping genes, with an overrepresentation of functions associated with nucleotide binding, protein degradation and activities associated with the cell cycle. We conclude that a conserved set of maternal transcripts and their associated functions may be a necessary starting point of early development in the Bilateria. For the wider community interested in discovering conservation of gene expression in early bilaterian development, the list of putative COMATs may be useful resource.

KEY WORDS: maternal to zygotic transition, mollusk, MBT, MZT, Spiralia

Introduction

Cell division requires that genome replication and assortment are achieved while cellular function is maintained. In somatic cells, there is continuity of cytoplasm from mother to daughter, so that new nuclei take up the reins of cellular control as transcription of their genomes is resumed after division. In contrast, in the formation of a new organism the early zygote has to perform a similar feat of taking control of a new cell, but the task is made more complex because the gametic pronuclei must be reprogrammed and coordinated before transcription initiation. In animal embryos the zygotic cytoplasm, provisioned by the mother, has been found to contain all the machinery necessary to drive the first stages of embryonic development. This maternal provisioning has been demonstrated through the blocking of transcription from the zygotic genome (Baroux et al., 2008). In transcriptionally-blocked...
embryos, maternal products are often sufficient to drive the first rounds of cell division, and even the first phases of differentiation (Baroux et al., 2008).

The switch between maternal and zygotic control is called the maternal-zygotic transition (MZT), or the midblastula transition (MBT), and spans the period from fertilisation to the point where maternally provisioned factors are no longer sufficient to deliver normal development (Baroux et al., 2008, Stitzel and Seydoux, 2007, Tadros and Lipshitz, 2009). The MZT is associated with the activation of the zygotic genome. In animal species where fine-scale analyses have been performed, zygotic gene activation has been modelled as two phases (Baroux et al., 2008, Tadros and Lipshitz, 2009). An early phase, involving a few loci, is associated with degradation of maternal proteins and mRNAs, while the second phase is much more extensive and includes genes involved in a wide range of biological processes (Schier, 2007, Tadros and Lipshitz, 2009). Initial, albeit limited, zygotic genome activation has been identified as early as the fertilised zygote (in the paternal pronuclei of mouse, sea urchin and the nematode Ascaris suum), as late as the 256-cell embryo stage (in Xenopus) (Baroux et al., 2008, Tadros and Lipshitz, 2009, Wang et al., 2013).

Experimental evidence indicates that the MZT is tightly regulated, and includes the birth of zygotic RNAs and the death of maternal RNAs (Schier, 2007, Stitzel and Seydoux, 2007, Tadros and Lipshitz, 2009), taking place at multiple levels and in a controlled and managed manner. Thus, while many embryos are able to transcribe experimentally introduced DNA, the early embryonic genome is maintained in a state that is incompatible with transcription. Changes in chromatin structure, combined with a dilution of factors such as transcriptional repressors by cell division, allow for the initiation of zygotic transcription. Nonetheless, despite the complexity, it has been suggested that the MZT can be simplified into two interrelated processes: the first whereby a subset of maternal mRNAs and proteins is eliminated, and the second whereby zygotic transcription is initiated (Schier, 2007, Tadros and Lipshitz, 2009).

In zebrafish, maternally-provisioned products from just three genes, Nanog, Pou5f1 and Sox2 (known for their roles in embryonic stem cell fate regulation), are sufficient to initiate the zygotic developmental program and to induce clearance of the maternal program by activating the expression of a microRNA (Lee et al., 2013, Leichsenring et al., 2013). In Xenopus, increasing nuclear to cytoplasmic ratio is believed to be the controlling element in the switch, with just four factors regulating multiple events during the transition (Collart et al., 2013). However, the generality of these findings remains unknown. Furthermore, while the regulation of RNA transcription (gene expression) has received considerable attention (primarily due to the advances in nucleic acid sequencing technologies), protein expression and turnover rates remain relatively under-studied (Stitzel and Seydoux, 2007). Our knowledge of maternal-to-zygotic transcription phenomena is also largely restricted to the dominant model animal species, with relatively few experimental studies existing for other metazoans.

Although there has been a recent upsurge in interest in the maternal control of embryonic development, especially the MZT (Benoit et al., 2009, De Renzis et al., 2007, Lee et al., 2013, Leichsenring et al., 2013, Tadros and Lipshitz, 2009), the study of maternal factors has played an important part in the history of embryology and development, particularly in the model animal taxa Drosophila melanogaster (phyllum Arthropoda from superphylum Ecdysozoa), Caenorhabditis elegans (Nematoda, Ecdysozoa), Strongylocentrotus purpuratus (Echinodermata, Deuterostomia), Mus musculus, Homo sapiens and Danio rerio (Chordata, Deuterostomia) (Gilbert, 2006). Missing from this roster of models are representatives of “the” superphylum Lophotrochozoa, a morphologically diverse group that includes the Mollusca and Annelida. Two annelid models, Platynereis dumerilii and Capitella telata, are becoming well established (Dill and Seaver, 2008, Giani et al., 2011, Hui et al., 2009), but model molluscs have been developed for their potential to answer particular questions (e.g. asymmetric distribution of patterning molecules during development; Lambert and Nagy, 2002), or their association with a particular disease (e.g. schistosome transmitting Biomphalaria: Knight et al., 2011).

As part of an ongoing effort to identify the maternal gene that determines left-right asymmetry in molluscs (Harada et al., 2004, Kuroda et al., 2009, Liu et al., 2013), we are developing Lymnaea stagnalis pond snails as a model because they are one of the few groups that exhibit genetically-tractable, natural variation in their left-right asymmetry, or chirality, and so are ideal systems in which to understand why chirality is normally invariant, yet also pathological when it does vary (Schilthuizen and Davison, 2005). In generating a maternal transcriptomic resource for this species (the chirality-determining gene is maternally expressed; Boycott and Diver, 1923, Sturtevant, 1923), we were surprised to discover that while there are general studies on the composition and regulation of maternal expression (Shen-Orr et al., 2010), there has been no comprehensive description of shared bilaterian maternal genes.
One reason may be that no maternal gene resource exists for the Lophotrochozoa, Spiralia or Mollusca. Instead, previous work has described early developmental transcription in the molluscs *Ilyanassa* sp. (Lambert et al., 2010) and *Crepidula fornicata* (Henry et al., 2010), but using combined developmental stage libraries. Here we compare a new 1 to 2-cell *L. stagnalis* transcriptome (presumed maternal) to maternal transcriptomes from selected ecdysozoan and deuterostome species to identify conserved maternally provisioned genes across the Bilateria.

## Results

**L. stagnalis embryonic transcriptome sequencing and assembly**

Roche 454 sequencing of the two *L. stagnalis* libraries (1 to 2-cell and ~32-cell) generated 192,758 and 218,893 reads respectively, of which 163,004 and 192,552 were 150 bases or longer. The 1 to 2-cell assembly generated more contigs than the ~32-cell assembly, despite having fewer sequences (Table 2). A GC content of 36% for both libraries was approximately the same as previously reported for *L. stagnalis* (Adema et al., 2006, Liu et al., 2013). Merging the two assemblies produced by Newbler and MIRA resulted in fewer, longer contigs. The 1 to 2-cell library generated 11,212 contigs, and the ~32 cell library 9,497 contigs.

### Comparison between maternal transcriptomes

We compared the two developmental transcriptomes of *L. stagnalis* to each other and to six published maternal transcriptomes of roughly comparable depth derived from four deuterostomes and two ecdysozoans (Table 3: Aanes et al., 2011, Azumi et al., 2007, Baugh et al., 2003, De Renzis et al., 2007, Eviskoy et al., 2006, Grondahl et al., 2010). For *M. musculus* and *C. elegans*, maternal-only transcripts (present in the oocyte or egg but not in developing embryos) and maternal-zygotic transcripts (found in both oocyte or egg, and after zygotic transcription has started) have been defined. For the mouse, 2,834 genes were maternal-only and 1,796 maternal-zygotic, while for *C. elegans* 2,794 were maternal-only and 2,285 maternal-zygotic (Baugh et al., 2003, Eviskoy et al., 2006).

By reciprocal tBLASTx analyses, we identified putatively orthologous genes present in each of the seven species. About one quarter of each of the other maternal transcriptomes, between 900 and 1,900 genes, overlapped with the maternal transcriptome of the pond snail, *L. stagnalis* (Table 4). Surprisingly, 481 of the *L. stagnalis* genes had putative orthologues in all seven taxa (Supplementary Table 1). These 481 orthologues in fact probably represent 439 or fewer distinct genes, as BLASTx analyses revealed that some matched the same sequence in the NCBI nr protein database. This result implies that 5-10% of the maternal transcriptome is conserved and shared across all of the representative taxa (*H. sapiens* 6.1%, *M. musculus* 9.9%, *D. rerio* 10.6%, *C. intestinalis* 11.4%, *D. melanogaster* 7.0%, *C. elegans* 9.0%). We refer to this conserved set as the “conserved maternal transcriptome” (COMAT).

We compared the *L. stagnalis* 1 to 2-cell transcriptome to maternal-only transcripts and maternal-zygotic transcripts from *M. musculus* and *C. elegans* (Baugh et al., 2003, Eviskoy et al., 2006) using tBLASTx. The *M. musculus* maternal-only data set matched 1069 *L. stagnalis* transcripts, whereas the *M. musculus* maternal-zygotic data set matched 884 *L. stagnalis* transcripts. Of the 481 COMATs from *L. stagnalis*, 219 were found in the *M. musculus* maternal-only data set and 261 in the *M. musculus* maternal-zygotic data set, indicating a relative over-representation of

---

Fig. 2. Visualisation of maternal gene product spatial distribution in uncleaved zygotes of *Lymnaea stagnalis* by whole mount *in situ* hybridisation. Eight maternal gene products were visualised in uncleaved zygotes relative to a negative control (β-tubulin). (A) β-tubulin is not detectable in uncleaved zygotes. A polar body is indicated by the horizontal arrow. (B) β-tubulin is clearly expressed in ciliated cells of older veliger larvae. (C) contig_2724: ATP-dependent RNA helicase dnx8. (D) contig_453: heat shock 70 kda protein cognate 4. (E) contig_7974: ADP-ribosylation factor 4. (F) contig_9053: proteasome alpha 6 subunit. (G) contig_9053: proteasome alpha 6 subunit. (H) contig_9016: eukaryotic translation initiation factor 3 subunit i. (I) contig_8075: eukaryotic translation elongation factor. (J) contig_8318: 78 kda glucose-regulated protein.
### TABLE 1

**PRIMER SEQUENCES USED TO ISOLATE GENE FRAGMENTS FOR RIBOPROBE SYNTHESSES**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer (5' to 3')</th>
<th>Reverse primer (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta-tubulin</td>
<td>TGTGGAAGTGATCCCCCCACAGATGCA</td>
<td>TCACCTAGGCTTGTGATCGCCTT</td>
</tr>
<tr>
<td>c2704 ATP-dependent RNA helicase</td>
<td>GCACGCGGTTTCCTCCGAGA</td>
<td>TTTTCCTGCTCCTTTTCGCTT</td>
</tr>
<tr>
<td>c453 heat shock 70 kDa protein</td>
<td>CCACGTCTGACCGTAACTGCTA</td>
<td>CTGATGACGACACCGCCGTTGA</td>
</tr>
<tr>
<td>c7974 ADP-ribosylation factor 4</td>
<td>CAAGTTGGAACGCTACCGGAGA</td>
<td>AAACTCCACACACACCCCCAAC</td>
</tr>
<tr>
<td>c9053 proteasome alpha 6 subunit</td>
<td>CGCCTCCTGATAGGGAAGGAGCTA</td>
<td>TCACTGATGACGACACCCCAAC</td>
</tr>
<tr>
<td>c579 erg and galp 2</td>
<td>CTGTCTGACAGTTGCGGGTTTTG</td>
<td>TCCGCTTGATTGACGCGCTTGA</td>
</tr>
<tr>
<td>c9016 eukaryotic translation initiation factor 3 subunit i</td>
<td>TGGTGCTGTTTGGTGTGATGGTGA</td>
<td>CTGAAAGGACGGCAGACTACAGCA</td>
</tr>
<tr>
<td>c8075 eukaryotic translation elongation factor</td>
<td>TACCTGCGAAGCCCGATGTTGGA</td>
<td>CGCAAAACGGCGACATATAAGCA</td>
</tr>
<tr>
<td>c8318 78 kDa glucose-regulated protein</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 2

**ASSEMBLY OF THE LYMNAEA STAGNALIS EMBRYO TRANSCRIPTOMES**

<table>
<thead>
<tr>
<th></th>
<th>1 cell transcriptome</th>
<th>32 cell transcriptome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Newbler 2.6</td>
<td>MIRA</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>13,201</td>
<td>15,419</td>
</tr>
<tr>
<td>Max contig length</td>
<td>4,258</td>
<td>2,937</td>
</tr>
<tr>
<td>Number contigs &gt;100bp</td>
<td>12,908</td>
<td>15,184</td>
</tr>
<tr>
<td>&gt;100bp GC content</td>
<td>36.3</td>
<td>35.8</td>
</tr>
<tr>
<td>Number contigs &gt;1000bp</td>
<td>1,685</td>
<td>1,375</td>
</tr>
<tr>
<td>&gt;1000bp GC content</td>
<td>36.4</td>
<td>36.8</td>
</tr>
<tr>
<td>Contigs versus SwissProt hits</td>
<td>27.60%</td>
<td>25.80%</td>
</tr>
</tbody>
</table>

### TABLE 3

**MATERNAL TRANSCRIPTOME DATASETS USED IN THIS STUDY**

<table>
<thead>
<tr>
<th>Taxonomic group / Species</th>
<th>Common name</th>
<th>Number of maternal genes</th>
<th>Method</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deuterostomia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>human</td>
<td>7,470</td>
<td>Array analysis of metaphase II oocytes</td>
<td>Grandahl et al.</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>mouse</td>
<td>4,645*</td>
<td>Sanger sequencing of oocyte cDNA library</td>
<td>Evsikov et al.</td>
</tr>
<tr>
<td>Danio rerio</td>
<td>zebrafish</td>
<td>4,375*</td>
<td>ABI Solid cDNA sequences of oocyte and early embryo</td>
<td>Aanes et al.</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
<td>Ciona / sea squirt</td>
<td>4,041*</td>
<td>Array analysis of early embryo</td>
<td>Azumi et al.</td>
</tr>
<tr>
<td>Ecdysozoa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>Drosophila / fly</td>
<td>6,582*</td>
<td>Array analysis of early embryo</td>
<td>De Renzis et al.</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>C. elegans / worm</td>
<td>5,081*</td>
<td>Array analysis of early embryo</td>
<td>Baugh et al.</td>
</tr>
<tr>
<td>Lophotrochozoa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymnaea stagnalis</td>
<td>snail</td>
<td>11,212</td>
<td>454 sequencing of cDNA library from 1 cell embryo</td>
<td>This study</td>
</tr>
</tbody>
</table>

* more sequences listed in paper, but not all retrievable or present in database (mouse ~5,400; worm 6,042; zebrafish 4,465)

### TABLE 4

**COMPARISON BETWEEN MATERNAL TRANSCRIPTOMES**

<table>
<thead>
<tr>
<th>Species</th>
<th>Maternal transcriptome</th>
<th>Number with orthologues in Lymnaea stagnalis transcriptome</th>
<th>% Unique hits</th>
<th>% Reciprocal hits</th>
<th>% Unique reciprocal hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>7,470</td>
<td>2,394</td>
<td>32%</td>
<td>1,852</td>
<td>25%</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>4,645</td>
<td>1,954</td>
<td>42%</td>
<td>1,442</td>
<td>31%</td>
</tr>
<tr>
<td>Danio rerio</td>
<td>4,375</td>
<td>1,913</td>
<td>44%</td>
<td>1,452</td>
<td>33%</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
<td>4,041</td>
<td>1,360</td>
<td>34%</td>
<td>954</td>
<td>24%</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>6,582</td>
<td>2,501</td>
<td>38%</td>
<td>1,980</td>
<td>30%</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>5,081</td>
<td>1,662</td>
<td>33%</td>
<td>1,220</td>
<td>24%</td>
</tr>
</tbody>
</table>
maternal-zygotic transcripts that are conserved between chordate and mollusc, compared with maternal-only (Fisher’s exact test, 2.834:1.796 maternal-only:maternal-zygotic *M. musculus* versus 1.069:884 maternal-only:maternal-zygotic *L. stagnalis*, *P* < 0.0001), especially when considering COMATs (Fisher’s exact test, 2.834:1.796 versus 219:261, *P* < 0.0001). A similar result was found in comparisons between *L. stagnalis* and *C. elegans* (Fisher’s exact test, 2794:2285 versus 733:929 or 222:259, *P* < 0.0001, *P* < 0.0002). Similar comparisons were also made for maternal transcripts identified as being actively degraded or not degraded in the early embryo (Baugh et al., 2003, Eviskov et al., 2006), but no differences were found.

**Gene ontology analyses**

About one-third (31% of the 1 to 2-cell and and 36% of the ~32-cell) *L. stagnalis* transcripts (~3,400 genes) had significant BLASTx matches in the SwissProt database (Table 2). Blast2GO was used to functionally annotate both *L. stagnalis* transcriptomes. Of the 11,212 1 to 2-cell contigs, 4,311 (38%) had a significant BLASTx match, and 3,481 (36%) were assigned GO identifiers. For the COMAT subset, all but one of the 481 sequences had a significant BLASTx match, and 435 (90%) were assigned GO identifiers (Supplementary Table 1).

The distribution of GO annotations into functional categories revealed no obvious qualitative differences between the 1 to 2-cell and ~32-cell *L. stagnalis* transcriptomes (Supplementary Figure 1). A Fisher’s exact test, with multiple correction for false discovery rate, confirmed that no functional categories were significantly enriched or depleted in the 1 to 2-cell transcriptome compared with the complete *L. stagnalis* 1 to 2-cell transcriptome (Fig. 1; Table 5; Supplementary Table 2). In particular, GO terms associated with nucleotide metabolism and binding in general were overrepresented in the COMAT subset (Figure 1; Table 5; Supplementary Table 2). The maternal expression of a selected set of the COMAT genes was validated in one-cell zygotes using *in situ* methods (Fig. 2).

**Comparison with human housekeeping genes**

The COMAT subset was compared to 3802 well-characterised human housekeeping genes (Eisenberg and Levanon, 2013). All but 38 of the 481 COMAT transcript had a significant match to this set (92%), indicating that the majority are housekeeping in function, at least in humans. In comparison, of the 4,311 *L. stagnalis* 1 to 2-cell transcripts that had a significant BLASTx match in the NCBI nr protein database, only 2,165 (50%) also had matches to the human housekeeping gene dataset. The conserved maternal gene dataset is therefore highly enriched for putative housekeeping genes (Fisher’s exact test, 2156:4311 versus 443:481, *P* < 0.0001).

We wished to understand if a particular subset of housekeeping genes are over-represented in the COMAT subset, or whether the genes are a random subset of all housekeeping genes. We therefore compared the GO annotations of the 3,802 human housekeeping genes against the subset of 300 human housekeeping genes (Table 6) that were found in the COMAT (a proportion of the COMATs hit the same human gene, hence fewer genes than expected). Similar GO annotations were enriched in this selected pairwise comparison compared with the COMAT as a whole (Supplementary Tables 3 and 4). At the highest level, the same first seven Molecular Functions were found in both *H. sapiens* housekeeping versus *H. sapiens* COMAT, and *L. stagnalis* 1 to 2-cell transcriptome versus *L. stagnalis* COMAT comparisons, with *P* < 5E-8 (Supplementary Table 4; ATP binding, GTPase activity, unfolded protein binding, protein serine/threonine kinase activity, GTP binding, threonine-type endopeptidase activity, and ATP-dependent DNA helicase activity). Similarly, the first seven terms relating to Biological Process were also found (*P* < 5E-8: anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process, protein polyubiquitination, negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle, DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle, antigen processing and presentation of exogenous peptide antigen via MHC class I, and TAP-dependent, GTP catabolic process). Thus, the overall conclusion is that the COMAT generally consists of housekeeping genes, but is particularly enriched for a particular subset, including those involved in nucleotide binding functions, protein degradation and activities associated with the cell cycle.

A final concern was that the COMATs

![Fig. 3. Frequency histogram of relative gene expression for human housekeeping genes. Conserved maternal transcripts (COMATs, red line) tend to have a higher gene expression (measured reads per kb per million mapped reads, RPKM) than non-COMATs (blue). However, COMATs still represent several orders of magnitude of gene expression. Gene expression data from Eisenberg & Levanon (2013).](image-url)
are simply conserved genes that tend to be highly expressed, and so are more likely to be detected in non-exhaustive sequencing experiments. We therefore used the expression data of Eisenberg & Levanon (2013) to compare the read depth of these two types of gene (COMATS and non-COMATS) in human tissues. Overall, COMATS tend to be more highly expressed, but they represent a set of genes that have a large range in their quantitative gene expression (Figure 3). Thus, while the mean gene expression in the conserved data set is higher (COMAT mean log geometric expression of gene (COMATS and non-COMATS) in human tissues. Overall, COMATS tend to be more highly expressed, but they represent a set of genes that have a large range in their quantitative gene expression (Figure 3). Thus, while the mean gene expression in the conserved data set is higher (COMAT mean log geometric expression). We find that COMATS tend to be more highly expressed in sequencing experiments than non-COMATS, but this is not a surprising result as the COMATs tend to be highly conserved regions enriched in the conserved maternal genome. A second, less surprising result is that the individual variation is considerable in both datasets (S.D. 0.51 and 0.47 respectively). Thus, a lack of depth in sequencing experiments cannot wholly explain the existence of COMATS.

Discussion

Much excitement has been caused by the discovery that the evolution of gene expression patterns seems to underpin the morphological hourglass pattern of both plants and animals (Kalinka et al., 2010, Meyerowitz, 2002, Quint et al., 2012). Thus, the long-standing observation that vertebrate morphology is at its most

**TABLE 5**

**HIGHEST LEVEL GENE ONTOLOGY TERMS ENRICHED IN THE CONSERVED MATERNAL DATASET**

<table>
<thead>
<tr>
<th>GO-ID</th>
<th>Term*</th>
<th>Category</th>
<th>FDR</th>
<th>P-Value after FDR</th>
<th>Number in test group</th>
<th>Number in cell reference</th>
<th>Number in reference total</th>
<th>Number not annotated in test</th>
<th>Number not annotated reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005524</td>
<td>ATP binding</td>
<td>F</td>
<td>2.93E-03</td>
<td>5.84E-06</td>
<td>119</td>
<td>136</td>
<td>255</td>
<td>271</td>
<td>1953</td>
</tr>
<tr>
<td>GO:0005544</td>
<td>ATPase family</td>
<td>F</td>
<td>2.93E-03</td>
<td>1.08E-07</td>
<td>42</td>
<td>28</td>
<td>70</td>
<td>348</td>
<td>2001</td>
</tr>
<tr>
<td>GO:0051082</td>
<td>unfolded protein binding</td>
<td>F</td>
<td>5.10E-11</td>
<td>2.75E-13</td>
<td>24</td>
<td>9</td>
<td>33</td>
<td>366</td>
<td>2080</td>
</tr>
<tr>
<td>GO:0008026</td>
<td>ATP-dependent helicase activity</td>
<td>F</td>
<td>6.39E-09</td>
<td>7.10E-11</td>
<td>24</td>
<td>15</td>
<td>39</td>
<td>366</td>
<td>2074</td>
</tr>
<tr>
<td>GO:0009234</td>
<td>GTPase activity</td>
<td>E</td>
<td>1.41E-08</td>
<td>1.61E-10</td>
<td>25</td>
<td>18</td>
<td>43</td>
<td>365</td>
<td>2071</td>
</tr>
<tr>
<td>GO:0007355</td>
<td>peptransferal cis-trans isomerase activity</td>
<td>E</td>
<td>5.29E-07</td>
<td>7.72E-09</td>
<td>14</td>
<td>4</td>
<td>18</td>
<td>376</td>
<td>2085</td>
</tr>
<tr>
<td>GO:0007676</td>
<td>sphingomyelin phosphodiesterase activity</td>
<td>E</td>
<td>1.05E-04</td>
<td>2.28E-06</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>383</td>
<td>2089</td>
</tr>
<tr>
<td>GO:0002498</td>
<td>threonine-type endopeptidase activity</td>
<td>E</td>
<td>1.21E-04</td>
<td>2.74E-06</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>392</td>
<td>2088</td>
</tr>
<tr>
<td>GO:0004842</td>
<td>ubiquitin-protein ligase activity</td>
<td>E</td>
<td>1.09E-03</td>
<td>2.96E-05</td>
<td>15</td>
<td>17</td>
<td>32</td>
<td>375</td>
<td>2072</td>
</tr>
<tr>
<td>GO:0052000</td>
<td>structural constituent of cytoskeleton</td>
<td>E</td>
<td>2.95E-03</td>
<td>8.91E-05</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>384</td>
<td>2088</td>
</tr>
<tr>
<td>GO:0008568</td>
<td>microtubule-severing ATPase activity</td>
<td>E</td>
<td>3.06E-03</td>
<td>9.43E-05</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>385</td>
<td>2089</td>
</tr>
<tr>
<td>GO:0042888</td>
<td>MHC class I protein binding</td>
<td>F</td>
<td>1.50E-02</td>
<td>6.05E-04</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>386</td>
<td>2089</td>
</tr>
<tr>
<td>GO:0055208</td>
<td>FK506 binding</td>
<td>F</td>
<td>1.50E-02</td>
<td>6.05E-04</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>386</td>
<td>2089</td>
</tr>
<tr>
<td>GO:0018999</td>
<td>enzyme binding</td>
<td>E</td>
<td>1.92E-02</td>
<td>8.08E-04</td>
<td>24</td>
<td>56</td>
<td>80</td>
<td>366</td>
<td>2033</td>
</tr>
<tr>
<td>GO:0006767</td>
<td>nucleic acid binding</td>
<td>E</td>
<td>2.13E-02</td>
<td>9.21E-04</td>
<td>80</td>
<td>293</td>
<td>373</td>
<td>310</td>
<td>1796</td>
</tr>
<tr>
<td>GO:0007264</td>
<td>small GTPase mediated signal transduction</td>
<td>P</td>
<td>1.78E-10</td>
<td>1.24E-12</td>
<td>25</td>
<td>12</td>
<td>37</td>
<td>365</td>
<td>2077</td>
</tr>
<tr>
<td>GO:0051258</td>
<td>protein phosphorylation</td>
<td>P</td>
<td>2.72E-07</td>
<td>3.75E-09</td>
<td>19</td>
<td>11</td>
<td>30</td>
<td>371</td>
<td>2078</td>
</tr>
<tr>
<td>GO:0006184</td>
<td>GTP catabolic process</td>
<td>P</td>
<td>8.65E-07</td>
<td>1.32E-06</td>
<td>24</td>
<td>23</td>
<td>47</td>
<td>366</td>
<td>2066</td>
</tr>
<tr>
<td>GO:0004143</td>
<td>protein peptidyl-prolyl isomerization</td>
<td>P</td>
<td>2.30E-06</td>
<td>3.94E-08</td>
<td>13</td>
<td>4</td>
<td>17</td>
<td>377</td>
<td>2085</td>
</tr>
<tr>
<td>GO:0006468</td>
<td>protein phosphorylation</td>
<td>P</td>
<td>2.76E-06</td>
<td>4.87E-08</td>
<td>34</td>
<td>52</td>
<td>86</td>
<td>356</td>
<td>2037</td>
</tr>
<tr>
<td>GO:0006200</td>
<td>ATP catabolic process</td>
<td>P</td>
<td>5.78E-04</td>
<td>1.50E-05</td>
<td>16</td>
<td>18</td>
<td>34</td>
<td>374</td>
<td>2071</td>
</tr>
<tr>
<td>GO:0031145</td>
<td>anaphase-promoting complex-dependent</td>
<td>P</td>
<td>1.83E-03</td>
<td>5.19E-05</td>
<td>9</td>
<td>5</td>
<td>14</td>
<td>381</td>
<td>2084</td>
</tr>
</tbody>
</table>

* ordered by category and significance
conserved during the embryonic pharyngula or phylotypic period is generally mirrored by conserved expression patterns of conserved genes at these stages (Kalinka and Tomancak, 2012, Kalinka et al., 2010). In contrast, active transcription in the early zygote is much more limited. Early animal embryos instead largely rely upon RNAs and proteins provided by the maternal gonad during oocyte maturation. This transcriptionally-quiescent period might, a priori, be considered evolutionarily constrained, as the maternally provided transcriptome is widely considered to fulfill one major role, the initiation and management of several rounds of rapid cell division. Every one of these early cell divisions is a critical event that must be faithfully completed to ensure the development of a healthy embryo (Evsikov et al., 2006).

Few studies have investigated the level of conservation of maternally provided genes (Shen-Orr et al., 2010), despite their well-recognised importance in early development (Wieschaus, 1996). Indeed there are few comprehensive datasets of maternally provisioned transcripts even in well-characterised taxa, and none in the Lophotrochozoa. Improvements in sequencing technologies mean that quantitative transcriptome studies are now possible on organisms that lack genomic resources. Our work therefore provides a list of conserved maternal transcripts, or COMATs (Table 6; Supplementary Table 1), that may be useful to the wider community interested in the study of early bilaterian development.

We identified a core set of COMATs from seven representatives of the three bilaterian superphyla, spanning >600 million years of evolution (Peterson et al., 2008). These species display highly divergent modes of development (from direct to indirect, and mosaic to regulative). Since the L. stagnalis maternal transcriptome we report here is unlikely to be complete, one possibility is that our estimate of 5-10% of all maternally provisioned transcripts being conserved across the Bilateria may rise upon deeper sampling of the snail transcriptome. Conversely, the number may reduce as maternal transcriptomes from more taxa are included in the analysis.

Unsurprisingly, we found that many of these genes had nucleotide (especially ATP and GTP) binding functions, were associated with protein degradation or had activities associated with the cell cycle (Table 6). The majority of functions ascribed are probably accurately defined as housekeeping (Eisenberg and Levenon, 2013). One possibility is that some of the most conserved maternal RNAs are those that cannot be provided (solely) as proteins. Cell cycle genes may be illustrative, because some cell cycle proteins are degraded every cycle and so maternal protein alone cannot be sufficient. Finally, the fact that the ~32-cell transcriptome was neither enriched nor underrepresented for any gene ontology relative to the 1 to 2-cell transcriptome, along with a relative over-representation of maternal-zygotic transcripts that are conserved between M. musculus / C. elegans and L. stagnalis suggests that the same transcripts are at least still present during early zygotic transcription (Supplementary Figure 1).

Given the wide variety of developmental modes and rates displayed by metazoan embryos, as well as the hourglass theory of evolution (Kalinka and Tomancak, 2012), one view is that we might expect to find relatively few deeply conserved maternal transcripts. Alternatively, as it has been documented that a relatively large fraction (between 45% and 75%) of all genes within a species’ genome can be found as maternal transcripts (see references within Tadros and Lipshitz, 2009), another view is that maternal transcripts that are conserved between different organisms may be a stochastic subset of a large maternal transcriptome. Instead, our analyses suggest that there is a core and specific set of maternal transcripts that may be essential for early cell divisions, irrespective of the precise mode of development.

While both our data and the others utilised in this study have obvious limitations, primarily the limited sequencing coverage, it is thus uncertain whether further investigation will reveal a greater or lesser proportion of conserved maternal transcripts. However, a simultaneous consideration is that we have detected those genes that are conserved and transcribed at a relatively high level across all taxa, since the study is at best partially quantitative. Further studies are warranted to reveal the true nature of this conservatism. Nonetheless, as we found that the conserved maternal part of a well annotated group of H. sapiens housekeeping genes is enriched for precisely the same functions (Table 6, Supplementary Table 3), we can robustly conclude that there is undoubtedly highly conserved gene expression in the early development of bilaterian embryos. There may also be a distinct set of genes, with mostly housekeeping and nucleotide metabolic functions, that is a necessary starting point of the maternal-to-zygotic transition.

Our analyses thus suggest that the ancestral function of maternal provisioning in animal eggs is to supply the zygote with the materials with which to perform the basic cellular functions of rapid cell division in the early stages of development. The extent of the provisioning is evolutionarily labile, with species that have evolved rapid development relying more on maternal products. Addition of patterning molecules is phylogenetically contingent: as different groups and species have evolved different mechanisms of patterning the embryo and been under selection for fast patterning (as in lineage-driven, or mosaic development) or delayed patterning (as in species with regulative development), so the role of maternal factors in driving patterning has changed.

**Materials and Methods**

**cDNA library construction**

Early development in the pond snail L. stagnalis has been described in exquisite morphological and cytological detail (Raven, 1966). However, the L. stagnalis MZT has not been mapped in the same detail as in model species, but transcription from zygotic nuclei was first detected in 8-cell embryos, and major transcriptional activity detected at the 24-cell stage (Morrill, 1982). While division cycles are not as rapid as development in C. elegans or D. melanogaster, the L. stagnalis embryo does not divide for ~3 hour at the 24-cell stage, suggesting this may represent a shift from maternal to zygotic control. We thus separately sampled 1 to 2-cell and ~32-cell stage L. stagnalis embryos from a laboratory stock maintained in Nottingham, representing the maternal component and the early stages of zygotic transcription. Zygotes were manually dissected out of their egg capsules and stored in RNAlater (Ambion). As one embryo was expected to yield ~0.5 ng RNA, more than one thousand individual embryos of each type were pooled. Total RNA was then extracted using the Qiagen RNeasy Plus Micro Kit. cDNA was then synthesised and two non-normalised cDNA libraries were constructed using the MINT system (Evrogen). The libraries were then processed for sequencing on the Roche 454 FLX platform in the Edinburgh Genomics facility, University of Edinburgh. The raw data have been submitted to the European Nucleotide Archive under bioproject PRJEB7773.

**Transcriptome assembly**

The raw Roche 454 data were screened for MINT and sequencing adapters and trimmed of low quality base calls. The reads from each library were
<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession</th>
<th>Description</th>
<th>Gene</th>
<th>Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTHR</td>
<td>NM_002454</td>
<td>S-methylthiolytrolate-homocysteine methyltransferase</td>
<td>MTHF1</td>
<td>NM_001016</td>
<td>S-methylthiolytrolate-homocysteine methyltransferase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reductase</td>
<td></td>
<td></td>
<td>reductase</td>
</tr>
<tr>
<td>ACAD9</td>
<td>NM_014409</td>
<td>Acyl-Coenzyme A dehydrogenase family, member 9</td>
<td>ACADVL</td>
<td>NM_006518</td>
<td>Acyl-Coenzyme A dehydrogenase, very long chain</td>
</tr>
<tr>
<td>ARF1</td>
<td>NM_001658</td>
<td>ARF-ribosylation factor 1</td>
<td>ARF5</td>
<td>NM_001662</td>
<td>ARF-ribosylation factor 5</td>
</tr>
<tr>
<td>ARF6</td>
<td>NM_001663</td>
<td>ARF-ribosylation factor 6</td>
<td>ARF6AP3</td>
<td>NM_001679</td>
<td>ARF-ribosylation factor</td>
</tr>
<tr>
<td>ARL1</td>
<td>NM_001777</td>
<td>ARL-ribosylation factor 1-like</td>
<td>AXL</td>
<td>NM_001124</td>
<td>AXL</td>
</tr>
<tr>
<td>AHS1A</td>
<td>NM_012111</td>
<td>AHA1, activator of heat shock 90Da protein ATPasome</td>
<td>APOE</td>
<td>NM_000961</td>
<td>APOE</td>
</tr>
<tr>
<td>ALDH8A</td>
<td>NM_006969</td>
<td>Aldehyde dehydrogenase 9 family, member A1</td>
<td>AIMP</td>
<td>NM_01087</td>
<td>AIMP</td>
</tr>
<tr>
<td>ANKRD17</td>
<td>NM_02217</td>
<td>Ankyrin repeat domain 17</td>
<td>ANKRD28</td>
<td>NM_011908</td>
<td>Ankyrin repeat domain 28</td>
</tr>
<tr>
<td>AR1A</td>
<td>NM_003491</td>
<td>AR10 homolog A, N-acetyltransferase (S. cerevisiae)</td>
<td>ACTR1A</td>
<td>NM_005736</td>
<td>ACTR1A, protein 1, alpha</td>
</tr>
<tr>
<td>ACTR1B</td>
<td>NM_001056</td>
<td>ACTR1B, protein 1, delta</td>
<td>ARNT</td>
<td>NM_001669</td>
<td>ARNT, auxiliary factor alpha</td>
</tr>
<tr>
<td>ARTS</td>
<td>NM_004046</td>
<td>ATP synthase, H+ transporting, mitochondrial F1 complex, alpha and beta</td>
<td>ATP5B</td>
<td>NM_001680</td>
<td>ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide</td>
</tr>
<tr>
<td>ATAD1</td>
<td>NM_023810</td>
<td>ATPasome-associated family, AAA domain containing 1</td>
<td>ATBC2</td>
<td>NM_012089</td>
<td>ATP-binding cassette, sub-family B (MRD-TAP), member 10</td>
</tr>
<tr>
<td>ABCB7</td>
<td>NM_004299</td>
<td>ATP-binding cassette, sub-family B (MRD-TAP), member 7</td>
<td>BDXC5</td>
<td>NM_025005</td>
<td>BDXC5, bixin domain containing 5</td>
</tr>
<tr>
<td>BDRD7</td>
<td>NM_013263</td>
<td>Bromodomain containing 7</td>
<td>BPF1</td>
<td>NM_004459</td>
<td>Bromodomain PHD finger transcription factor</td>
</tr>
<tr>
<td>BUB3</td>
<td>NM_004725</td>
<td>BUB3, budding uninitiated by benznidazole 3 homolog</td>
<td>CAS9</td>
<td>NM_012899</td>
<td>CAS9, calcium binding protein 34</td>
</tr>
<tr>
<td>CALU</td>
<td>NM_001219</td>
<td>Calumenin</td>
<td>CYB5B</td>
<td>NM_035083</td>
<td>CDC2-related kinase, arginine/serine-rich</td>
</tr>
<tr>
<td>CJKD</td>
<td>NM_003418</td>
<td>CDC5-type zinc finger, nucleo acid binding protein</td>
<td>CNDP1</td>
<td>NM_001780</td>
<td>CNDP1, CDC5-like (Pombe)</td>
</tr>
<tr>
<td>CDDX</td>
<td>NM_007460</td>
<td>CDDX, cyclin-dependent kinase 9</td>
<td>CDC37</td>
<td>NM_007085</td>
<td>CDC37, homolog (S. cerevisiae)</td>
</tr>
<tr>
<td>CDC42</td>
<td>NM_004730</td>
<td>CDC42, GTP-binding protein, 28kDa</td>
<td>CDC5L</td>
<td>NM_012535</td>
<td>CDC5L, CDC5-like (S. pombe)</td>
</tr>
<tr>
<td>CDC63</td>
<td>NM_003992</td>
<td>CDC-like kinase 3</td>
<td>CTCF</td>
<td>NM_006565</td>
<td>CTCF, CCCTC-binding factor (zinc finger protein)</td>
</tr>
<tr>
<td>CDNB</td>
<td>NM_002436</td>
<td>CDNB, constitutive phosphomorphogen homolog 2 (Arabidopsis)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDNS</td>
<td>NM_001172</td>
<td>Chaperonin containing TCP1, subunit 3 (gamma)</td>
<td>CTC7</td>
<td>NM_005998</td>
<td>Chaperonin containing TCP1, subunit (gamma)</td>
</tr>
<tr>
<td>CTCP</td>
<td>NM_001762</td>
<td>Chaperonin containing TCP1, subunit 6a (zeta)</td>
<td>CTCT7</td>
<td>NM_004259</td>
<td>Chaperonin containing TCP1, subunit 7 (eta)</td>
</tr>
<tr>
<td>CTCT8</td>
<td>NM_006855</td>
<td>Chaperonin containing TCP1, subunit 8 (theta)</td>
<td>CDH4</td>
<td>NM_001273</td>
<td>Chromodomain helicase DNA-binding domain 4</td>
</tr>
<tr>
<td>CTR7</td>
<td>NM_010207</td>
<td>Chaperonin containing TCP1, subunit 5 (epsilon)</td>
<td>CTR7A</td>
<td>NM_012899</td>
<td>Chromatidic helicase DNA-binding domain 4</td>
</tr>
<tr>
<td>CTR8</td>
<td>NM_010749</td>
<td>Chromatidic helicase DNA-binding domain 4</td>
<td>CTR8</td>
<td>NM_012899</td>
<td>Chromatidic helicase DNA-binding domain 4</td>
</tr>
<tr>
<td>CTS1</td>
<td>NM_013240</td>
<td>Cleavage stimulation factor, 3, pre-RNA, subunit 1, 50kDa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CST2</td>
<td>NM_015235</td>
<td>Cleavage stimulation factor, 3, pre-RNA, subunit 2, 64kDa, tau variant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPA</td>
<td>NM_004371</td>
<td>Cotomin protein complex, subunit alpha</td>
<td>COP2S</td>
<td>NM_004309</td>
<td>COP2S, constitutive phosphomorphogen homolog 2 (Arabidopsis)</td>
</tr>
<tr>
<td>CTR4</td>
<td>NM_001121</td>
<td>Cyclin-dependent kinase 9</td>
<td>CTDS2P</td>
<td>NM_005730</td>
<td>CTD (carboxy-terminal domain, RNA polymerase ii, polypeptide A) small phosphatase 2</td>
</tr>
<tr>
<td>CBX5</td>
<td>NM_003079</td>
<td>Cytochrome b5 type B (outer mitochondrial membrane)</td>
<td>CPY2U1</td>
<td>NM_183075</td>
<td>Cytochrome P450, family 2, subfamily U, polypeptide 1</td>
</tr>
<tr>
<td>CYP2U3</td>
<td>NM_011997</td>
<td>Cytochrome P450, family 2, subfamily U, polypeptide 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXA1</td>
<td>NM_008459</td>
<td>DDX1, DEAD-associated protein A</td>
<td>DXA1B</td>
<td>NM_008596</td>
<td>DDX1, DEAD-associated protein A</td>
</tr>
<tr>
<td>DXA2</td>
<td>NM_008247</td>
<td>DDX2, DEAD-associated protein 2</td>
<td>DXA2</td>
<td>NM_008247</td>
<td>DDX2, DEAD-associated protein 2</td>
</tr>
<tr>
<td>DXA3</td>
<td>NM_008248</td>
<td>DDX3, DEAD-associated protein 3</td>
<td>DXA3</td>
<td>NM_008248</td>
<td>DDX3, DEAD-associated protein 3</td>
</tr>
<tr>
<td>DXA4</td>
<td>NM_008249</td>
<td>DDX4, DEAD-associated protein 4</td>
<td>DXA4</td>
<td>NM_008249</td>
<td>DDX4, DEAD-associated protein 4</td>
</tr>
<tr>
<td>DXA5</td>
<td>NM_008250</td>
<td>DDX5, DEAD-associated protein 5</td>
<td>DXA5</td>
<td>NM_008250</td>
<td>DDX5, DEAD-associated protein 5</td>
</tr>
</tbody>
</table>

**Table 6:** The 300 Human Genes in the Conserved Maternal Dataset

- **Gene:** The gene symbol.
- **Accession:** The accession number from the database.
- **Description:** The description of the gene function.

- **Protein:** The protein function.
- **Function:** The functional category.
- **Gene:** The gene symbol.
- **Accession:** The accession number from the database.
- **Description:** The description of the gene function.

- **Protein:** The protein function.
- **Function:** The functional category.
### TABLE 6 (CONTINUED)

#### THE 300 HUMAN GENES IN THE CONSERVED MATERNAL DATASET

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_003753</td>
<td>Eukaryotic translation initiation factor 3, subunit D</td>
<td>NM_014409</td>
<td>TAF5-like RNA polymerase II, p300/CBP-associated factor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_002227</td>
<td>Neurofibromin 2 (bilateral acoustic neuroma) regulated 8</td>
</tr>
<tr>
<td>NM_013642</td>
<td>Mitogen regulated 8</td>
</tr>
<tr>
<td>NM_006597</td>
<td>Minichromosome maintenance complex component 7</td>
</tr>
<tr>
<td>NM_006625</td>
<td>Heterogeneous nuclear ribonucleoprotein D (AU box and WD repeat domain containing 1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_018981</td>
<td>DNAJ (Hsp40) homolog, subfamily C, member 10</td>
</tr>
<tr>
<td>NM_018163</td>
<td>DNAJ (Hsp40) homolog, subfamily C, member 17</td>
</tr>
<tr>
<td>NM_025219</td>
<td>DNAJ (Hsp40) homolog, subfamily C, member 5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_030640</td>
<td>Dual specificity phosphatase 16</td>
</tr>
<tr>
<td>NM_01419</td>
<td>ELAV (elamiph) embryonic lethal, abnormal vision, Drosophila-like 1 (Hu anti-r)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EFA</th>
<th>Electron-transfer-flavoprotein, alpha polypeptide (glutamic aciduria SPG7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_004092</td>
<td>Enoyl-CoA hydratase, short chain, 1, mitochondrial</td>
</tr>
<tr>
<td>NM_015700</td>
<td>ERYC and gold 2</td>
</tr>
<tr>
<td>NM_01961</td>
<td>Eukaryotic translation elongation factor 2</td>
</tr>
<tr>
<td>NM_021300</td>
<td>F-box and WD repeat domain containing 11</td>
</tr>
<tr>
<td>NM_021623</td>
<td>F-box and WD repeat domain containing 1</td>
</tr>
<tr>
<td>NM_002013</td>
<td>FK506 binding protein 3, 25kDa</td>
</tr>
<tr>
<td>NM_030280</td>
<td>Fyn homolog 1 (E. coli)</td>
</tr>
<tr>
<td>NM_006625</td>
<td>Fusulating protein (serine/arginine-rich-1)</td>
</tr>
<tr>
<td>NM_001514</td>
<td>General transcription factor IB</td>
</tr>
<tr>
<td>NM_006625</td>
<td>Glucosamine-6-phosphate deaminase 1</td>
</tr>
<tr>
<td>NM_012820</td>
<td>Glutamate-rich WD repeat containing 1</td>
</tr>
<tr>
<td>NM_012820</td>
<td>Glutamate-rich WD repeat containing 1</td>
</tr>
<tr>
<td>NM_02341</td>
<td>GTP binding protein 4</td>
</tr>
<tr>
<td>NM_02701</td>
<td>Histone acetyltransferase 1</td>
</tr>
<tr>
<td>NM_021422</td>
<td>Histone histone family, member V</td>
</tr>
<tr>
<td>NM_005916</td>
<td>Histone H3-like (C. elegans)</td>
</tr>
<tr>
<td>NM_019648</td>
<td>Heat shock 10kDa protein 1 (Drosophila)</td>
</tr>
<tr>
<td>NM_05347</td>
<td>Heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)</td>
</tr>
<tr>
<td>NM_05347</td>
<td>Heat shock 70kDa protein 8</td>
</tr>
<tr>
<td>NM_00712</td>
<td>Hepatocyte growth factor-regulated tyrosine kinase substrate</td>
</tr>
<tr>
<td>NM_003246</td>
<td>Heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein)</td>
</tr>
<tr>
<td>NM_003246</td>
<td>Histone acetyltransferase 1</td>
</tr>
<tr>
<td>NM_004460</td>
<td>HLA-B associated transcript 1</td>
</tr>
<tr>
<td>NM_03416</td>
<td>IMP4, US small nucleolar ribonucleoprotein, homolog (yeast)</td>
</tr>
<tr>
<td>NM_002227</td>
<td>Janus kinase 1 (a protein tyrosine kinase)</td>
</tr>
<tr>
<td>NM_002264</td>
<td>Karyopherin alpha 1 (importin alpha 5)</td>
</tr>
<tr>
<td>NM_021488</td>
<td>Leucine rich repeat containing 47</td>
</tr>
<tr>
<td>NM_014288</td>
<td>Mitochondrial ribosomal protein L4</td>
</tr>
<tr>
<td>NM_02745</td>
<td>Mitogen-activated protein kinase 1</td>
</tr>
<tr>
<td>NM_00754</td>
<td>Mitochondrial maintenance complex component 7</td>
</tr>
<tr>
<td>NM_005916</td>
<td>Mitochondrial ribosomal protein L4</td>
</tr>
<tr>
<td>NM_002745</td>
<td>Mitogen-activated protein kinase 1</td>
</tr>
<tr>
<td>NM_002745</td>
<td>Mitogen-activated protein kinase 10</td>
</tr>
<tr>
<td>NM_00754</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_002754</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_001274</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_002083</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_002083</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_002083</td>
<td>Mitogen-activated protein kinase 2</td>
</tr>
<tr>
<td>NM_003148</td>
<td>Leucine-rich repeat containing 47</td>
</tr>
<tr>
<td>NM_002487</td>
<td>Myc-associated protein Y (serine/threonine kinase receptor associated protein)</td>
</tr>
<tr>
<td>NM_003148</td>
<td>Myc-associated protein Y (serine/threonine kinase receptor associated protein)</td>
</tr>
<tr>
<td>NM_005934</td>
<td>Multiformyloid or mixed-lineage leukemia (trichorax homolog, Drosophila)</td>
</tr>
<tr>
<td>NM_001657</td>
<td>Myoneurin</td>
</tr>
<tr>
<td>NM_004988</td>
<td>Myosin IE</td>
</tr>
<tr>
<td>NM_004988</td>
<td>Myosin IE</td>
</tr>
<tr>
<td>NM_003828</td>
<td>Myotubulin related protein 1</td>
</tr>
<tr>
<td>NM_003288</td>
<td>Myotubulin related protein 1</td>
</tr>
<tr>
<td>NM_002496</td>
<td>NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa</td>
</tr>
<tr>
<td>NM_003421</td>
<td>Nucleolar protein 28kDa, ID1 regulated</td>
</tr>
<tr>
<td>NM_000286</td>
<td>Neurofibrin 2 (bilateral acoustic neuroma)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_0012746</td>
<td>Ribosomal protein 56, kinase, 90Da, polypeptide 2</td>
</tr>
<tr>
<td>NM_003459</td>
<td>Ribosomal protein 56, kinase, 90Da, polypeptide 3</td>
</tr>
<tr>
<td>NM_003368</td>
<td>S-adenosylhomocysteine hydrolase</td>
</tr>
<tr>
<td>NM_0015356</td>
<td>Schnibb homolog (Drosophila)</td>
</tr>
<tr>
<td>NM_007178</td>
<td>Serine/threonine kinase receptor associated protein</td>
</tr>
<tr>
<td>NM_003832</td>
<td>SET domain containing (lysine methyltransferase)</td>
</tr>
<tr>
<td>NM_005903</td>
<td>SMAD family member 5</td>
</tr>
<tr>
<td>NM_018126</td>
<td>Smu 1 suppressor of mec-8 and unc-52 homolog (C. elegans)</td>
</tr>
<tr>
<td>NM_003562</td>
<td>Solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier)</td>
</tr>
<tr>
<td>NM_016016</td>
<td>Solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier)</td>
</tr>
<tr>
<td>NM_006979</td>
<td>Solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_03119</td>
<td>Spastic paraplegia 7 (pure and complicated autosomal recessive)</td>
</tr>
<tr>
<td>NM_02463</td>
<td>Spermatogenesis associated 5-like</td>
</tr>
<tr>
<td>NM_003016</td>
<td>Splicing factor, arginine/serine-rich 2</td>
</tr>
<tr>
<td>NM_005500</td>
<td>SUMO1 activating enzyme subunit 1</td>
</tr>
<tr>
<td>NM_005499</td>
<td>SUMO1 activating enzyme subunit 2</td>
</tr>
<tr>
<td>NM_014409</td>
<td>TAF5-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa</td>
</tr>
<tr>
<td>NM_025235</td>
<td>Tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2</td>
</tr>
<tr>
<td>NM_003685</td>
<td>T-cell 1</td>
</tr>
<tr>
<td>NM_0012473</td>
<td>TFIIA</td>
</tr>
<tr>
<td>NM_005783</td>
<td>Thoreidoxin domain containing 9</td>
</tr>
<tr>
<td>NM_003056</td>
<td>Triosephosphate isomerase 1</td>
</tr>
<tr>
<td>NM_0172049</td>
<td>TRAF3 receptor-associated protein 1</td>
</tr>
<tr>
<td>NM_014820</td>
<td>Translase of the maternal mitochondrial membrane 70 homolog A (S. cerevisiae)</td>
</tr>
<tr>
<td>NM_003365</td>
<td>Tyrosine-3-monoxygenase/tryptophan 5-monoxygenase activation protein, beta polypeptide</td>
</tr>
<tr>
<td>NM_003335</td>
<td>Ubiquitin A-52 residue ribosomal protein fusion product 1</td>
</tr>
<tr>
<td>NM_018955</td>
<td>Ubiquitin B</td>
</tr>
<tr>
<td>NM_012009</td>
<td>Ubiquitin C</td>
</tr>
<tr>
<td>NM_014671</td>
<td>Ubiquitin protein ligase E3C</td>
</tr>
<tr>
<td>NM_003988</td>
<td>Ubiquitin-activating enzyme E1C (UBA3 homolog, yeast)</td>
</tr>
<tr>
<td>NM_021898</td>
<td>Ubiquitin-conjugating enzyme E2 variant 1</td>
</tr>
<tr>
<td>NM_003336</td>
<td>Ubiquitin-conjugating enzyme E2A (RAD6 homolog)</td>
</tr>
<tr>
<td>NM_003337</td>
<td>Ubiquitin-conjugating enzyme E2B (RAD6 homolog)</td>
</tr>
<tr>
<td>NM_003340</td>
<td>Ubiquitin-conjugating enzyme E2D (UBC4/5 homolog, yeast)</td>
</tr>
<tr>
<td>NM_003344</td>
<td>Ubiquitin-conjugating enzyme E2G (UBC4/5 homolog, yeast)</td>
</tr>
<tr>
<td>NM_00100454</td>
<td>Yippe-like 2 (Drosophila)</td>
</tr>
<tr>
<td>NM_014263</td>
<td>YME1-like 1 (S. cerevisiae)</td>
</tr>
<tr>
<td>NM_003403</td>
<td>YtK1 transcription factor</td>
</tr>
<tr>
<td>NM_000249</td>
<td>Multiform, homolog 1, colon cancer, nonpolyposis type 2 (E. coli)</td>
</tr>
<tr>
<td>NM_003365</td>
<td>Zinc finger and BTB domain containing 6</td>
</tr>
<tr>
<td>NM_00131058</td>
<td>Zinc finger protein 37A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_00131058</td>
<td>Zinc finger protein 37A</td>
</tr>
</tbody>
</table>
assembled using gsAssembler (version 2.6; also known as Newbler; 454 Life Sciences) and MIRA (Chevreux et al., 2004) separately, and then the two assemblies were assembled together using CAP3 (Huang and Madan, 1999), following the proposed best practice for transcriptome assembly from 454 data (Kumar and Blaxter, 2010). gsAssembler assemblies were run with the –cdna and –urt options. MIRA assemblies used job options ‘denovo, est, accurate, 454’ and with clipping by quality off (-CL:qc=no). CD-HIT was then used to remove redundant sequences from the merged CAP3 assemblies (Li and Godzik, 2006), running cd-hit-est with sequence identity threshold 0.98 (-c 0.98) and clustering to most similar cluster (-g 1). The assembly has been made available on afterParty (http://afterparty.bio.ed.ac.uk).

**Maternal transcriptomes from other species**

We identified a number of published, high-throughput, maternal transcriptome studies from Ciona intestinalis (Urochordata, Deuterostomia), Danio rerio, Mus musculus, Homo sapiens (Chordata, Deuterostomia), C. elegans (Nematoda, Ecdysozoa) and D. melanogaster (Arthropoda, Ecdysozoa). A “maternal transcript” is an mRNA that is present in the embryo before the initiation of major zygotic transcription. This does not mean that these mRNAs are not also later also transcribed from the zygotic genome in the developing embryo.

We carried out a reciprocal tBLASTx comparison of the L. stagnalis 1 to 2-cell transcriptome against each of the other datasets, using a threshold of expect value of $10^{-16}$. By identifying L. stagnalis transcripts that had homologues in all of the species we identified a putative set of conserved bilaterian maternal transcripts.

**Functional annotation of transcriptome**

The 1 to 2-cell and 32-cell transcriptome assemblies were annotated with gene ontology (GO) terms using Blast2GO v.2.7.0 against the NCBI non-redundant (nr) protein database, with an E-value cutoff of 1e-05. GO term distribution was quantified using the Combined Graph function of the GOseq R package against each of the other datasets, using a threshold of $10^{-16}$. By identifying L. stagnalis transcripts that had homologues in all of the species we identified a putative set of conserved bilaterian maternal transcripts.

**In situ validation of representative transcripts**

We validated the maternal expression of a selection of sequences in L. stagnalis 1-cell embryos by using whole mount in situ hybridisation (WISH). Primers were designed to amplify fragments of selected genes, which were then cloned into pGEM-T and verified by standard Sanger sequencing. Supplementary riboprobes were prepared from these templates as described in Jackson et al., (2007a). The WISH protocol we employed here for L. stagnalis is similar to previously described protocols for molluscan embryos and larvae (Jackson et al., 2006, Jackson et al., 2007b) with some important modifications (described elsewhere; in review). The colour reactions for all hybridisations (including the negative [l-tubulin control) were allowed to proceed for the same length of time, and all samples cleared in 60% glycerol and imaged under a Zeiss Axio Imager Z1 microscope. The primers used are shown in Table 1.

**Acknowledgements**

The authors would like to thank The GenePool Genomics Facility (now Edinburgh Genomics), University of Edinburgh for generating the DNA sequences used in this study. Thanks to Karim Gharbi, Marian Thompson and colleagues at the GenePool, Azzz Aboobaker, as well as Eli Eisenberg for helpful advice on the human housekeeping data. Two anonymous referees provided helpful comments and advice. The work was principally funded by Biotechnology and Biological Sciences Research Council grant BB/F018940/1 to AD and MLB with additional funding provided by the Universities of Edinburgh and Nottingham, the Wellcome Trust Sanger Institute (WT098051), Biotechnology and Biological Sciences Research Council grants G00661X and F021135; Medical Research Council grant (G0900740) and Natural Environmental Research Council grant (R8/H10/56) to MLB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. DJJ is funded by the German Excellence Initiative and DFG project JA 2108/1-2.

**References**


Further Related Reading, published previously in the *Int. J. Dev. Biol.*

**Maternal RNAs encoding transcription factors for germline-specific gene expression in *Drosophila* embryos**
Jun Yatsu, Makoto Hayashi, Masanori Mukai, Kayo Arita, Shuji Shigenobu and Satoru Kobayashi
*Int. J. Dev. Biol.* (2008) 52: 913-923
http://www.intjdevbiol.com/web/paper/082576jy

**Centroid, a novel putative DEAD-box RNA helicase maternal mRNA, is localized in the mitochondrial cloud in *Xenopus laevis* oocytes**
Malgorzata Kloc and Agnes P. Chan
*Int. J. Dev. Biol.* (2007) 51: 701-706
http://www.intjdevbiol.com/web/paper/072293mk

**Differences in maternal supply and early development of closely related nematode species**
Magdalena Laugsch and Einhard Schierenberg
http://www.intjdevbiol.com/web/paper/031758ml

**Identification and characterization of maternally expressed genes with mRNAs that are segregated with the endoplasm of early ascidian embryos**
K Imai, N Satoh and Y Satou
*Int. J. Dev. Biol.* (1999) 43: 125-133
http://www.intjdevbiol.com/web/paper/10235388

**HrWnt-5: a maternally expressed ascidian Wnt gene with posterior localization in early embryos**
Y Sasakura, M Ogasawara and K W Makabe
*Int. J. Dev. Biol.* (1998) 42: 573-579
http://www.intjdevbiol.com/web/paper/9694628

5 yr ISI Impact Factor (2011) = 2.959